INTERACTIONS BETWEEN HOST AND PHAGE ENCODED FACTORS SHAPE PHAGE INFECTION

A Dissertation

by

DENISH KUMAR PIYA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ryland F. Young
Co-Chair of Committee,	Jason J. Gill
Committee Members,	Jennifer Herman
	Lanying Zeng
Head of Department,	Gregory D. Reinhart

December 2018

Major Subject: Biochemistry

ABSTRACT

Evolution of phages and their bacterial hosts are directed by interaction between phage and host-encoded factors. These interactions have resulted in the development of several defense and counter-defense strategies such as DNA restriction and antirestriction systems. Type I restriction-modification (R-M) systems present a barrier to foreign DNA, including phage, entering the bacterial cell, by cleaving inappropriately modified DNA in a sequence-specific manner. Phages have evolved diverse mechanisms to overcome restriction systems. The temperate coliphage P1 encodes virion-associated proteins that protect its DNA from host type I R-M systems. By using genetic and biochemical analysis, it has been established that the P1 Dar (Dar for *d*efense *a*gainst *r*estriction) system is comprised of at least six virion-associated proteins: DarB, Ulx, Hdf, DarA, DdrA and DdrB. DarB protects P1 DNA from *Eco*B and *Eco*K restriction *in cis* by an unknown mechanism and is incorporated into the virions only in the presence of Hdf, DarA and DdrA. Hdf and DarA have also been found to affect capsid morphogenesis, as their absence results in phage progeny with predominantly aberrant small heads. Examination of purified P1 proheads shows that Dar system proteins are incorporated into the virion before DNA packaging, and an N-terminal signal is required for DarB packaging. Twenty-four additional P1 genes of unknown function were disrupted and none were found to alter the antirestriction phenotype. A purification protocol for the ~250 kDa antirestriction protein DarB has been optimized, which will facilitate biochemical approaches for determining its mechanism of action.

While the phage-host interactions of classical phages such as P1 are better understood, relatively little is known about how phages outside of the common paradigm organisms interact with their

hosts. A high throughput genetic screen of the T1-like coliphage LL5 and the rV5-like coliphage LL12 was conducted against the *E. coli* Keio collection. This screen revealed host receptors required for both phages to initiate infection and two chaperones, PpiB and SecB, required for efficient propagation of phage LL5.

DEDICATION

I would like to dedicate this work to my late grandfather. He was my first teacher and taught me to read and write, and always encouraged me to pursue higher education.

ACKNOWLEDGEMENTS

I would like to thank my advisers, Dr. Jason J. Gill and Dr. Ry Young for believing in me and providing me with opportunity to work under their tutelage. Their support and words of wisdom have been very crucial in my journey through graduate school. The work presented in this dissertation would not have possible without their mentorship. Both of you have greatly inspired me to pursue science and excellence. I would also like to thank my committee members, Dr. Jennifer Herman and Dr. Lanying Zeng for their guidance, helpful discussion and comments throughout graduate school.

I would like to thank past and present members of Gill Lab (Jackie Grimm, Season Xie, Abby Korn, Dr. Justin Leavitt, Justin Boeckman, Miguel F. Gonzales) for creating an awesome Lab environment and always being there for helpful discussions and assistance. I would also like to thank my undergraduate research mentees Brian Koehler, Judith Salazar, Leonardo Vara and Ashley Stonecipher for their help in conducting experiments and CPT staff Lauren Lessor for providing phages LL5 and LL12 used in this study. I also thank Dr. Matthew Taylor from Department of Animal Science at Texas A&M University for providing STEC strains used in this study. I also thank Dr. Junjie Zhang and Dr. Timothy Devarenne from Department of Biochemistry & Biophysics and Dr. Carlos F. Gonzalez from Department of Plant Pathology and Microbiology at Texas A&M University for allowing me to use their lab equipment. I also thank Dr. William Russell from Department of Chemistry at Texas A&M University (Currently at The University of Texas Medical Branch at Galveston) and Dr. Lawrence Dangott from Department of Biochemistry & Biophysics at Texas A&M University for assistance with mass spectrometry.

I would like to thank all CPT members and Supergroup members for their feedback and comments on data and presentations. Special thanks to Adriana C. Hernandez, Dr. Hem Thapa, Dr. Karthik Chamakura and Dr. Manoj Rajaure for teaching me several experimental techniques. I would also like to thank all administrative staffs of TAMU BioBio department for their assistance. Special mention to Mrs. Daisy Wilbert for always taking care of everything.

I also thank my teachers, instructors and professors from Notre Dame School (Bandipur, Nepal), Small Heaven School (Narayanghat, Nepal), Budhanilkantha School (Kathmandu, Nepal) and Caldwell University (Caldwell, NJ, U.S.) for providing me with a solid foundation required for this journey. I would also like to thank all of my friends from my schools in Nepal and especially from Caldwell University and Texas A&M University for making life more fun and for being there at tough times.

Finally, thanks for my parents for their encouragement and support throughout my educational journey. This journey has been only possible because their constant motivation since my first day at preliminary school. Thanks for my sisters as well for their support and love. Also, thanks to my wife for her patience and love and for sharing this journey with me.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Ry Young (advisor), Dr. Jennifer Herman and Dr. Lanying Zeng of the Department of Biochemistry & Biophysics and Dr. Jason J. Gill (co-advisor) of the Department of Animal Science. All work for the dissertation was completed independently by the student. Graduate study was supported by start-up funding from Texas AgriLife Research and Texas A&M University to Dr. Jason J. Gill and from Deerland Enzymes.

NOMENCLATURE

E. coli	*Escherichia coli*
R-M	Restriction-Modification

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Bacteriophages

Bacteriophages (phages) are natural predators of bacteria. It has been estimated that ~96% of phages existing in nature belong to order *Caudovirales*, characterized by the presence of double stranded DNA as genomic material in the capsids and tails as infection apparatus (Brussow & Hendrix, 2002; Hatfull & Hendrix, 2011). Tailed phages are the most abundant life forms as they have been estimated to be present in numbers $>10^{30}$ and outnumber their bacterial hosts by five- to ten-fold in environment (Brussow & Hendrix, 2002; Hendrix, 2002). Phages are categorized into different taxonomic groups based upon a system proposed by the International Committee on Taxonomy of Viruses (ICTV) (Nelson, 2004). Phages of order *Caudovirales* are further classified, based on tail morphology into different families: *Myoviridae* (long, straight, contractile tail), *Siphoviridae* (long, flexible, non-contractile tail) and *Podoviridae* (short, stubby, noncontractile tail) (Fig. 1.1) (Maniloff & Ackermann, 1998). This system of classification, based on shared characteristics, however, has various limitations. This system does not consider the genomic and proteomic information, resulting in resulting in grouping of unrelated phages based on their morphology only (Botstein & Herskowitz, 1974; Nelson, 2004). Since phage replication involves genetic recombination, phage genomes are mosaic, which is not considered in this classification system. It is also not feasible to isolate phages for determining its morphological

characteristics when the phage genomes are obtained from community sequencing project (Nelson, 2004).

Advances in structural biology has made analysis of phage structure in atomic resolution possible. Based on the conserved folds of structural proteins, *Caudovirales* phages are suggested to have a common evolutionary origin. The assembly of these virions occur

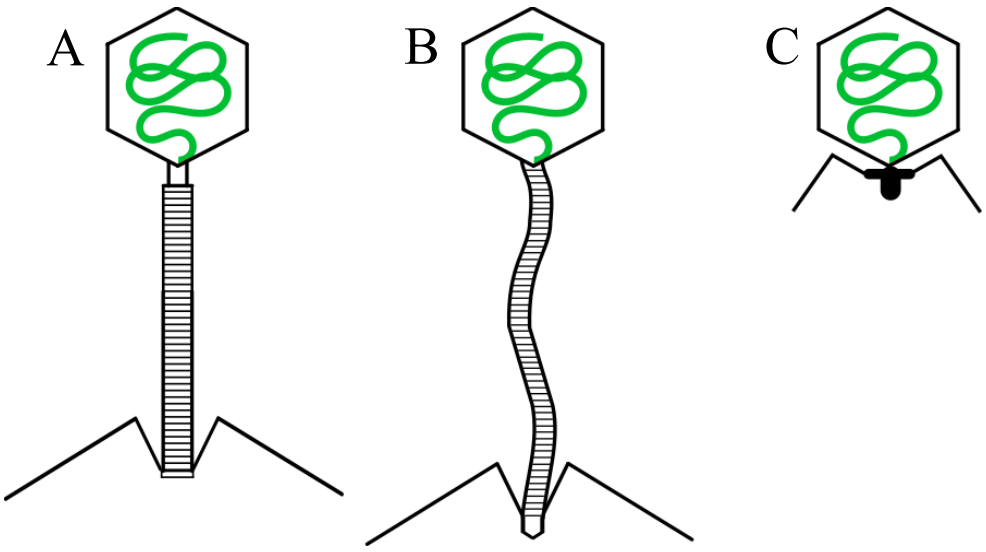


Fig. 1.1. Morphotypes of *Caudovirales* phages. Phages belonging to order *Caudovirales* are divided into three families, based upon characteristics of their tails. **Panel A:** Phages of family *Myoviridae* have long contractile tail. **Panel B:** Phages of family *Siphoviridae* have long, noncontractile but flexible tails. **Panel C:** Phages of family *Podoviridae* have short tails.

from conserved components in a similar pathway (Fig. 1.2) (Fokine & Rossmann, 2014). In general, the capsids are icosahedral and are composed of eleven pentameric, and twenty hexameric capsomers (Mateu, 2013). One pentameric vertex is occupied by a dodecameric ring of portal protein and is called the portal vertex (Aksyuk & Rossmann, 2011). The assembly of capsids start at the portal vertex in presence of scaffolding proteins and major capsid proteins, forming proheads (Fig. 1.2). The scaffolding proteins facilitate the assembly of capsid proteins into a definite geometry (Dokland, 1999). DNA is packaged into procapsids through the ring of portal proteins by DNA packaging machine composed of small and large terminases (Casjens, 2011). Once the DNA is packaged, the terminase complex disassociates and the head completion proteins are attached to the portal vertex. The head completion proteins serve as the attachment site for tails of *Myoviridae* and *Siphoviridae* phages, which are assembled in an independent pathway. The assembly of tails in *Myoviridae* and *Siphoviridae* start at the base plate and the hexamers of tail tube and tail sheath are sequentially added to a certain length as determined by tape measure protein. However, in *Podoviridae* phages, tails are assembled on the portal vertex (Aksyuk & Rossmann, 2011; Fokine & Rossmann, 2014).

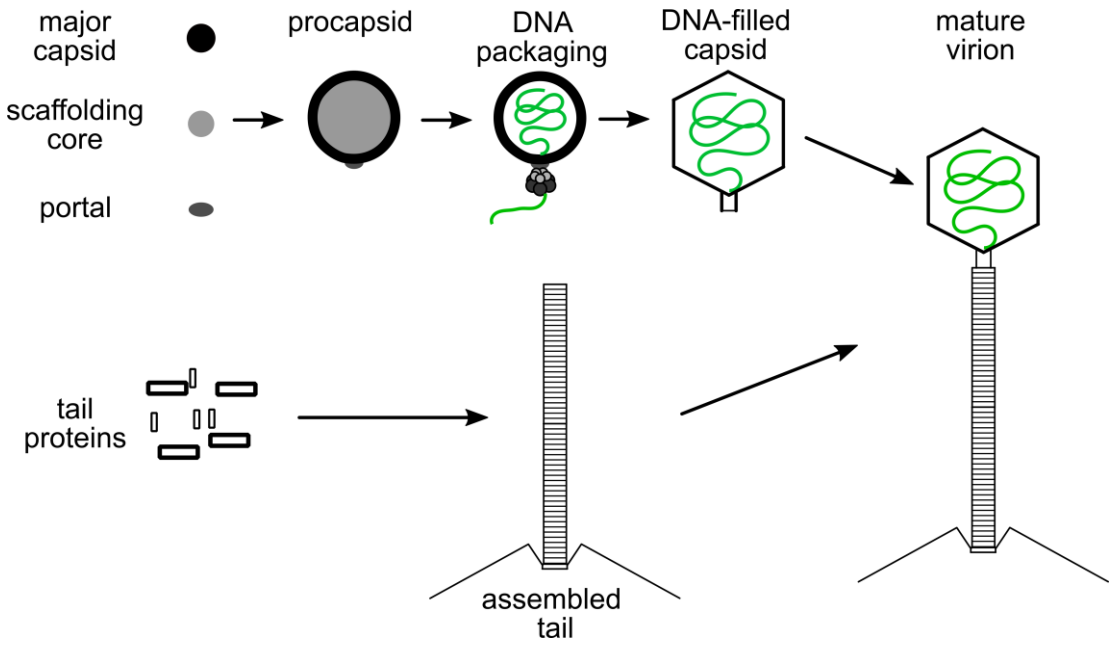


Fig. 1.2. General assembly pathway of *Caudovirales* phages. Schematics represent assembly pathway of *Myoviridae* and *Siphoviridae* phages. Major capsid proteins, scaffolding core and portal proteins assemble into procapsids. DNA is packaged into procapsids by DNA packaging complex, followed by maturation of capsids into icosahedral geometry. Tails are assembled in an independent pathway and are attached to DNA-filled capsids to form complete virions. In *Podoviridae* phages, tails are assembled directly into the portal vertex.

Phage infection cycle starts with adsorption of phages to the specific receptors on cell surface of a host, mediated by the phage receptor-binding protein, also called tail fiber proteins (Dowah & Clokie, 2018). This adsorption event is followed by ejection of phage genetic material into host cytoplasm. Depending upon if the infecting phage is virulent or temperate, a specific life cycle is pursued (Fig. 1.3) (Gill & Hyman, 2010). Virulent phages always pursue lytic life cycle, which begins with sequential expression of phage proteins to take over host cell, which usually follows by DNA replication and synthesis of structural components. Complete virions are assembled and host cells are lysed to liberate progeny phages to surroundings (Echols, 1972). Contrary to the lifecycle of virulent phages, temperate phages can pursue lytic or lysogenic life cycle. The lytic cycle is pursued in a similar fashion as described for virulent phages, whereas in lysogenic cycle, phage genome is integrated into host chromosome, in most cases and phage-host genome replicate simultaneously. During lysogenic state, genes responsible for synthesis of structural proteins and host lysis are repressed. Under certain conditions, lysogenic state can be induced to lytic cycle and phage progeny are released into surroundings upon host lysis (Echols, 1972).

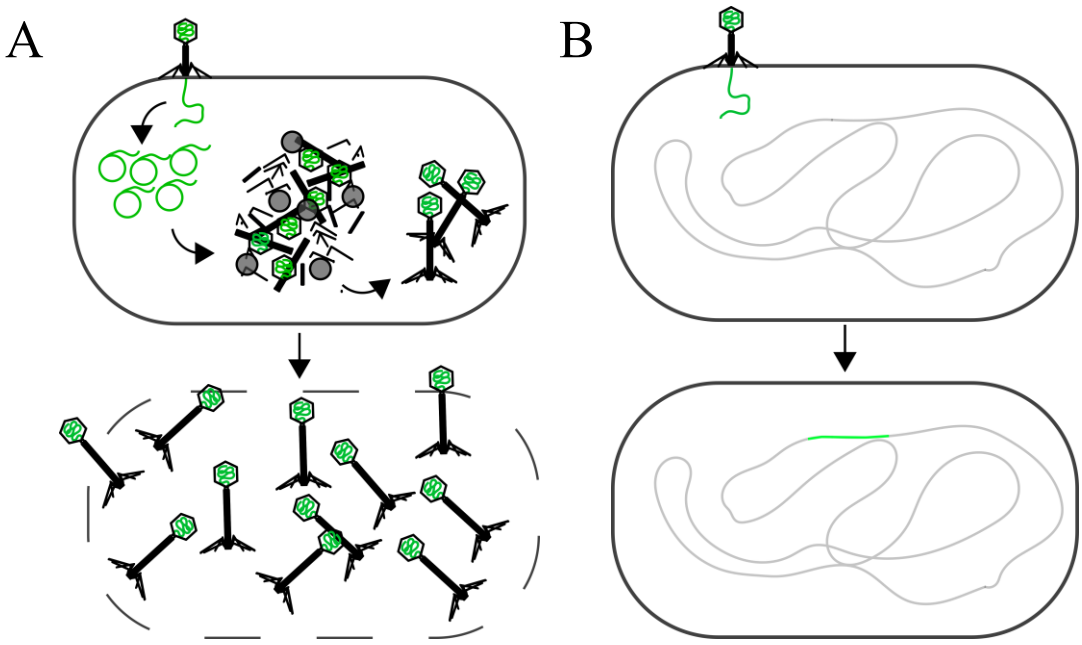


Fig. 1.3. Lytic vs lysogenic life cycle. Phage infection cycle starts with adsorption of a phage to its host receptors, followed by ejection of its DNA into host cytoplasm. Depending upon whether infecting phages are virulent or temperate, phages can pursue either lytic or lysogenic life cycle. **Panel A**: In lytic life cycle, phage DNA replicates and simultaneously various components necessary for virion morphogenesis are synthesized. Phage DNA is packaged into procapsids by DNA packaging complex. The phage capsids and tails are attached forming mature virions. At a genetically determined time, the host cell lyses releasing progeny virions into the surroundings. **Panel B:** In lysogenic life cycle, phage DNA is mostly integrated into host chromosome. Phage DNA and host DNA are replicated simultaneously. Virulent phages can pursue only lytic life cycle, whereas temperate phages can pursue either lytic or lysogenic life cycle. Temperate phages can also switch to lytic life cycle from lysogenic state.

Phage P1

Phage P1 was discovered as a resident prophage in *E. coli* strain "Li" , along with two other phages, P2 and P3, by G. Bertani in 1951 (Bertani, 1951). These three phages, P1,

P2, and P3, could be differentiated by plaque morphology as well as serum inactivation assay. The plaques formed by P1 were small, with size of less than 1 mm in diameter (Bertani, 1951). Along with other *E. coli* phages such as λ, T4 and T7, P1 has played a crucial role in the development of molecular biology and the study of phage infection cycle. T4 and T7 have a virulent lifecycle whereas both λ and P1 have temperate lifecycles (Lobocka *et al.*, 2004). Unlike phage λ that is integrated into host chromosome during lysogeny, P1 is maintained as an extra-chromosomal plasmid (Ikeda & Tomizawa, 1968). P1 has been used as a workhorse of gene transduction as it can non-specifically package host chromosome (Lennox, 1955; Lobocka *et al.*, 2004).

Virion morphology

P1 virions have been analyzed extensively by negative-stained electron microscopy. P1 is a member of *Myoviridae* family and has an icosahedral head, and a long contractile tail that terminates in six tail fibers (Walker & Anderson, 1970; Lobocka *et al.*, 2004). It has been reported that the lysates of P1 consist of virions of three different head sizes: P1B with ~86 nm, P1S with ~65 nm, and P1M with ~45 nm side-to-side diameter. The ratio of sizes of P1B:P1M:P1S is 4:3:2, which corresponds to *T* values of 16, 9 and 4 for P1B, P1M and P1S respectively (Walker & Anderson, 1970). Only the virions of head-size class P1B can package the full-length genome, hence are infectious. On the contrary, virions of both P1S and P1M cannot package full-length genome and thus, do not have a complete set of genetic elements required to establish a successful infection (Walker & Anderson, 1970).

The host specificity of P1 is determined by a 4.2 kb invertible C segment which involves a similar mechanism as in phage Mu (Iida, 1984). The C segment consists of tail fiber genes *Sv-U* or *Sv'-U'*, encoding the variable region of tail fibers, which is fused to the constant region encoded by *Sc* (Guidolin *et al.*, 1989a). Cin, a site-specific recombinase, is encoded by a gene adjacent to the C segment and mediates recombination between the 0.6 kb inverted repeats flanking the C segment, which form the *cix* (for *C in*version *cross*-over) sites (Iida *et al.*, 1982). P1 virions induced from lysogens have both C(+) and C(-) orientations, whereas P1 virions produced from lytic infection have only C(+) orientation (Iida *et al.*, 1982). P1 with C(+) orientation can infect both *E. coli* K12 and *E. coli* C, whereas P1 with C(-) orientation only infects an *E. coli* mutant (Iida *et al.*, 1982).

DNA packaging

The infective virions of P1 contain of ~94 kb cyclically permuted, linear, double-stranded DNA with a terminal redundancy of 10-15 kb (Sternberg, 1990; Lobocka *et al.*, 2004). P1 genome consists of 117 predicted genes, organized into 45 operons, of which 112 are translated into proteins and five code for untranslated RNA's (Lobocka *et al.*, 2004). The genomic DNA is produced as linear concatemers and packaged into P1 procapsids by a headful mechanism by the P1-encoded pacase proteins, PacA and PacB (Skorupski *et al.*, 1994a). P1 DNA contains a consensus *pac* sequence, which is recognized and cleaved by P1 pacase proteins, followed by packaging of the cleaved DNA into P1 procapsids. After the first headful packaging, the pacase proteins cleave

DNA in a non-sequence-dependent manner and initiates packaging of the remaining linear concatemeric DNA substrate into a new procapsid. After DNA packaging, P1 tails and DNA-filled capsids are attached to form complete P1 virions (Sternberg, 1990; Skorupski *et al.*, 1994b).

Cyclization of P1 DNA following infection

After recognizing a specific bacterial host, P1 ejects its DNA into host cytoplasm. Once in the cytoplasm, unlike other phage DNA, P1 DNA exists as a self-replicating, extra-chromosomal, circular form. P1 encoded recombinase protein Cre (for *c*auses *re*combination) and *loxP* [for *lo*cus of crossing over (*x*), *P*1] sites, at which recombination occurs, are necessary for circularization of P1 DNA following infection (Sternberg & Hamilton, 1981; Hochman *et al.*, 1983). The conversion of the linear DNA in the P1 capsid to the covalently closed circular form in host cytoplasm, after infection, occurs by an intramolecular recombination event that leads to the removal of the redundant portion of the DNA (Segev *et al.*, 1980). This recombination event results in the genetic map of P1 being linear, despite P1 DNA being cyclically permuted, in contrast to the circular genetic maps of other phages such as P22 or T4 (Sternberg & Hamilton, 1981). Because of the linearity of P1 genome, P1 genes on the ends of the genome are not linked together (Sternberg & Hamilton, 1981). The Cre-*loxP* system also mediates the segregation of P1 prophage into daughter host cells during lysogenic growth by resolving a P1 dimer composed of two P1 monomers (Sternberg *et al.*, 1986). After it was discovered that this Cre-*loxP* system can function in yeast *Saccharomyces*

cerevisiae, it has been used extensively in other eukaryotes for genome editing (Sauer, 1987; Lambert *et al.*, 2007). P1 DNA is also circularized by the host RecBCD nuclease complex. There is a total of 50 recombinational host spot Chi sites in positive and negative strands of P1 DNA. RecBCD complex functions as a 3'-to-5' nuclease, but when this complex encounters Chi sites in P1 genome, the enzymatic activity of the complex is switched to recombination-promoting form and thus helps in homologous recombination mediated circularization of P1 DNA (Zabrovitz *et al.*, 1977; Lobocka *et al.*, 2004).

Lysis-lysogeny decision

After circularization of DNA, P1, being a temperate phage, can pursue either lytic or lysogenic infection cycle. This lytic-lysogenic decision is controlled by genes in P1 immunity circuitry, which is distributed in three regions of P1 genome: ImmC, ImmI and ImmT (Fig. 1.4) (Yarmolinsky, 2004). Among the temperate phages studied, the immunity region of P1 and P7 is the most complex one (Heinrich *et al.*, 1995). The ImmC region consists of the *c*1 (master repressor) and *coi* (C1 inhibitor). The ImmI region include genes that code for translational repressor RNA, C4, and antirepressor proteins, Ant1and Ant2. The ImmT region has a gene, *lxc* that codes for a corepressor (Fig. 1.4) (Yarmolinsky, 2004). C1 and Coi are antagonistic to each other (Yarmolinsky, 2004). C1 repressor binds to ~17 operators throughout P1 genome and negatively

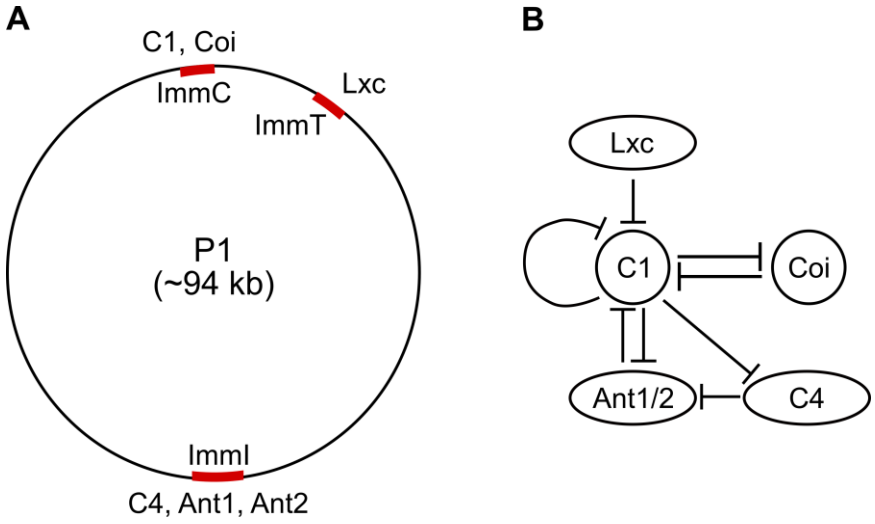


Fig. 1.4. Lysis-lysogeny decision in P1. Panel A: The decision to pursue lytic or lysogenic lifecycle in P1 is determined by the immunity circuitry, located in three regions of temperate phage P1 genome (represented by circle): ImmC, ImmT and ImmI. ImmC encodes C1 (master repressor) and Coi, ImmT encodes Lxc and ImmI encodes C4 (RNA), Ant1 and Ant2. **Panel B:** The lytic or lysogenic lifecycle is decided by the interplay between different factors. If C1 prevails, lysogeny is established, whereas if Coi prevails, lytic replication is established (Heinrich *et al.*, 1995), (Yarmolinsky, 2004), (Lobocka *et al.*, 2004).

regulates the expression of genes required for lytic pathway, including Coi, whereas Coi inactivates C1 by direct interaction. The RNA, C4 blocks the ribosome binding site in the transcript of the antirepressor proteins, thereby reducing the cellular concentration of the antirepressor proteins. These antirepressor proteins inactivate the C1 repressor (Heinrich *et al.*, 1995; Yarmolinsky, 2004). Lxc forms a complex with C1, and enhances the binding affinity of C1 to its operators, consequently resulting in the lowered expression of C1 itself (Yarmolinsky, 2004). Moreover, the ability of Coi to inhibit C1 repressor activity is inhibited in the presence of Lxc. The decision to pursue lytic or lysogenic lifecycle, upon infection of host by P1, is determined by the competition between C1 and Coi proteins (Heinrich *et al.*, 1995). If C1 synthesis is maintained, lysogeny growth is established, whereas if Coi synthesis is maintained, lytic growth is established (Lobocka *et al.*, 2004). In our study, we have used P1 with temperature sensitive C1 repressor. At non-permissive temperature, C1 is degraded and lytic infection cycle is induced (Lobocka *et al.*, 2004).

Host lysis

Phages of Gram-negative bacterial hosts are released into the surroundings after the disruption of the cellular membrane and the peptidoglycan layer. Usually, host lysis is accomplished by actions of three proteins. The holins create “holes” in the inner membrane releasing endolysins to the periplasmic space where the peptidoglycan layer is degraded by enzymatic activity of the endolysins (Young, 2013). In the final step, spanin complexes fuse the inner and the outer membrane resulting in the host lysis

(Rajaure *et al.*, 2015). The host lysis steps mediated by P1 is unique in that the export of endolysins to the periplasmic space is not dependent on holins (Xu *et al.*, 2004). P1 endolysin, Lyz, has a N-terminal SAR (for *s*ignal-*a*nchor and *r*elease) sequence and is exported and tethered in an inactive form to the inner membrane by the *sec* translocon. The SAR sequence is also sufficient for the release of Lyz from the inner membrane into the periplasmic space, without any proteolytic cleavage of the sequence, and this release occurs because of the collapse of the membrane potential(Xu *et al.*, 2004).

Host defense against phages

It has been estimated that ~10^{23} phage infections occur every second (Hatfull & Hendrix, 2011), which provide enormous selection pressure on bacterial host. Since bacteria need to survive constant predation by phages, bacteria have evolved a wide arsenal of defense systems to thwart the phage infection cycle (Fig. 1.5). In turn, phages are constantly evolving with antidefense systems to overcome these bacterial defense systems. This constant co-evolution of phages and their bacterial hosts have led to an evolutionary arms race, resulting in significant diversity in defense systems of bacteria and phages against each other (Van Valen, 1973; Labrie *et al.*, 2010; Stern & Sorek, 2011).

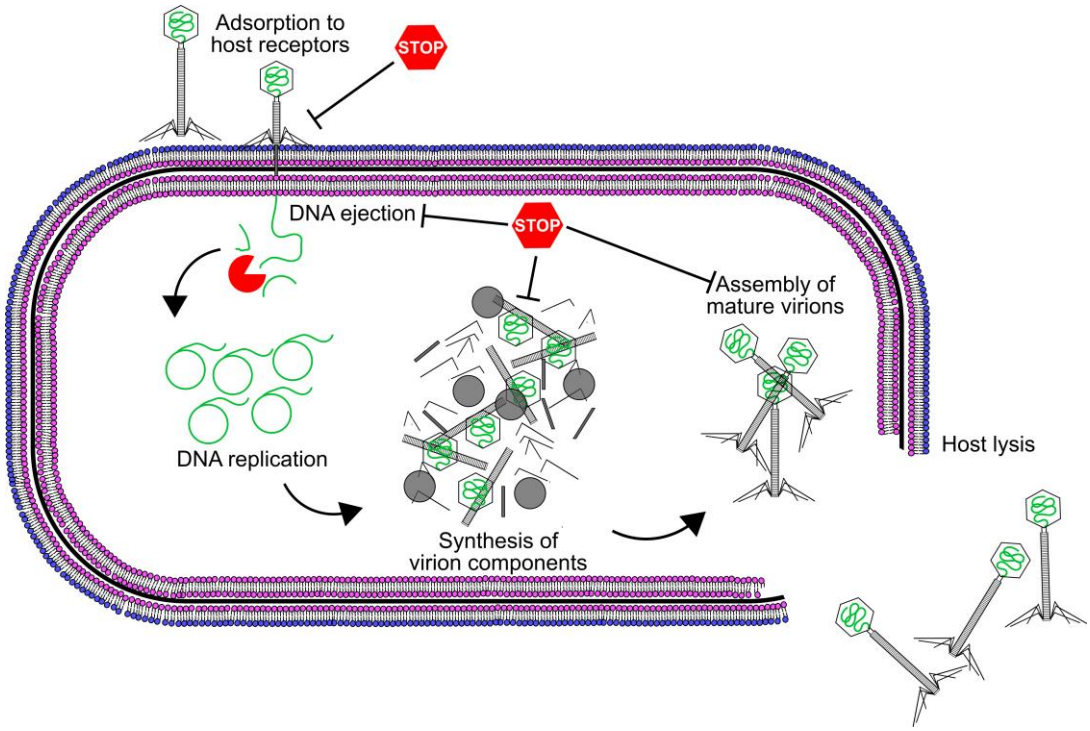


Fig. 1.5. Host defense against different stages of phage infection cycle. Bacteria have defense mechanisms to thwart phage infection at various stages. Bacteria can prevent adsorption of phages to the receptors by masking receptors with extracellular materials or by modifying receptors themselves. Even after successful adsorption, bacteria can prevent the ejection of phage DNA into cytoplasm by obstructing the DNA entry point. In the cytoplasm, phage DNA can be cleaved by CRISPR-Cas systems or Restriction-Modification systems. Moreover, bacteria possess abortive infection or assembly interference mechanisms that eventually prevent the formation of mature virions (Murray, 2000; Labrie *et al.*, 2010; Seed, 2015). The image is adapted from Seed, 2015.

Changes in receptors

The first step of phage infection cycle is the recognition and adsorption to the specific receptors on the surface of host cell by the receptor binding proteins of the phages. The receptors are present on the outer surface of the host cells and may be comprised of proteins, polysaccharides and lipopolysaccharides (Samson *et al.*, 2013). Bacteria have different strategies to prevent the adsorption of phages. The common theme of these strategies is to mask the specific receptor component. In *Staphylococcus aureus*, increased secretion of protein A has been shown to mask specific phage receptors (Nordstrom & Forsgren, 1974). Similarly, the K1 polysaccharide capsule of *E. coli* has been shown to prevent T7 infection (Scholl *et al.*, 2005). In another instance, in *Bordetella* spp., phase variation, under the control of the BvgAS two-component regulatory system, determines the expression of receptors and hence, phage infection (Liu *et al.*, 2002). Similar phase variation has also been reported in *Vibrio cholerae* O1 serogroup strains, in which the expression of O1 antigen determines if the cells are susceptible to phage infection or not (Seed *et al.*, 2012). Likewise, in other cases, extracellular components such as alginates, hyaluronan and glycoconjugates act as a barrier to prevent phages from accessing their specific receptors (Labrie *et al.*, 2010).

Phages have evolved with strategies to overcome this barrier provided by bacterial cells which masks receptors. By generating mutations in the host specificity determining gene *mtd*, *Bordetella* phage BPP-1 can infect host expressing different receptors (Liu *et al.*, 2002). Besides, *Pseudomonas* phage F116 and AF has been shown to degrade the

exopolysaccharide matrix surrounding the host, which might be necessary to reach the cell surface and initiate infection (Hanlon *et al.*, 2001; Cornelissen *et al.*, 2012). Moreover, several instances have also been reported where phages have acquired the ability to recognize these extracellular barriers as host receptors (Gross *et al.*, 1977; Steinbacher *et al.*, 1997).

Blocking DNA entry into cytoplasm

Only after recognition of the cell surface receptors, phage genetic material is ejected into host. This stage of phage infection cycle can be prevented by superinfection exclusion (Sie) systems. The Sie systems are encoded by phage genomes; thus, these systems specifically define the phage defense systems to prevent infection by closely related phages. The Sie proteins are localized in the membranes and prevent the phage genetic material from being delivered to the cytoplasm (Labrie *et al.*, 2010). The well-studied coliphage T4 has two Sie proteins, Imm and Sp. Imm prevents the transfer of T4 DNA into the cytoplasm by interfering with the injection site (Vallee & Cornett, 1972; Lu & Henning, 1989). T4 tail-associated protein gp5 assists in DNA ejection process with its lysozymic activity by degrading peptidoglycan layer. The other Sie protein, Sp inhibits the activity of T4 gp5, thereby blocking DNA ejection into cytoplasm (Kao & McClain, 1980). Other Sie systems are also described in coliphage P1 (Kliem & Dreiseikelmann, 1989), HK97 (Cumby *et al.*, 2012) and *Salmonella enterica* phage P22 (Susskind *et al.*, 1971; Susskind *et al.*, 1974a; Susskind *et al.*, 1974b).

The Sie system of P1 is encoded by *sim* gene in *immC* region of P1 genome (Kliem & Dreiseikelmann, 1989). Sim is synthesized as a ~25 kDa protein, which is proteolytically processed at its N-terminus to yield a ~24 kDa protein in a SecA-dependent manner. Sim is predicted to localize either in the membrane or the periplasmic space and prevents P1 superinfection (Maillou & Dreiseikelmann, 1990).

Degradation of phage DNA

Once successfully introduced to the host cytoplasm, the phage genetic material still must overcome host defenses such as CRISPR (for *C*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats)- Cas (for *C*RISPR-*as*sociated) and R-M (for *R*estriction-*M*odification) systems mediated cleavage of the genetic material (Labrie *et al.*, 2010).

CRISPR-Cas

CRISPR-Cas systems are the nucleic-acid based adaptive immunity systems that provide protection against foreign DNA including phage DNA, in which DNA fragments of an infecting phage are incorporated into CRISPR arrays and subsequent infection from the same phage is prevented (Barrangou *et al.*, 2007). CRISPR arrays consist of several partially palindromic repeats (CRISPR repeats, 23-55 nt) separated by variable (CRISPR spacers, 21-72 nt). These CRISPR arrays are often adjacent to *cas* genes (Barrangou *et al.*, 2007; Barrangou & Marraffini, 2014). In general, protection provided by CRISPR-Cas systems involve three distinct steps. In the first step, the spacer sequences are acquired from the genome of the phages that have infected the host. In the next step, the

CRISPR arrays are transcribed and processed into CRISPR RNAs, which form a complex with Cas proteins. Eventually, upon infection by the same phage, the complementary base-pairing of CRISPR RNAs and phage genome results in the degradation of the target (Stern & Sorek, 2011; Barrangou & Marraffini, 2014; Leon *et al.*, 2018). CRISPR-Cas systems are categorized into two major classes based on the organization of effector proteins, that mediate CRISPR RNA processing, and target recognition and cleavage (Koonin *et al.*, 2017).

Phage have diverse mechanisms to prevent CRISPR-Cas mediated DNA degradation. Any mutations in the spacer targeted (protospacer) region or the PAM (for *p*rotospacer-*a*djacent *m*otif) will render CRISPR-Cas ineffective (Deveau *et al.*, 2008; Pawluk *et al.*, 2018). Moreover, recent studies have discovered several anti-CRISPR proteins encoded in phage genomes, that inhibit CRISPR-Cas systems (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2016a; Pawluk *et al.*, 2016b; Rauch *et al.*, 2017).

Restriction-Modification systems

The R-M systems are the innate host defense systems which specialize in cleaving the foreign DNA based on its modification status. The effects of R-M systems were initially observed as the host specificity imparted to phages after they were propagated on different hosts (Bertani & Weigle, 1953). It was later demonstrated that the factors determining host specificity were physically linked to phage DNA (Arber & Dussoix, 1962) and degradation of phage DNA resulted in the host specificity of phages (Dussoix

& Arber, 1962). The R-M systems are broadly categorized into three types, type I, II and III, based on subunit composition, cofactor requirements, recognition site and cleavage position (Murray, 2000; Tock & Dryden, 2005). This study focuses on the P1 components that are required to protect phage DNA from the restriction function of type I R-M system.

The genes encoding components of type I R-M system were first identified in 1970 (Hubacek & Glover, 1970). The type I R-M systems are hetero-oligomeric complexes, encoded by three genes *hsdS, hsdM* and *hsdR* (*hsd* for *h*ost *s*pecificity *d*eterminant) (Fig. 1.6) (Murray, 2000). Unlike type I R-M systems, type II and III are composed of two subunits only. The main characteristics of these two systems and their differences with type I R-M systems are highlighted in (Table 1.1). HsdS (S) is the specificity subunit that enables any type I R-M complex to recognize specific base sequences. HsdM (M) is the modification subunit that adds methyl group in a sequence-dependent manner and HsdR (R) is the restriction subunit that cleaves the DNA under certain conditions (Tock & Dryden, 2005). The three subunits of type I R-M system can form two forms of complexes; modification complex and restriction complex. Both complexes recognize an asymmetric, bipartite sequence. The modification complex is composed of M and S subunits in the stoichiometric ratio of 2:1 (M_2S_1), whereas the restriction complex is composed of all R, M and S subunits in the stoichiometric ratio of 2:2:1 ($R_2M_2S_1$). The modification complex can only modify the DNA in a sequence-dependent manner, whereas the restriction complex can either modify or cleave DNA depending upon the

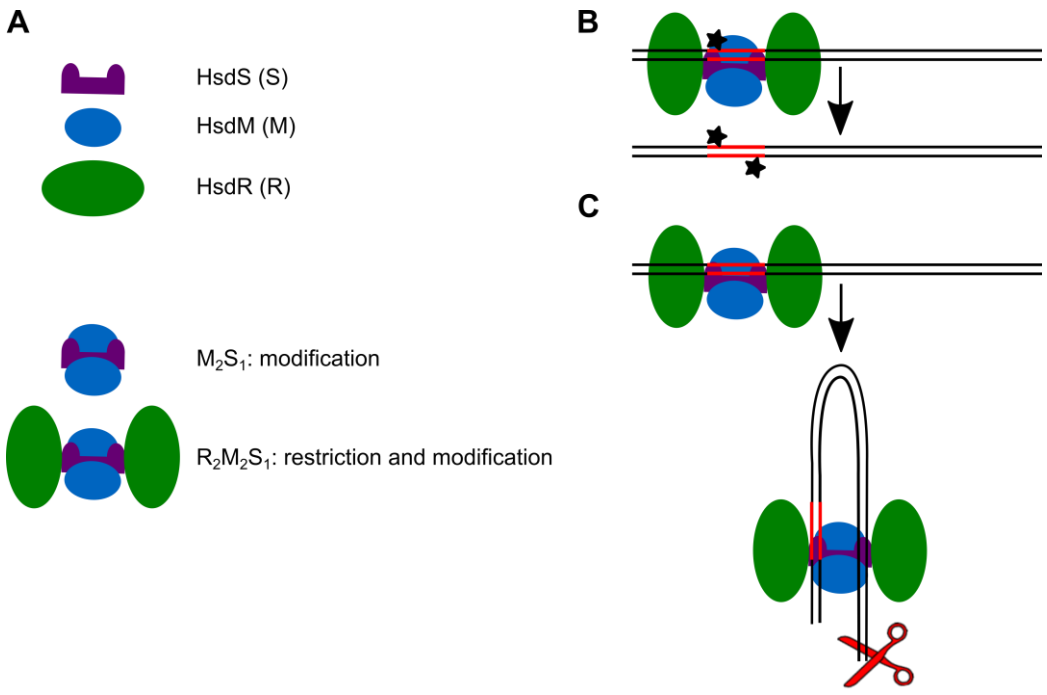


Fig. 1.6. Fate of foreign DNA depends on the methylation status of the recognition sequence. Panel A: The type I R-M system is a complex composed of three subunits: HsdS (S), HsdM (M) and HsdR (R). The S subunit provides sequence specificity to the complex, the M subunit has the active site to methylate specific bases within the recognition site, whereas the R subunit has the active site to cleave the DNA. The three subunits can form two stoichiometric complexes: M_2S_1 and $R_2M_2S_1$. M_2S_1 is the modification subunit that can only methylate DNA, whereas $R_2M_2S_1$ can either methylate or cleave the DNA depending upon its methylation status. **Panels B and C:** The foreign DNA can be hemi-methylated or unmethylated (methylation status denoted by ★). **(B)** If the DNA is hemi-methylated , then the $R_2M_2S_1$ complex, methylates the other strand. **(C)** If the DNA is unmethylated, the $R_2M_2S_1$ complex translocates the DNA and cleaves the DNA a few kb away from the recognition site .

Table 1.1. Characteristics of different R-M systems. The characteristics of different R-M systems are highlighted (Kruger & Bickle, 1983; Tock & Dryden, 2005). The methylation activity of the R-M systems described in the table occurs with the recognition sequence, whereas the site for DNA restriction varies according to the type of R-M systems. Adenine is methylated at N6 position, whereas cytosine is methylated at either N4 or C5 position (Tock & Dryden, 2005).

R-M Systems	Subunits	Subunit arrangement	Co-factors	DNA recognition sequence	DNA cleavage site	DNA translocation
type I	HsdS (S), HsdM (M), HsdR (R)	Modification - M_2S_1	AdoMet	Asymmetric, bipartite	NA	No
		Restriction - $R_2M_2S_1$	AdoMet Mg^{2+}, ATP		~1-5 kb away from recognition site	Yes
type II	Restriction (R), Modification (M)	Modification - M	AdoMet	mostly palindrome	NA	No
		Restriction - R_2	Mg^{2+}		mostly at the recognition site	No
type III	Mod (M), Res (R)	Modification -M_2	AdoMet	asymmetric	NA	No
		Restriction- R_2M_2	Mg^{2+}, ATP		~25-27 bp away from recognition	Yes

methylation status of the recognition sequence. If the DNA is hemi-methylated, the restriction complex methylates a specific base in the complementary strand, and only if the DNA is unmodified or inappropriately modified in both strands, the restriction complex cleaves the DNA (Fig. 1.6) (Murray, 2000). The type I R-M complex requires AdoMet, Mg^{2+} and ATP as cofactors. HsdM provides binding interface for AdoMet, whereas HsdR provides binding interface for Mg^{2+} and ATP. AdoMet acts as the methyl donor and is the only cofactor required for the modification activity. However, all of AdoMet, Mg^{2+}, and ATP are required for restriction activity (Table 1.1) (Meselson & Yuan, 1968; Tock & Dryden, 2005).

Restriction of foreign DNA by type I R-M systems leading to specific activity involve a series of steps. In general, the enzyme complex gets activated by binding to AdoMet and the activated complex binds to DNA. Upon binding to ATP, the complex can recognize methylation status in recognition site. If DNA is methylated, the complex is released from DNA. If DNA is hemi-methylated, a specific base in recognition site in the complementary strand in methylated. If DNA is unmethylated, DNA is translocated and cleaved several kb away from the recognition site, accompanied by ATP hydrolysis (Fig. 1.6) (Yuan *et al.*, 1975; Burckhardt *et al.*, 1981).

Bacterial host may contain any specific type I R-M system. Depending upon the restriction background of the host, the DNA of progeny phage will have sequence specific modification (Fig. 1.7). On the contrary, if the bacterial host does not have any functional type I R-M system, the DNA of progeny phages will not have any sequence specific modification. If the sequence specificity of the type I R-M system of the new host is same, then the modified phage DNA, upon ejection into host cytoplasm, is not cleaved, and the phage can establish infection efficiently. In another case, if the phage DNA is unmethylated at the recognition sequence of the type I R-M system of the new host, phage DNA is cleaved and phages have lower plating efficiency (Fig. 1.7).

Phages have evolved with diverse mechanisms to overcome the type I R-M system mediated DNA cleavage. The mechanisms of how paradigm phages such as T3, T4, T7, Mu, λ, and P1 protect their genome from host type I R-M system mediated DNA cleavage have been well studied (Table 1.2) (Labrie *et al.*, 2010).

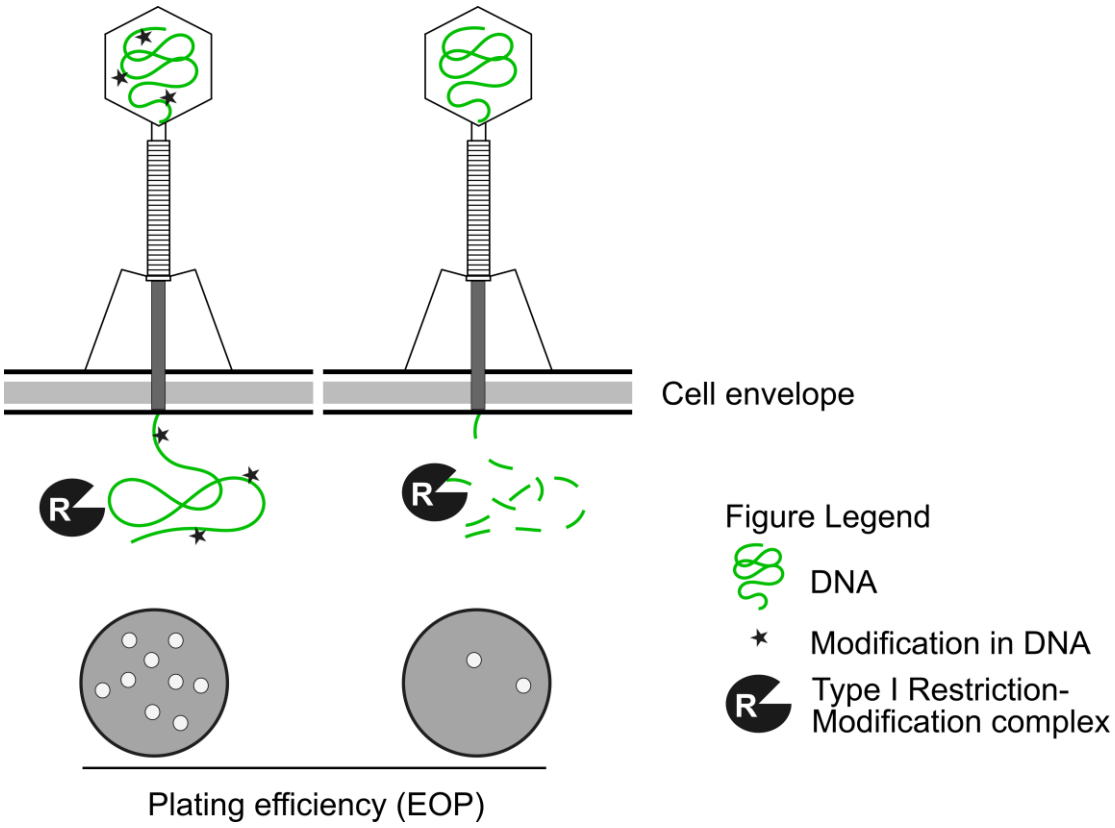


Fig. 1.7. Unmethylated phage DNA is cleaved by host type I R-M system. Depending upon the restriction background of host, the phage propagated in the host could have modification in a particular recognition sequence (modification denoted by ★). Upon infection of a new host, phage DNA is encountered by the type I R-M system of the new host. If the recognition sequence of this type I R-M system in the phage DNA is modified, the DNA is not cleaved and the phage can plate efficiently. In other case, if phage DNA is unmodified at the recognition sequence, then it is cleaved and the phage cannot plate efficiently.

Table 1.2. Antirestriction mechanism of phages. Phages have diverse mechanisms of overcoming the type I R-M system of their bacterial host. Some mechanisms of well-studied phages have been mentioned in the table. See text for details.

Phages	Proteins	Mechanism of protection	References
T3 and T7	Ocr	Dimer of Ocr mimics the phosphate backbone of B-form DNA and binds to the type I R-M complex, inactivating its function. T3 ocr also hydrolyzes SAM, a co-factor for type I R-M system.	(Studier & Movva, 1976; Kruger & Bickle, 1983; Atanasiu *et al.*, 2002)
T-even	N/A	Hypermodification of DNA: The cytosine of genome is replaced by hydroxymethyl cytosine, which is again glucosylated.	(Volkin, 1954; Kornberg *et al.*, 1961; Kruger & Bickle, 1983)
Mu	Mom	Acetimidation of adenine in genome.	(Kruger & Bickle, 1983)
λ	Ral	Enhances methylation function of type I R-M complex.	(Zabeau *et al.*, 1980; Kruger & Bickle, 1983)
P1	DarB	Proposed to methylate DNA upon ejection into host cytoplasm.	(Iida *et al.*, 1987; Piya *et al.*, 2017)

Phages T3 and T7 protect DNA from type I R-M system by expressing protein gp0.3. The protein gp0.3, also called Ocr (Ocr for *o*vercome *c*lassical *r*estriction), is one of the early proteins expressed following T3 and T7 infection (Studier, 1973). The dimer form of Ocr mimics the size, shape and charge of bent B-form DNA and inhibits the type I R-M complex by directly binding to the DNA binding region of the complex (Walkinshaw *et al.*, 2002). The binding affinity of the type I R-M complex to Ocr is stronger than that for DNA (Samson *et al.*, 2013). Because of this direct interaction between Ocr and type I R-M complex, the type I R-M complex cannot bind to DNA, and hence it is unable to degrade foreign DNA (Tock & Dryden, 2005). T3 Ocr protein also inhibits the type I R-M system by hydrolyzing AdoMet, which is a co-factor required for the biological activity of type I R-M system (Studier & Movva, 1976; Spoerel *et al.*, 1979). In the absence of AdoMet, the type I R-M system cannot be activated to degrade foreign DNA (Tock & Dryden, 2005).

The ability of the T-even phages, T2, T4 and T6, to evade restriction provides an example of how phage-host interactions has shaped evolution. These phages incorporate an unusual base hydroxymethylcytosine (HMC) in its genomic DNA (Volkin, 1954; Lehman & Pratt, 1960; Kornberg *et al.*, 1961; Kruger & Bickle, 1983). Because of the presence of this unusual base, DNA of these phages are not susceptible to cleavage from

restriction complexes that have cytosine in the recognition sequence. To overcome this feature of T-even phages, host cells have acquired the modification-dependent systems (MDSs) such as McrA, McrBC and Mrr that can cleave hydroxymethylated DNA (Raleigh & Wilson, 1986). Interestingly, DNA of T-even phages are also glucosylated at HMC, hence these DNAs are also protected against these MDSs (Volkin, 1954; Lehman & Pratt, 1960; Kornberg *et al.*, 1961; Kruger & Bickle, 1983). However, a prophage encoded two-component system consisting of proteins, GmrS and GmrD (Gmr for *g*lucose-*m*odified *r*estriction), which can degrade glucosylated HMC, have been found in *E. coli* strain CT596. To protect DNA from GmrS-D restriction systems, T-even phages inject hundreds of virion-associated internal protein I (IPI), which inactivate the nuclease activity of the GmrS-D restriction systems (Abremski & Black, 1979; Rifat *et al.*, 2008; Labrie *et al.*, 2010).

A peculiar epigenetic modification is present in phage Mu DNA, which make its DNA insensitive to restriction from type I R-M complex because of the function of phage gene *mom* (*mom* for *m*odification *o*f *M*u). Owing the gene function of *mom*, ~15% of adenine residues in Mu DNA is modified by acetimidation (Kruger & Bickle, 1983).

The genome of phage λ contains the gene *ral* (for *r*estriction *al*leviation) that contributes to protecting DNA from type I R-M systems (Zabeau *et al.*, 1980). The antirestriction protein Ral enhances the modification activity of type I R-M complex, thereby lowering the number of the unmodified recognition sites in λ genome (Zabeau *et al.*, 1980;

Loenen & Murray, 1986; Samson *et al.*, 2013). Because of this, cleavage of λ DNA by the type I R-M complex is alleviated.

Following infection of its *E. coli* host, phage P1 injects Dar (for *d*efense *a*gainst *r*estriction) proteins to protect its DNA from host type I R-M system (Iida *et al.*, 1987). Iida *et al*. discovered that two P1 loci provided restriction protection against a subset type I R-M system and hence, were named accordingly. DarA protected P1 DNA from *Eco*A-mediated restrictioon, and DarB protected P1 DNA from *Eco*B- and *Eco*K-mediated DNA cleavage. Dar proteins are virion-associated and act only *in cis*. However, they can protect any DNA that is packaged into P1 capsids. Packaging of DarB into virions was known to be dependent on the presence of DarA because *darA*$^{-}$ phages were phenotypically *darB*$^{-}$ (Iida *et al.*, 1987). DarA is synthesized as a high-molecular-weight precursor and is proteolytically processed at the N-terminus. Only the processed form of DarA is packaged into P1 virions (Streiff *et al.*, 1987). The genes encoding these Dar proteins were found in two distinct regions of P1 genome (Lobocka *et al.*, 2004). The genes near *darA* and including *darA* itself, were also shown to affect virion morphogenesis, but the phenotype was not distinctly assigned to any gene (Iida *et al.*, 1998). All three of these type I R-M systems have specific recognition site and different number of those recognition sites in P1 genome (Table 1.3).

Table 1.3. Recognition sequence of type I R-M systems. Three type I R-M systems used in our study has their own sequence specificity. The recognition sequence for any type I R-M system is bipartite. There are two groups of specific sequences: One in the 5' end and the other in the 3' end. These two groups are separated by non-specific bases. The red face A (5' group) is methylated by the type I R-M complex in the positive strand, whereas the complementary base A in the to the bold face T (3' group) is methylated in the negative strand (Loenen *et al.*, 2014). There are different number of recognition sites in the P1 genome for these type I R-M systems.

Type I R-M System	Recognition sequence (5'-3')	Number of sites in P1 genome
*Eco*A	G**A**G(N7)G**T**CA	2
*Eco*B	TG**A****(**N8)**T**GCT	20
*Eco*K	A**A**C(N6)G**T**GC	11

Abortive infection

Even when phages survive the R-M and CRISPR-Cas systems, other bacterial defense systems such as abortive infection and assembly interference can still prevent phage infection cycle (Chopin *et al.*, 2005; Seed, 2015). Compared to other bacterial defense systems, abortive infection results in the death of host cell, thus this defense system is also referred to as altruistic death of host cells to prevent phage multiplication so that the surrounding bacterial population can be protected from phage predation (Fineran *et al.*, 2009). The phage λ encoded Rex system is the best characterized abortive infection system. Upon activation, the Rex system depolarizes the membrane resulting in lowered cellular ATP concentration, which ultimately stops all cellular processes and thus prevents phage multiplication (Parma *et al.*, 1992). Several other abortive systems have also been reported to directly target different stages of the phage infection cycle such as DNA replication (Emond *et al.*, 1997), phage transcripts (Parreira *et al.*, 1996), and inducing early lysis (Durmaz & Klaenhammer, 2007).

Assembly interference

The assembly of certain phage particles can also be prevented by the action of phage-inducible chromosomal islands (PICIs) such as *Staphylococcus aureus* pathogenicity islands (SaPIs) (Seed, 2015). Unlike above described abortive infection, this type of assembly interference still allows cellular processes to proceed, resulting in the formation of phage-like, but non-infectious particles (Tallent *et al.*, 2007; Fillol-Salom *et al.*, 2018). These SaPIs reside dormant in bacterial chromosomes, until they are induced

by the infection of specific helper phages and contribute to lateral gene transfer (Lindsay *et al.*, 1998). Upon induction of PICIs, replication of target phage is limited by the elements of PICIs, which redirect phage replication to PICI replication and use phage encoded structural proteins to package PICI-specific DNA, resulting in the formation of non-infectious, phage-like particles (Ram *et al.*, 2012). The PICIs have a well-conserved genetic organizations and are specially characterized by absence of structural and lysis genes (Fillol-Salom *et al.*, 2018).

A similar PICI-like element (PLE) has been reported in *Vibrio cholera* as well. It has been shown that phage ICPI overcomes *Vibrio* PLE by encoding a CRISPR-Cas system in the phage genome, which targets the PLE genome (Seed *et al.*, 2013).

Assembly of capsid-associated proteins

The Dar antirestriction proteins of P1 and IP1 antirestriction proteins of T4 are comparable in that both of these proteins are assembled into the mature virions and are injected into the host cells during the infection cycle to protect phage DNA from restriction (Iida *et al.*, 1987; Rifat *et al.*, 2008). Contrary to scarce information available on P1 head morphogenesis, there is wealth of information available on T4 head morphogenesis. There are several instances where phages are found to deliver proteins to host cells, in addition to the genetic material. The well-studied *Salmonella enterica* phage P22 also injects proteins following infection of the host cells (Jin *et al.*, 2015). The capsids of P22 contain three internal proteins (also called “pilot”, “ejection” or “E”

proteins"), which are essential for infection, and have been shown to be injected into the host cells (Jin *et al.*, 2015). Both IPs of T4 and E proteins of P22 are assembled in the virion capsids during initial stages of capsid morphogenesis (Thomas & Prevelige, 1991; Leiman *et al.*, 2003).

The assembly pathway of prokaryotic viruses shares a significant similarity. The capsid morphogenesis in most bacteriophages and eukaryotic viruses such as herpesviruses and adenoviruses start with the formation of a protein shell that is later packed with the genetic material. The steps involved in capsid morphogenesis of different phages T4, P22, HK97 are different, but in general this process requires portal, major capsid and scaffolding proteins. The portal proteins occupy one of the twelve vertices of icosahedral phage capsids and form a dodecameric ring (Aksyuk & Rossmann, 2011; Veesler & Johnson, 2012). The portal proteins act as a nucleation factor and recruits major capsid and scaffolding proteins to form a protein procapsid shell of correct size and shape. The scaffolding protein core, which could be made of one or multiple proteins, determines the size and shape of the capsid for a phage. The scaffolding proteins are also required for the incorporation of internal proteins such as IPs of T4 and E proteins of P22 in the procapsid shell (Veesler & Johnson, 2012; Mateu, 2013). Since the morphogenesis pathway of P1 has not been studied in a great detail, there is no information on the role of portal and scaffolding proteins in capsid size determination and incorporation of other capsid proteins. However, by genetic and biochemical characterization, we have shown

that the virion-associated antirestriction components of P1 are assembled in the capsid following a definite pathway (Chapter II).

Questions to be addressed

Interest in P1 antirestriction system, DarB in particular, stems from the results of bioinformatic analysis that indicates that P1 DarB-like proteins are conserved (Gill *et al.*, 2011). Most of these DarB-like proteins are associated with mobile DNA elements such as plasmids, conjugative transposons, insertion sequences, integrative conjugative elements or genomic islands. This suggests that these DarB-like proteins might protect these DNA elements when they are mobilizing across hosts with different restriction background. In some cases, these DNA elements are associated with virulence factors, antibiotic resistance or phenotypic conversion, which raises public health concern, thus the genetic and biochemical study of DarB is of great significance (Gill *et al.*, 2011).

Using P1 and its *E. coli* host with different type I R-M systems, the mechanism of how P1 antirestriction system protects its genome from type I R-M system mediated DNA cleavage will be studied. As described above, phage P1 ejects Dar proteins into its host to protect its DNA from host type I R-M systems. The original study reported by Iida *et al.* in 1987 implies DarA and DarB as the constituent proteins of P1 antirestriction system. In addition, in 1998, *Iida et al.* reported that some genes in the *darA* operon affected capsid morphogenesis. However, the phage mutants used in those studies were generated using classic genetic tools, in which multiple genes could have been affected,

thus ascribing a phenotype to a particular gene was ambiguous. Throughout this study a λ Red-mediated genetic recombineering approach was used to construct isogenic gene deletions in P1 prophage and plasmid-based complementation system to verify the phenotypes. In Chapter II, several genes are interrogated to determine if components other than the *darA* and *darB* operons are important in the antirestriction system. It has been demonstrated that *hdf, ddrA* and *ddrB* from the *darA* operon and *ulx* from the *darB* operon form this multicomponent antirestriction system. Besides, it has been shown that these genes encode virion-associated proteins and these proteins are incorporated in the mature virions following a definite pathway. Moreover, the roles of *hdf* and *darA* in head-size determination of P1 virions have also been established. Building up from our findings described in Chapter II, P1 antirestriction system has been further characterized in Chapter III. Chapter III has three parts: In the first part, P1 genes of unknown functions have been tested for their role in protecting P1 DNA from type I R-M systems. In the second part, it has been determined that the antirestriction proteins are incorporated in the virions before packaging of DNA. More importantly, a few N-terminal residues of the antirestriction protein DarB were found to be significant in directing the incorporation of DarB into P1 virions. In the third part, expression and purification of DarB have been optimized to obtain DarB for testing the biochemical mechanism of antirestriction function.

Understanding phage-host relationship is crucial for several reasons. Food and biotechnology industries use phage resistant bacterial strains to manufacture several

products. Thus, the emerging phage population, which could propagate on the selected bacterial strains, have to be constantly evaluated to ensure continuous product development (Labrie *et al.*, 2010). Moreover, due to the emergence of antibiotic resistant bacteria, phage therapy, the use of phages to treat bacterial pathogens, is receiving renewed interest. Classical phage biology has relied on isolating novel phages from environmental sources for any bacterial strains. With the advances in systems and synthetic biology, the concept of synthesizing phages with a broader host range and host-independence have been proposed. As co-evolutionary arms race between phages and bacteria result in spectacular host-phage interactions, it is crucial to understand these relationships to successfully engineer any phages. Since bacteria can readily become resistant to phages by mutating the receptors, it is important to characterize the receptors so that phages targeting different receptors can selected for therapeutic purposes. Because of the adaptability of phage genomes, most of the sequenced phage genes are of unknown function. Since the genetic material of phages is ejected into the host cells in each infection cycle, efforts should be made to understand the essentiality of these genes. More importantly, phages also use internal host factors such as chaperones, transcription factors during infection cycle. These interactions must be understood especially in context of therapeutic phages so that robust phages that do not rely completely on host factors can be engineered and selection of phage resistant bacteria can be limited (Labrie *et al.*, 2010; Young & Gill, 2015).

Chapter IV discusses application of high-throughput genetic screen to study host-phage interactions of novel *E. coli* phages, LL5 and LL12, belonging to TLS- and rV5-like group, respectively. Receptors for these two phages have been characterized and it has been discovered that two chaperones, PpiB and SecB, are required for efficient infection cycle of phage LL5. In Chapter V, future directions of these research projects have been discussed.

CHAPTER II

THE MULTICOMPONENT ANTIRESTRICTION SYSTEM OF PHAGE P1 IS LINKED TO CAPSID MORPHOGENESIS[1]

Introduction

Bacteriophage P1 was discovered by G. Bertani in 1951 as a temperate phage residing in Escherichia coli strain "Li" (Bertani, 1951). P1 is a myophage of the order Caudovirales, with a 94 kilobase (kb) unit genome of linear dsDNA (Lobocka et al., 2004); unlike most other known temperate phages, P1 lysogenizes *E. coli* as a circular plasmid maintained at one copy per host chromosome (Ikeda & Tomizawa, 1968; Rosner, 1972). P1 has also played a major role in molecular genetics as the premier generalized transducing phage of *E. coli* (Lennox, 1955; Calendar, 1988).

In *Caudovirales* phages such as P1, the infection cycle begins with the adsorption of the phage to the host surface and the ejection of phage DNA into the host cell. Bacteria have several mechanisms to defend themselves against phage infection, including restriction and modification (R-M) systems that recognize and cleave foreign DNA in a site-specific manner (Fig. 2.1A). These systems are broadly divided into three categories, type I, II and III, which are distinguished by their subunit makeup, DNA cleavage

[1] Reprinted with permission from "The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis" by Piya D., Vara L., Russell W.K., Young R. and Gill J.J., 2017. *Mol Microbiol.* 2017 August; 105(3): 399–412. Copyright by John Wiley and Sons, Inc.

mechanisms and the nature of their DNA recognition sequences (Kruger & Bickle, 1983; Loenen *et al.*, 2014).

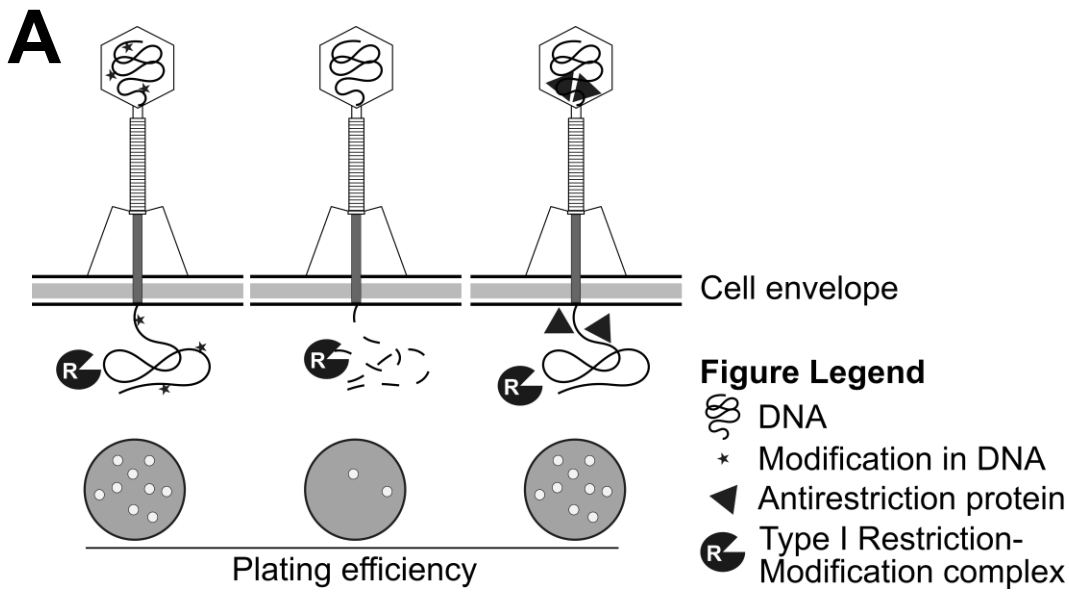


Fig. 2.1. The antirestriction system of phage P1 protects DNA from type I restriction and modification (R-M) mediated cleavage. Reprinted with permission from Piya *et al*., 2017. Panel A. Phage infection begins with the translocation of the linear 95 kb P1 genome from the phage head into the host cytoplasm. Depending on the bacterial host where phages are propagated, specific bases in the phage DNA are methylated (denoted as stars). If the methylation pattern in the DNA matches the specificity of the host type I R-M system, the DNA is protected from cleavage (left), resulting in normal plating efficiency. If the phage DNA is unmethylated (or improperly methylated), the DNA is cleaved by the host type I R-M system (middle), resulting in reduced plating efficiency. Iida *et al.* (1987) demonstrated that in P1, virion-associated proteins (triangles) can protect P1 DNA from type I restriction, even if it is unmodified (right). It is believed that at least some of these virion-associated proteins must translocate into the cell in order to function.

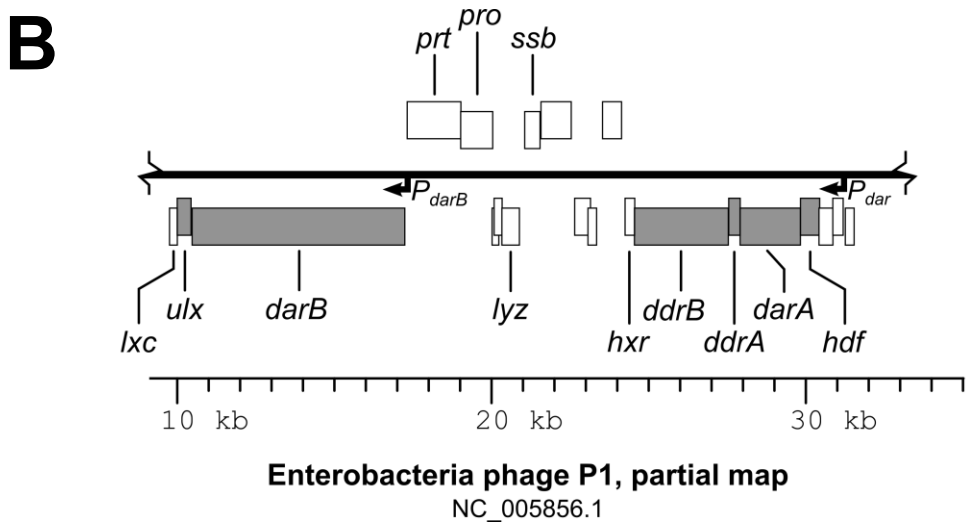


Fig. 2.1 Continued. Panel B. Partial genomic map of P1 shows genes associated with the antirestriction system. Rectangular blocks represent genes, and those above the black line are transcribed rightwards, and genes below the line are transcribed leftwards. Besides the previously described *darA* and *darB*, several other genes are associated with the P1 antirestriction system. The dar genes (grey) are organized in two separate predicted transcriptional units. One transcript contains *darB* and *ulx* (*darB* operon, driven by the P_{darB} promoter) and the other contains *hdf*, *darA*, *ddrA* and *ddrB* (*darA* operon, driven by P_{dar}). The ruler at the bottom of the figure corresponds to the nucleotide position as it appears in the published P1 genome (NC_005856.1).

Type I R-M systems are composed of three subunits, HsdR (R), HsdM (M) and HsdS (S) (Hsd for host specificity of DNA). These subunits can form two different oligomeric complexes; $R_2M_2S_1$, which can catalyze both restriction and modification, and M_2S_1, which can only catalyze modification (Suri *et al.*, 1984; Murray, 2000). The HsdS subunit recognizes specific, bipartite DNA sequences of 13-15 bp, the HsdM subunit recognizes the modification status in the recognition DNA sequences and the HsdR subunit catalyzes DNA cleavage at a non-specific location up to a few kb away from the recognition sequence, if DNA is unmethylated or is methylated inappropriately (Murray, 2000). Unlike type I R-M systems, type II R-M systems have independent restriction and modification enzymes. These restriction enzymes cleave DNA basepairs within the recognition sites if the DNA is not appropriately modified (Bickle & Kruger, 1993). Similar to type II R-M systems, type III systems are comprised of separate modification and restriction subunits. The modification subunit can function by itself as a modification methylase. However, a complex of modification and restriction subunits is required for endonuclease function, in which DNA cleavage occurs 25-27 bp away from the recognition site (Tock & Dryden, 2005).

P1 encodes the virion-associated proteins DarA and DarB (*dar* standing for *d*efense *a*gainst *r*estriction), which are required to protect the P1 genomic DNA (gDNA) from restriction by host type I R-M systems (Fig. 2.1A) (Iida *et al.*, 1987). DarA, with a predicted molecular mass of 69 kDa, is expressed late in the lytic cycle and is processed into a smaller 60 kDa form when incorporated into the virion (Streiff *et al.*, 1987). DarB

has a predicted molecular mass of 251 kDa and appears to be incorporated into the virion intact (Iida *et al.*, 1987). Iida *et al*. demonstrated by plaque assay that DarA is required for protection of DNA against restriction by the *Eco*A type I R-M system and DarB is required for protection against the *Eco*B and *Eco*K systems. The efficiency of plating (EOP) of P1 *darA* mutants was reduced in *Eco*A, *Eco*B and *Eco*K strains whereas the EOP of P1 *darB* mutants was reduced only in *Eco*B and *Eco*K strains. Because complementation studies indicated that the Dar proteins function only *in cis*, it has been proposed that these are ejected into the host cell along with phage DNA at the initiation of infection, where they exert their antirestriction activity (Fig. 2.1A) (Iida *et al.*, 1987). The antirestriction proteins synthesized during the phage lytic cycle only act in the following infection cycle, and the antirestriction activity is not specific to P1 gDNA, but any DNA that is packaged into the P1 virion (Iida *et al.*, 1987). Since *darA*$^{-}$ phages also exhibit the *darB*$^{-}$ phenotype and *darA*$^{-}$ virions lack both DarA and DarB proteins, it appears that DarA is required for the incorporation of DarB into the P1 capsid (Iida *et al.*, 1987). The 2,255-residue DarB contains an identifiable N-terminal methyltransferase domain and a central DExD-like helicase domain, enzymatic functions that could be imagined to play a direct role in the protection of DNA from restriction (Lobocka *et al.*, 2004). Bioinformatic analysis of DarB revealed that it is widely distributed among bacteria, primarily associated with mobile DNA elements including phages, plasmids and conjugative transposons (Gill *et al.*, 2011). It has been proposed that DarB could represent a larger class of proteins that facilitate DNA mobility by protecting mobile

DNA elements from type I R-M system-mediated cleavage during their transfer between hosts.

To further explore the antirestriction system of phage P1, we used a recombineering approach to create isogenic deletion mutants of *darA, darB,* and other P1 genes with previously unknown function including *hdf*, *ddrA*, *ulx* and *ddrB*. The results are discussed in terms of a network of proteins involved in the assembly of P1 antirestriction system and their effects on capsid morphogenesis.

Materials and methods

Bacterial strains and phages

The bacterial strains and the parent phage P1CM*clr*100 used in this study were obtained from the Coli Genetic Stock Center, Yale University, and are listed in Table 2.1. P1CM is a derivative of P1*kc* that acquired chloramphenicol resistance from the R-factor R_{14} (Kondo & Mitsuhashi, 1964). P1CM*clr*100, hereafter referred to simply as P1, is a thermoinducible mutant of P1CM (Rosner, 1972). All phage mutants used in this study are single-gene deletions of P1. Unless otherwise noted, *E. coli* strains were cultured on LB broth [10 g L^{-1} Bacto Tryptone (BD), 5 g L^{-1} Bacto yeast extract (BD), 10 g L^{-1} NaCl (Avantor)] or LB agar (LB broth amended with 15 g L^{-1} Bacto agar) at 37 °C. P1 lysogens were cultured and maintained at 30 °C on LB amended with 10 µg mL^{-1} chloramphenicol (LB cm) or 10 µg mL^{-1} chloramphenicol plus 30 µg mL^{-1} kanamycin (LB cm+kan).

Table 2.1. Bacterial strains, phages and plasmids. Reprinted with permission from Piya *et al.*, 2017.

Strains, phages or plasmids	Genotype or relevant characteristic	Reference/source
***E. coli* strains**		
WA2379	*leu*$^-$ *met*$^-$ *lac*$^-$ r_A^- m_A^-	(Arber & Wauters-Willems, 1970) / The Coli Genetic Stock Center (CGSC), Yale University
W3110	F^- λ^- *rpoS*(*Am*) *rph-1 Inv*(*rrnD-rrnE*) r_K^+ m_K^+	(Iida *et al.*, 1987; Hayashi *et al.*, 2006) / CGSC
WA921	*thr*$^-$ *leu*$^-$ *met*$^-$ *lac*$^-$ r_K^- m_K^-	(Wood, 1966; Arber & Wauters-Willems, 1970) / CGSC
WA960	thr_B^+ *leu*$^-$ *met*$^-$ *lac*$^-$ r_B^+ m_B^+	(Wood, 1966; Arber & Wauters-Willems, 1970) / CGSC
BW25113(pKD46)	F^- l^- *rpoS*(*Am*) *rph-1 rrnB3* Δ*lacZ4787 hsdR514* Δ(*araBAD*)*567* Δ(*rhaBAD*)*568*	(Baba *et al.*, 2006) / CGSC
BW25141(pKD4)	*lacI*^{q} $rrnB_{T14}$ Δ$lacZ_{WJ16}$ Δ*phoBR580 hsdR514* Δ$araBAD_{AH33}$ Δ$rhaBAD_{LD78}$ *galU95* Δ$endA_{BT333}$ *uidA*(Δ*MluI*)::*pir*$^+$ *recA1*	(Datsenko & Wanner, 2000) / CGSC
MG1655	*lacI*^{q1}*tonA*::Tn*10*	(Park *et al.*, 2006) / Lab stock
Phages		
P1Δ*darB*	in-frame deletion of *darB* in P1CM*clr*100	This study
P1Δ*ulx*	in-frame deletion of *ulx* in P1CM*clr*100	This study
P1Δ*darA*	in-frame deletion of *darA* in P1CM*clr*100	This study
P1Δ*hdf*	in-frame deletion of *hdf* in P1CM*clr*100	This study
P1Δ*ddrA*	in-frame deletion of *ddrA* in P1CM*clr*100	This study
P1Δ*ddrB*	in-frame deletion of *ddrB* in P1CM*clr*100	This study
P1Δ*hxr*	in-frame deletion of *hxr* in P1CM*clr*100	This study
Plasmids		
p*darB*	P1 *darB* cloned into pBAD24	This study
p*ulx*	P1 *ulx* cloned into pBAD24	This study
p*darA*	P1 *darA* cloned into pBAD24	This study
phdf_{am}_*darA*_$ddrA_{am}$	P1 *hdf*_*darA*_*ddrA*gene fragment cloned into pBAD24 and amber mutations introduced to 9th codon in *hdf* and 28th codon in *ddrA*	This study
p*hdf*	P1 *hdf* cloned into pBAD24	This study
p*ddrA*	P1 *ddrA* cloned into pBAD24	This study
p*ddrB*	P1 *ddrB* cloned into pBAD24	This study
p*ulx-CFLAG*	P1 *ulx* with C-terminal FLAG-tag fusion cloned into pBAD24	This study

Production of phage lysates

Phage lysates were produced by thermal induction of P1 lysogens. Lysogenic strains were grown at 30 °C in LB cm to OD_{550} 0.5 - 0.6. P1 was thermally induced by shifting the culture to 42 °C in a shaking water bath for 60-75 min (Iida & Arber, 1977). The crude lysate was harvested when the OD_{550} fell to ~0.2, by centrifugation of the culture at 10,000 x g, 10 min, 4°C and sterilized by passage through a 0.22 µm syringe filter (Millipore).

Lysogenization

Lysogens of P1 and its mutants were produced as previously described (Rosner, 1972) with minor modifications. A fresh overnight culture of the desired *E. coli* lysogenization host was supplemented with 5mM $CaCl_2$, and 100 µL of the culture was mixed with 100 µL of an undiluted phage lysate (~10^8-10^9 PFU mL^{-1}) and incubated at RT for ~20 minutes. The phage-host mixture was then plated to LB cm and a CM^R colony was selected after overnight incubation at 30 °C and purified by restreaking. The same procedure was used for producing lysogens of P1 deletion mutants, with the exception that plating was conducted on LB cm+kan.

Generation of PCR fragments for single-gene deletions

All primers (table 2.2) used in PCR were ordered from Integrated DNA Technologies. The plasmid pKD4 has an *FRT*-flanked *kan* gene which was used at the source of the kanamycin resistance cassette used for gene deletions (Datsenko & Wanner, 2000). To

Table 2.2. Primers. Reprinted with permission from Piya *et al*., 2017.

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
tttgtcaaAaccgacctgtccg	ggacaggtcggtTttgacaaaaag	To mutagenize *Eco*P1 site in kanamycin resistance gene in pKD4. The nucleotides in capital letters represent the mutation that is introduced.
ggttttatggacagcaagcg	gcttccatccgagtacgtg	Sequencing primers to verify site-directed mutation in pKD4
tggcgaactcaaacgtcgtctta agcaactgaaagcaggaaattaa catggtgtaggctggagctgctt c	gaggtccattaactcaatagct gatagtgtcatgctgctaccc cgctttatgggaattagccatg gtcc	To knockout *darA* from P1 genome
gtctttcaatcaacagcagatca c	acaccatcatgttccgaagg	Sequencing primers to verify *darA* deletion in P1 genome
tcgtgtttgaacggaatttaaca ctagtcacttgttaaggattacc aatggtgtaggctggagctgctt c	tgcgacgcccggcgaaccgggc gctcctgttatgcgtattgttg gatgacatgggaattagccatg gtcc	To knockout *darB* from P1 genome
aggaggatgttgtcccgttc	tttcagtaatcgcccgcgtag	Sequencing primers to verify *darB* deletion in P1 genome
gcagttacgtaataaatctcgca caggatgtgtcagatgacgaaaa ataagtgtaggctggagctgctt c	atccctttattgatattgaact gttccatgttaatttcctgctt tcagttatgggaattagccatg gtcc	To knockout *hdf* from P1 genome
tatatgggtaaaggagagcgact g	tccgtagcacaccagtggatc	Sequencing primers to verify *hdf* deletion in P1 genome
acctggctgatctgctggtagca atccagaaagcgggggtagcagc atgagtgtaggctggagctgctt c	tgatcgcttaagctcatccgtg ttacgccttatgctgcctcttt aatgttatgggaattagccatg gtcc	To knockout *ddrA* from P1 genome
ggatatgcaaagcactgacatgg	ccgtccagagtattggatacc	Sequencing primers to verify *ddrA* deletion in P1 genome
tgctggaactgttattaggtgtg gcaaaggagctaaataatgcaga ttaagtgtaggctggagctgctt c	gaacaggttatttttgcactat caggtagttaatatactatctg gtctacatgggaattagccatg gtcc	To knockout *hxr* from P1 genome
tatacccgccagcctcagtaag	ctgcggcggttataggttcc	Sequencing primers to verify *hxr* deletion in P1 genome
cccggttcgccgggcgtcgcata atatggccacactatctgataca ataagtgtaggctggagctgctt c	cccatgctttactgtgtagtat cgcttttttcaagttttctctcc agcctgatgggaattagccatg gtcc	To knockout *ulx* from P1 genome

Table 2.2. Primers. Continued

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
cgaaagacgcaatcaagacg	tgatctgctcccattcttcg	Sequencing primers to verify *ulx* deletion in P1 genome
catggatagccgccaacattaaagaggcagcataaggcgtaacacggatggtgtaggctggagctgcttc	atctccgcctaatggcgctgcgattttaatctgcattatttagctcctttatgggaattagccatggtcc	To knockout *ddrB* from P1 genome
agtggcgaggagaaacaacc	tcaagggtgtcaacgaatcc	Sequencing primers to verify *ddrB* deletion
atcgtctagagtaataaaggaggtatcgatatgaacaagctatctatgggggtg	atcgaagcttttatgcgtattgttggatgacggc	Clone *darB* into pBAD24
atcgatctagagtaaaggagatcgatcatgtgtcagatgacgaaaaataagtatgc	atcgaaagcttttatgctgcctctttaatgttggc	Clone *hdf*, *darA* and *ddrA* into pBAD24
atcgatctagagtaaaggagatcgatcatgtgtcagatgacgaaaaataagtatgc	acacaaagcttttaatttcctgctttcagttgcttaag	Clone *hdf* into pBAD24
atcgatctagagtaaaggagatcgatcatggaacagttcaatatcaataaaggg	atcgaaagcttagctgatagtgtcatgctgc	Clone *darA* into pBAD24
atcgatctagagtaaaggagatcgatcatgacactatcagctattgagttaatgg	atcgaaagcttttatgctgcctctttaatgttggc	Clone *ddrA* into pBAD24
atcgatctagagtaaaggagatcgatcatggccacactatctgatacaataaaacc	atcgaaagctttcaagttttctctccagcctgtg	Clone *ulx* into pBAD24
atcgatctagagtaaaggagatcgatcatgagcttaagcgatcaggtgg	atcgaaagcttttaatctgcattatttagctcctttgcc	Clone *ddrB* into pBAD24
gaaaaataagtaGgcaacggtcgattttg	tcgaccgttgcctacttatttttcg	To make amber mutation in the 9th codon of *hdf*
cgagtggctaggagttgctg	caactcctagccactcgcgg	To make amber mutation in the 28th codon of *ddrA*
atcgatctagagtaaaggagatcgatcatggccacactatctgatacaataaaacc	atcgaaagctttcacttgtcgtcatcgtctttgtagtcagaaccagttttctctccagcctgtgc	To clone *ulx* with C-terminal FLAG-tag fusion into pBAD24

avoid restriction by the Type III R-M system resident in the P1 chromosome, an *Eco*P1 site normally present in the *kan* gene was removed by following protocol from QuikChange site-directed mutagenesis kit (Agilent Technologies) changing G608 to A in a silent mutation. To generate FRT-flanked *kan* insert for gene deletion, the 3'-ends of the forward and reverse primers were designed to match the priming sites gtgtaggctggagctgcttc and atgggaattagccatggtcc of pKD4 (Datsenko & Wanner, 2000). The 5' ends of both forward and reverse primers were synthesized to include 50 nt homology to the P1 genome up- and downstream of the targeted gene. The flanking regions were chosen to retain the first and last several codons of the targeted genes to avoid polar effects (Datsenko & Wanner, 2000). PCR reactions were performed using Phusion Hi-Fidelity PCR Master Mix (New England Biolabs), following the manufacturer's recommended protocol. The PCR products were gel purified by using the QIAquick Gel Extraction Kit (Qiagen).

Generation of single-gene knockout mutants

The phage lambda Red recombinase mediated homologous recombination method was used to generate isogenic single-gene knock-out P1 mutants (Datsenko & Wanner, 2000). P1 was lysogenized into BW25113(pKD46) and a colony resistant to both chloramphenicol and ampicillin was selected. BW25113(pKD46) lysogenized with P1 was grown in LB amended with 100 μg mL^{-1} ampicillin and 10 μg mL^{-1} chloramphenicol to OD_{550} 0.1, 1 mM L-arabinose was added to induce Red proteins, and the culture was grown to OD_{550} 0.6. The cells were harvested and made

electrocompetent as described previously (Datsenko & Wanner, 2000). Each electroporation reaction contained 300 ng of gel-purified DNA and 100uL competent cells in a 0.1 cm cuvette and was transformed in a Bio-Rad MicroPulser^{TM} electroporator following the manufacturer's protocol. Cells were recovered in 1 mL SOC for 2 hours at 30 ^{o}C and plated to LB cm+kan at 30 °C overnight. CM^{R} and Kan^{R} colonies were selected and mutations were verified by PCR amplifying a DNA region spanning the *kan* insertions and sequencing across the insertion junctions (Baba *et al.*, 2006). These lysogens were then induced as above and the mutant P1 phages were lysogenized into *E. coli* strain MG1655 as described above for maintenance of the strains.

Complementation of phage mutants

Phage knockouts were complemented *in trans* by induction of lysogens containing both the mutant prophages and vectors expressing the deleted genes. Complementing genes were amplified by PCR from a P1 DNA template (Table 2.1) and cloned into pBAD24 at its *Xba*I and *Hind*III sites using standard molecular biology techniques (Guzman *et al.*, 1995). The primers used for PCR are listed in table 2.2. Ligation products were transformed into competent *E. coli* 5-alpha cells (New England Biolabs) and selected by plating on LB agar amended with 100 µg mL^{-1} ampicillin. The plasmids were extracted as described above and were verified by sequencing before transformation into the $r^{-}m^{-}$ strain WA921. Complemented phages were prepared by thermal induction of mutant phage from WA921 lysogens containing the corresponding complementing vector expressing the deleted gene *in trans* (Table 2.1). 1 mM L-arabinose (Sigma-Aldrich) was

added to the culture at the time of temperature shift to 42 °C to induce protein expression from the complementing plasmids. The phage lysate was used to determine Efficiency of Plating (EOP) on restricting and non-restricting hosts as described below.

Efficiency of Plating (EOP) Assay

All phages used in EOP assays were induced from modification-deficient WA921 lysogens. Assays were conducted as previously described with few modifications (Arber & Dussoix, 1962; Mise & Arber, 1976; Iida *et al.*, 1987). Host cells, grown to OD_{550} 0.4 - 0.5 in tryptone broth (10 g L^{-1} Bacto tryptone, 5 g L^{-1} NaCl) amended with appropriate antibiotics, were incubated with 10 mM $CaCl_2$ for 30 minutes at RT. Phage were adsorbed to 300 μL of host cells for 20 minutes at RT. The cells and phage were then plated using the soft agar overlay method using LB (Lennox) plates (10 g L^{-1} Bacto tryptone, 5 g L^{-1} NaCl, 5 g L^{-1} Bacto yeast extract, g L^{-1} Bacto agar) containing 2.5 mM $CaCl_2$ as the bottom plates with 4 mL lawns of tryptone top agar (10 g L^{-1} Bacto tryptone, 5 g L^{-1} NaCl, 7 g L^{-1} Bacto agar). Plaques were counted after overnight incubation at 42 °C. EOP was calculated as the ratio of plaques appearing on the lawn of the restricting strain to the number of plaques on WA921 lawns. The EOP of each phage mutant was normalized to the EOP of parental P1 on the same plating strain. In complementation assays, EOPs were normalized to P1 induced from WA921(pBAD24). All experiments were replicated three times.

Purification of virions by CsCl isopycnic centrifugation

P1 or P1 mutant was induced in 1L LB as described above. The crude lysate was centrifuged in JA10 rotor for 15 minutes at 17000 x g and the supernatant was filter sterilized. The lysate was concentrated by 24-hour centrifugation in JA10 rotor at 14000 x g at 4°C. The pellet was soaked in SM buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$) at 4°C and extracted after 48 hours. 10 µg mL^{-1} DNase (Sigma) was added to the concentrated phage and left at RT for 30 minutes. Phages were then purified by equilibrium centrifugation in Cesium chloride as previously described (Boulanger, 2009). The phage bands were extracted with 18-gauge needles and dialyzed against SM buffer in Slide-A-Lyzer 3500 MWCO dialysis cassettes (Thermo Scientific). Small volumes of pre-dialysis samples were saved and used to measure refractive index in an Abbe Refractometer. Refractive indices were converted to density (g cc^{-1}) by comparing to standard CsCl density-refractive index correlation table.

SDS-PAGE analysis

For SDS-PAGE analysis of phage proteins, samples were prepared as described previously, with slight modification (Boulanger, 2009). Approximately 2 x 10^{10} PFU of CsCl purified P1 were loaded per lane. For antirestriction P1 mutants, protein loading was normalized to equal amounts of tail sheath protein (57 kDa). Phages were heated in boiling water for 10 minutes to release DNA from the capsid and samples were then treated with DNase I at 37 °C for 2 hours. The samples were denatured by heating in boiling water for 5 minutes in Laemmli sample buffer and loaded on a 4-20% Tris-

glycine SDS-PAGE gel (Life Technologies) (Laemmli, 1970). PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific) was used as molecular mass standard. The gel was stained with SYPRO Ruby (Thermo Scientific), following the manufacturer's recommended protocol for maximum sensitivity. The gel was imaged with Fotodyne gel imager.

Proteomic analysis

Proteins associated with the virion were identified by mass spectrometry (MS) of the protein bands excised from SDS-PAGE gel as described previously (Gill *et al.*, 2011). Briefly, ~10^{11} PFU of phage P1 was prepared and loaded into 4-20% Tris-glycine SDS-PAGE as described above. Coomassie-stained bands were excised, subjected to reduction with dithiothreitol and alkylation with iodoacetamide. The samples were treated with ~0.4 µg trypsin (Thermo Scientific) and digested peptides were concentrated and desalted using C_{18} ZipTips (Millipore). The samples were spotted manually onto a matrix-assisted desorption ionization (MALDI) target (Genomic Solutions) using α-cyano 4-hydroxycinnamic acid. MALDI-MS and MS/MS analysis were conducted against the NCBI-nr database as described previously (Gill *et al.*, 2011). P1 DarB, DarA and Hdf were identified by this method.

The virion-associated DdrB was identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in collaboration with the Protein Chemistry Lab at Texas A&M University. Four lanes each of ~2 x 10^{10} PFU P1 and P1Δ*ddrB* were loaded into a

4-20% Tris-glycine SDS-PAGE and stained with SyproRuby as described above. The gel region missing protein band in P1Δ*ddrB* lanes and corresponding protein band from P1 lanes were excised. The samples were processed for LC-MS/MS as described before (Shevchenko *et al.*, 2006). LC-MS/MS was performed on a LTQ Orbitrap Velos/ETD mass spectrometer (Thermo Fisher) equipped with a NanoLC 2-D HPLC system (Eksigent). The results were analyzed using Mascot (Matrix Science) and X!Tandem (The GPM). Scaffold (Proteome Software) was used to validate MS/MS based peptide and protein identifications. The SDS-PAGE band was annotated based upon its presence in P1 and absence in P1Δ*ddrB* lanes.

Transmission Electron Microscopy

Phages were stained with 2% uranyl acetate and imaged in a JEOL 1200 EX transmission microscope under 100 kV accelerating voltage as previously described (Valentine *et al.*, 1968; Gill *et al.*, 2011). Side-to-side head diameters perpendicular to the axis of the tail were measured electronically using ImageJ (Schneider *et al.*, 2012) and converted to nm against images of a carbon grating replica of known dimensions (Ted Pella, cat# 607).

Results and discussion

Genomic context of the P1 antirestriction system

The genome of phage P1 was completed in 2004 (NC_005856.1), showing the positions of the *darA* and *darB* structural genes within the P1 chromosome (Lobocka *et al.*, 2004).

As shown in the partial P1 genetic map (Fig. 2.1B) the 1,920 bp *darA* gene is located in a polycistronic operon with its expression driven by LP*dar* (LP for late promoter) (Guidolin *et al.*, 1989b), and *darB* (6,768 bp) is located ~10 kb downstream of *darA* in a separate operon driven by P*darB* (Lobocka *et al.*, 2004). The *darA* gene is located immediately downstream of *hdf,* and upstream of *ddrA, ddrB* and *hxr*. The genes *lydA* and *lydB,* upstream of *hdf,* are known to function as the phage holin and antiholin, respectively (Schmidt *et al.*, 1996). Two operons are located between the *dar* operons: one containing the transcript of the phage's SAR endolysin *lyz* (Xu *et al.*, 2004) and another on the opposite strand encoding the predicted phage portal, portal protease and single-stranded DNA binding protein (Lobocka *et al.*, 2004). The *darB* gene is upstream of *ulx* and the lysogeny maintenance gene *lxc* (Lobocka *et al.*, 2004).

The genes shown in grey in Fig. 2.1B were deleted and replaced with *kan* markers by a lambda Red recombineering approach (Datsenko & Wanner, 2000). Initially *darA* and *darB* were deleted, followed by *hdf, ddrA, ddrB, hxr* and *ulx*. Because *lydAB, lyz, lxc* and the phage structural proteins Prt and Pro already had assigned functions, they were not considered to be relevant to this study and were not studied further. Deletion of *hxr* produced no detectable antirestriction phenotype against *Eco*A, *Eco*B or *Eco*K (data not shown) and genes downstream of *hxr* were not examined further.

darB and ulx provide protection against EcoB and EcoK

The antirestriction phenotypes associated with P1 *darA* and *darB* were first reported in 1987 (Iida *et al.*, 1987): P1, when propagated on a modification-deficient host, exhibited a slight (less than ten-fold) reduced efficiency of plating (EOP) on *Eco*A and *Eco*K lawns and an EOP of ~10^{-2} on *Eco*B. Under the same conditions, P1*darA*$^{-}$ plated at an EOP ~10^{-4} to 10^{-5} on all three restrictive hosts, whereas P1*darB*$^{-}$ was restricted (EOP ~ 10^{-4}) only on *Eco*B and *Eco*K.

The previously reported *darB*$^{-}$ phenotype, defined as severely reduced plating efficiency on *Eco*B and *Eco*K hosts, was recapitulated in an isogenic deletion of *darB* (Fig. 2.2A). As previously reported, P1Δ*darB* did not exhibit an antirestriction defect in the *Eco*A host. In addition to reproducing the *darB* phenotype, the antirestriction defect in P1Δ*darB* could be efficiently complemented *in trans* by inducing the mutant phage from a lysogen of the non-modifying host strain WA921 carrying the plasmid p*darB,* expressing DarB from an arabinose-inducible promoter (Fig. 2.2A). In this complementation system, the progeny phage could incorporate the DarB protein expressed from the plasmid in the propagating host and produce the antirestriction phenotype upon infection of a restricting strain. Phages complemented in this way could only express the antirestriction phenotype for their first infection cycle. Complemented mutant phage that were re-propagated on WA921 exhibited the same EOP as the original P1Δ*darB* when plated to all three of the restricting strains (data not shown), indicating the absence of recombination between the phage and complementing vector during

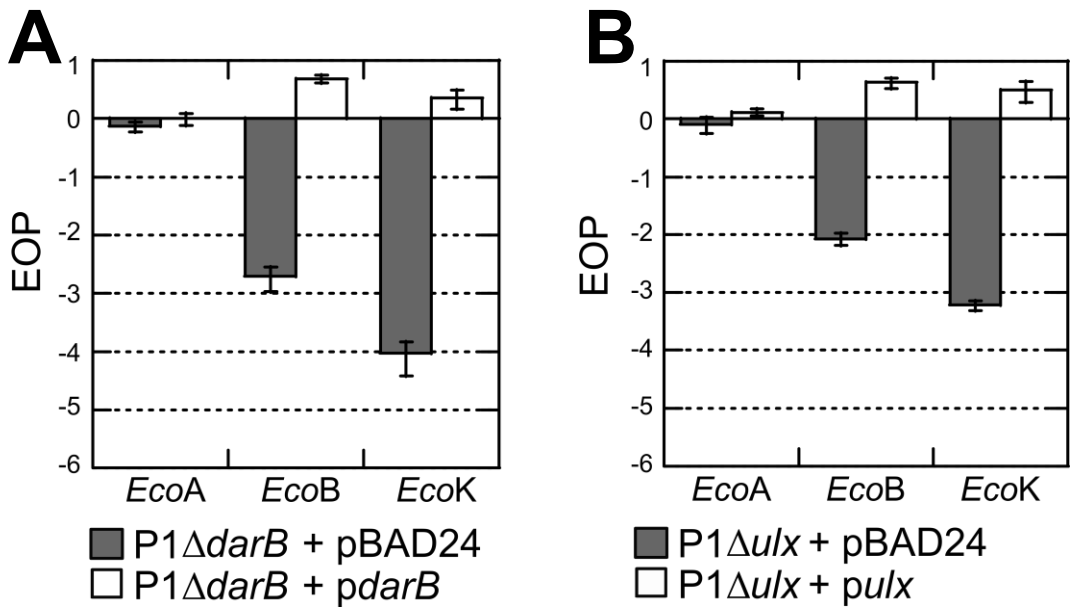


Fig. 2.2. *darB* and *ulx* are required for protection of P1 against *Eco*B and *Eco*K restriction. Reprinted with permission from Piya *et al*., 2017. Parental P1 or isogenic deletion mutant phages were thermally induced from lysogenized modification-deficient WA921 containing either empty pBAD24 or the respective complementing plasmid. Phages were then plated on *E. coli* hosts containing the Type I *Eco*A, *Eco*B, or *Eco*K R-M systems and WA921; efficiency of plating (EOP) for each phage was calculated as the ratio of plaques formed on lawns of the R-M-containing strains to the plaques formed on WA921. The EOP data shown have been normalized to the EOP of parental P1 induced from WA921 containing pBAD24. **Panel A:** the relative EOP of P1Δ*darB* was reduced by ~10^{-3} and 10^{-4} in *Eco*B and *Eco*K strains, respectively (gray bars). **Panel B:** the relative EOP of P1Δ*ulx* was reduced by ~10^{-2} and 10^{-3} in *Eco*B and *Eco*K strains respectively (gray bars). The restriction phenotype associated with both *darB* and *ulx* could be complemented *in trans* (white bars). No notable phenotype was observed in the *Eco*A strain. The data shown are averages of three biological replicates and error bars represent standard deviation.

phage propagation. The reduced EOP of P1Δ*darB* in *Eco*B and *Eco*K cells indicates that in the absence of DarB, P1 DNA, after being ejected into host cells, is susceptible to *Eco*B and *Eco*K restriction, but is protected from *Eco*A restriction.

Other genes near *darB* were examined for their possible role in antirestriction. The gene *ulx* (146 codons; *ulx* for *u*pstream of *lx*c (Lobocka *et al.*, 2004)) is located between *darB* and *lxc* and was previously of unknown function. P1Δ*ulx* showed a partial *darB*-like restriction phenotype, with EOP attenuated approximately 10-fold less than for P1Δ*darB* (Fig. 2.2B). The *ulx*-associated restriction phenotype could also be complemented *in trans* (Fig. 2.2B). In both *darB* and *ulx* deletions, complementation produced slightly higher EOP's than the parent P1 phage, suggesting complementation could provide a greater copy number of protein available for incorporation into the virion and subsequent greater protection against restriction. The plaque sizes for both P1Δ*darB* and P1Δ*ulx* were comparable to P1. Moreover, a double deletion mutant of *darB* and *ulx* showed plating deficiency similar to P1Δ*darB* (data not shown). Thus, both *darB* and *ulx* are required for full protection of P1 DNA against *Eco*B and *Eco*K restriction.

darA, hdf and ddrA contribute to protection against EcoA, EcoB and EcoK

In agreement with previous work, the isogenic deletion P1Δ*darA* had reduced EOP when plated on strains expressing *Eco*A, *Eco*B or *Eco*K (Fig. 2.3A). Relative to P1, the EOP of P1Δ*darA* was decreased by ~10^{-3}, 10^{-2} or 10^{-4} in cells expressing *Eco*A, *Eco*B or *Eco*K, respectively. The *Eco*A-associated *darA* restriction phenotype could be only

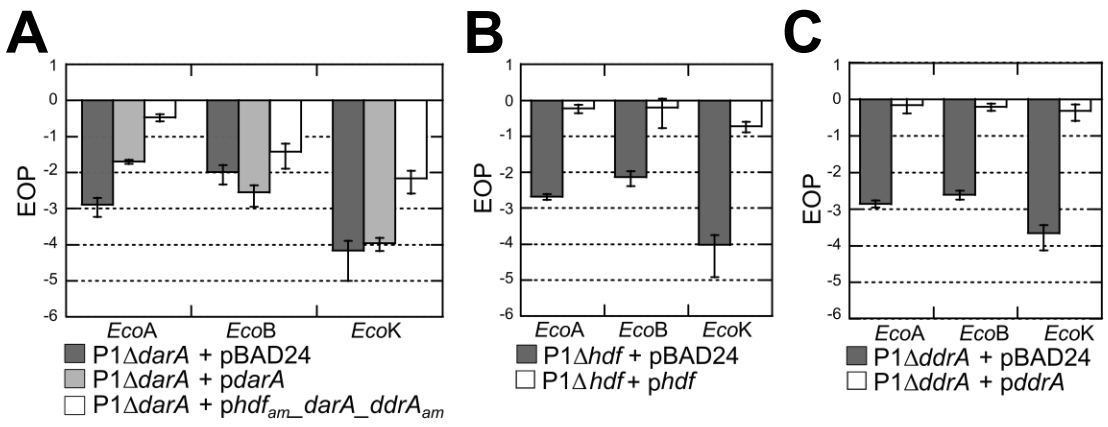


Fig. 2.3. *hdf*, *darA* and *ddrA* are required for protection of P1 against *Eco*A, *Eco*B and *Eco*K restriction. Reprinted with permission from Piya *et al*., 2017. Phages were induced and EOP assays were performed as described previously. The EOP data shown have been normalized to EOP of parental P1 induced from modification deficient WA921 containing pBAD24. **Panel A:** The relative EOP of P1Δ*darA* was reduced by ~10^{-3}, 10^{-2} and 10^{-4} in *Eco*A, *Eco*B and *Eco*K strains respectively (dark gray bars). The *darA*-associated restriction phenotype could be complemented partially by a plasmid bearing only *darA* (light gray); stronger complementation was observed with a plasmid expressing a longer transcript containing *darA*, along with the upstream gene hdf_{am}, and downstream $ddrA_{am}$ (white bars). **Panels B and C:** Both P1Δ*hdf* and P1Δ*ddrA* exhibited a *darA*-like restriction phenotype (gray bars). The restriction phenotypes of both *hdf* and *ddrA* could be complemented *in trans* (white bars). The data shown are averages of three biological repeats and the error bars represent standard deviation.

partially complemented *in trans* by *darA* alone, and the *Eco*B and *Eco*K phenotypes were not affected. It was previously reported that *darA* mutants could be complemented to nearly *wt* levels *in trans* by a P1 DNA fragment containing *darA*, along with the upstream gene *hdf* and the downstream gene *ddrA* of the *darA* operon (Iida *et al.*, 1998). In our experiments, P1Δ*darA* was more efficiently complemented *in trans* by a construct

containing *darA*, *hdf* and *ddrA* (data not shown). Better complementation in presence of *hdf* and *ddrA* raised a possibility of the polar effects of *darA* deletion, in which the restriction phenotype observed could either be associated with *hdf* or *ddrA*, the other genes present in the complementing plasmid. In order to rule out the roles of Hdf and DdrA in complementation, nonsense mutations were introduced in both *hdf* and *ddrA* to prevent their expression. Same degree of complementation was observed with phdf_{am}_*darA*_$ddrA_{am}$ (Fig. 2.3A). The low efficiency of complementation exhibited by the p*darA* vector may be due to instability of the mRNA transcript containing *darA* alone, as the longer DNA fragments included on phdf_{am}_*darA*_$ddrA_{am}$ and the previously reported P1 DNA fragment provided better complementation.

Previous studies of *darA* and *darB* function used P1 mutants that contained either *IS* insertions or deletions that often spanned *hdf* and *ddrA* (Iida *et al.*, 1987). To determine if these genes could also play a role in the protection of P1 DNA from host restriction, the isogenic mutants P1Δ*hdf* and P1Δ*ddrA* were created and evaluated for their antirestriction phenotypes. Both P1Δ*hdf* and P1Δ*ddrA* showed *darA*-like phenotypes, with sensitivity to *Eco*A, *Eco*B and *Eco*K restriction similar to that exhibited by P1Δ*darA* (Fig. 2.3B and 2.3C). Unlike *darA*, the *hdf* and *ddrA* phenotypes could both be fully complemented *in trans* by vectors containing *hdf* or *ddrA* alone (Fig. 2.3B and 2.3C). The plaque sizes for both P1Δ*hdf* and P1Δ*darA* were smaller than that of P1, whereas the plaques of P1Δ*ddrA* appeared to be of normal size.

ddrB modulates the EcoB and EcoK antirestriction phenotype

Having determined the restriction phenotypes in *ulx, hdf* and *ddrA* mutants, we tested other genes located within the *darA* and *darB* operons (Fig. 2.1B). Surprisingly an isogenic deletion of *ddrB* was found to have a ~10-fold higher EOP on cells expressing *Eco*B or *Eco*K than that of the parental P1 (Fig. 2.4). This phenotype could again be complemented *in trans* by a vector containing *ddrB* alone, with the EOP returning to ~3 in the complemented phage. Deletion and complementation of *ddrB* produced no notable phenotype on *Eco*A strains. Both *ulx* and *darB* are similar to *ddrB* in this regard, in that their deletion affects susceptibility to *Eco*B and *Eco*K restriction, with no apparent effect on *Eco*A strains (Fig. 2.2). It seems likely that *ddrB* exerts its phenotype via action on Ulx and/or DarB, either by regulating the copy number of these proteins incorporated into the virion, or by modulating their activity after the antirestriction system is deployed in the host cell.

Virion proteomics of P1 and its mutants

It was shown previously that P1*darA*$^{-}$ phages fail to package DarB (Iida *et al.*, 1987), leading to the conclusion that, in addition to being required for protection of DNA against *Eco*A, DarA is also required for incorporation of DarB into the virion. In the present study, multiple genes in addition to *darA* and *darB* have been identified with roles in the antirestriction phenotype of P1. These phenotypes can be placed into two broad categories: *darA*-like, which affects phage EOP on *Eco*A, *Eco*B and *Eco*K strains,

and *darB*-like, in which EOP is affected only on *Eco*B and *Eco*K strains. In order to determine how these multiple P1 genes could produce similar restriction phenotypes,

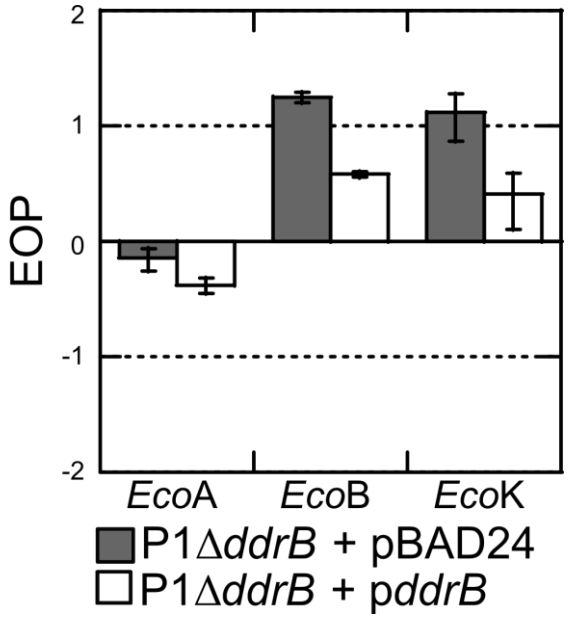


Fig. 2.4. *ddrB* negatively affects the *darB* phenotype. Reprinted with permission from Piya *et al*., 2017. Phages were induced and EOP assays were performed as described previously. The EOP data shown have been normalized to EOP of parental P1 induced from modification-deficient WA921 containing pBAD24. Unlike the other restriction phenotypes described, disruption of *ddrB* increased the relative EOP of the phage in *Eco*B and *Eco*K strains by approximately 10-fold compared to wild-type P1 (gray bars). This effect did not extend to *Eco*A strains, suggesting that DdrB negatively impacts the activity of DarB. The *ddrB* phenotype could also be complemented *in trans* (white bars).

proteomic analysis of purified virions was conducted (Fig. 2.5). Protein bands corresponding to DarB, Hdf, DarA and DdrB could be assigned by mass spectrometry analysis (underlined in Fig. 2.5A). The protein band corresponding to DdrA is assigned based on missing band density corresponding to its predicted protein molecular mass in lanes of P1Δ*ddrA* virions compared to the parental P1. The protein band corresponding to Ulx was assigned based on western blotting of purified P1Δ*ulx* virions complemented with a Ulx-FLAG fusion (Inclan *et al.*, 2016) (Fig. 2.6).

Disruption of *darB* resulted in the loss of DarB and Ulx (Fig. 2.5), indicating only Ulx is dependent on DarB for incorporation. Coupled with the conserved domains present in DarB, the *Eco*B and *Eco*K phenotypes observed for P1Δ*darB* (Fig. 2.2A) suggests that DarB is directly responsible for *Eco*B and *Eco*K antirestriction activity. Band densitometry indicates P1Δ*ulx* packages only ~15% of the DarB as the parental P1 (data not shown). Thus, DarB can be packaged into the virion in the absence of Ulx but packaging efficiency is greatly reduced. This reduced packaging of DarB is consistent with the ~10-fold weakened antirestriction phenotype exhibited by P1Δ*ulx* compared to P1Δ*darB* (Fig. 2.2).

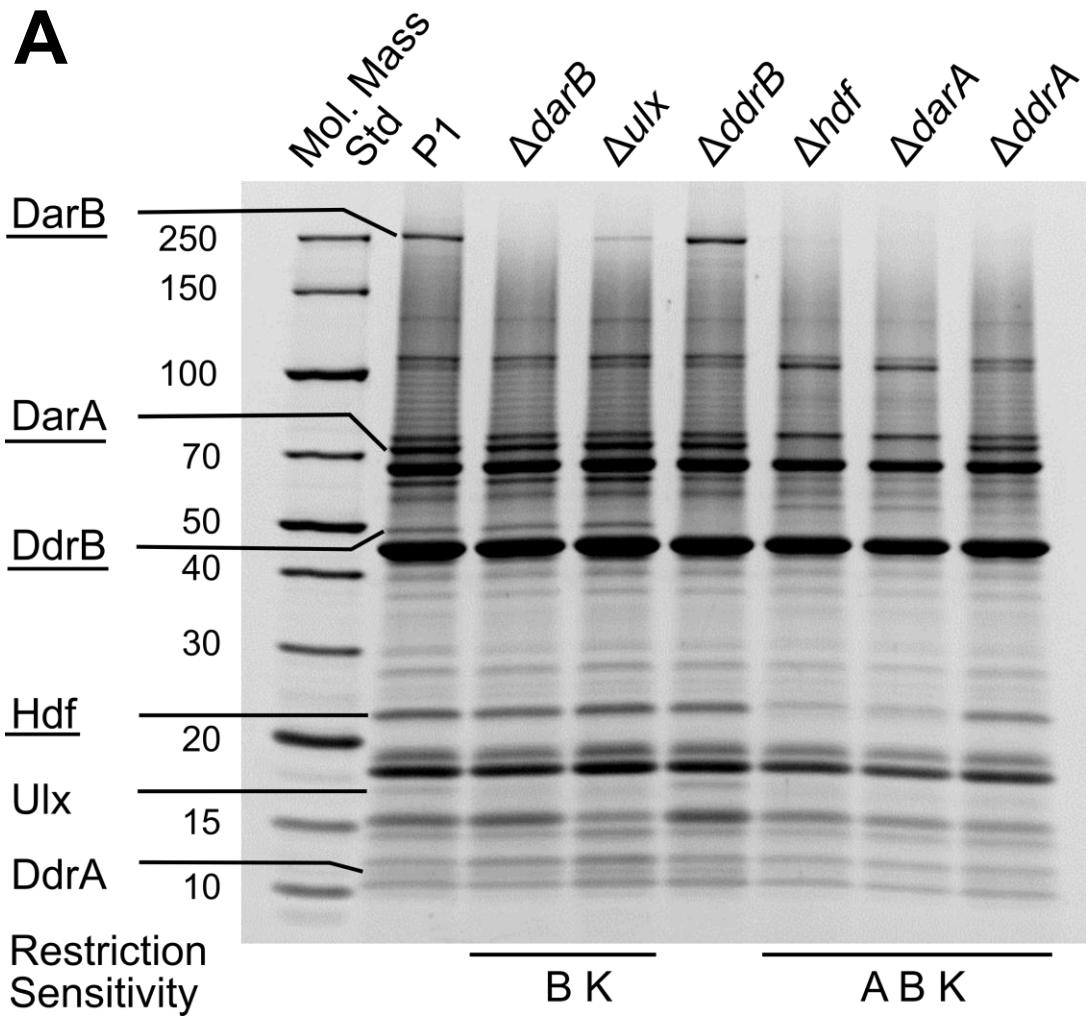


Fig. 2.5. Comparative SDS-PAGE of parental P1 and isogenic mutants suggests cascading dependencies for protein incorporation into the P1 virion. Reprinted with permission from Piya *et al*., 2017. Panel A: SDS-PAGE of parental P1 and antirestriction mutants. Molecular masses of the size standard are provided in kDa on the left. Relevant protein bands are labeled on the left; proteins that have been identified by mass spectrometry are underlined. DdrA is assigned based on the missing band density in the gel image and predicted molecular mass, and Ulx is assigned based on western blotting of P1Δ*ulx* complemented with Ulx-FLAG with anti-FLAG antibody. At the bottom of the figure, the restriction sensitivity phenotypes of the P1 mutants are indicated: B, K (sensitive to *Eco*B and *Eco*K restriction), or A, B, K (sensitive to *Eco*A, *Eco*B and *Eco*K restriction).

B

P1 proteins	P1 and antirestriction mutants						
	P1	Δ*darB*	Δ*ulx*	Δ*ddrB*	Δ*hdf*	Δ*darA*	Δ*ddrA*
DarB	+	-	(+)	+	-	-	-
DarA	+	+	+	+	-	-	+
DdrB	+	+	+	-	-	-	-
Hdf	+	+	+	+	-	-	+
Ulx	+	-	-	+	-	-	-
DdrA	+	+	+	+	-	-	-

Fig. 2.5 Continued. Panel B: The table summarizes protein presence or absence in P1 mutants. P1Δ*ulx* shows a faint DarB band, which suggests that DarB can be incorporated in the absence of Ulx, albeit inefficiently. When either *hdf* or *darA* is disrupted, P1 fails to package both Hdf and DarA, suggesting that Hdf and DarA are co-dependent for virion incorporation.

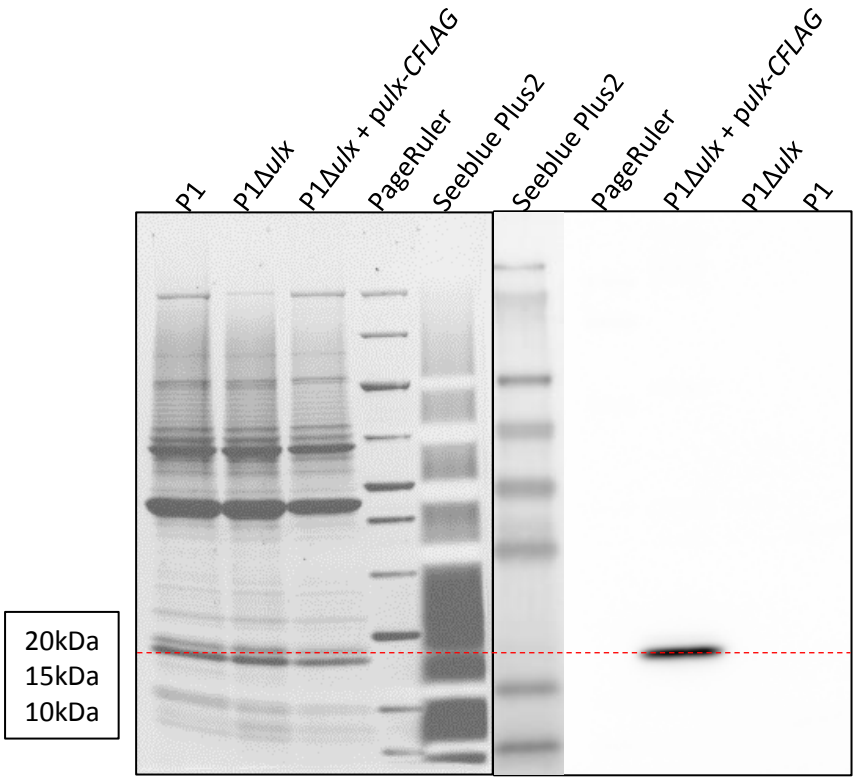


Fig. 2.6. Identification of Ulx band in SDS-PAGE of P1 virions. Reprinted with permission from Piya *et al*., 2017. Ulx protein band in P1 virions was identified by western blot on P1Δ*ulx* complemented with functional Ulx fused to FLAG-tag on C-terminal end. The protein samples for SDS-PAGE were normalized and prepared as described in materials and methods. The samples were loaded in 4-20% Tris-glycine (Thermo Scientific) in order as shown above. After SDS-PAGE, the gel was cut into two halves. One half was stained with SyproRuby as described earlier and the other half was used for western blot following standard protocol (Green & Sambrook, 2012). In brief, SDS-PAGE resolved proteins were transferred to PVDF membrane (GE Healthcare) overnight at 30V. The membrane was blocked with 1% blocking buffer and probed with Anti-FLAG primary antibody (Thermo Scientific) and HRP-conjugated secondary antibody. The membrane was developed with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) and imaged (BioRad).

Compared to the parental P1, a single ~50 kDa band is missing in the P1Δ*ddrB* virion; this species was confirmed as DdrB by mass spectroscopy analysis. Aside from this 50 kDa species, all bands including those corresponding to DarB, Ulx, DarA, Hdf, and DdrA are present, indicating that no other virion proteins are dependent upon DdrB for their incorporation. Since *ddrB* encodes a protein with a predicted molecular mass of ~109 kDa, the appearance of DdrB at a position corresponding to 50 kDa suggests that the protein is proteolytically processed for incorporation into the virion; no ~109 kDa band corresponding to DdrB was observed in SDS-PAGE. Based on MS-MS analysis of the gel slice containing this predicted DdrB band from the parental P1, high peptide coverage (5 to 27-fold, covering 88% of residues) was observed for the N-terminal portion of the protein sequence, with coverage dropping abruptly after residue 421 (1 to 3-fold, covering ~42% of residues), suggesting that it is primarily the N-terminal portion of the DdrB protein that is present in the P1 virion. Cleavage of the protein at a position at or shortly after 421 would result in a ~46 kDa product, which matches closely the observed 50 kDa DdrB band. Presumably some small amount of the C-terminal fragment somehow remains with the virions, and this fragment would also be ~50 kDa and co-migrate with the C-terminal portion. The phenotype associated with *ddrB* deletion is an increase in protection against *Eco*B and *Eco*K restriction. One possibility is that the absence of DdrB allows for incorporation of more DarB into the virion, providing greater antirestriction activity. However, band densitometry does not convincingly show greater DarB band intensity in *ddrB* mutants (data not shown). An alternative hypothesis is that DdrB actively suppresses DarB-associated antirestriction activity. In

either case, the role of DdrB in the P1 antirestriction system is not clear, as its presence appears to counterintuitively hinder the fitness of P1 as measured by its ability to overcome host restriction.

As shown in Fig. 2.5, P1Δ*hdf*, P1Δ*darA* and P1Δ*ddrA* virions all fail to incorporate DarB, DdrB and Ulx. *Eco*B and *Eco*K restriction sensitivity of these mutants can be explained by the absence of DarB because DarB is most likely responsible for *Eco*B and *Eco*K antirestriction activity as discussed above. However, *hdf*, *darA* and *ddrA* mutants are also all sensitive to *Eco*A restriction, which suggests that one of these three genes are responsible for the *Eco*A antirestriction activity of P1. Deletion of either *hdf* or *darA* results in virions missing both Hdf and DarA, which indicates that these proteins are co-dependent for virion incorporation. Both the P1Δ*hdf* and P1Δ*darA* virions are also missing DdrA in addition to DarB, DdrB and Ulx mentioned above. It was suggested that DarA is responsible for EcoA protection (Iida *et al.*, 1987), however, DarA was found to be present in P1Δ*ddrA* virions, which are sensitive to *Eco*A restriction (Figs. 2.5, 2.3C). This suggests that DarA is not solely responsible for protection from *Eco*A restriction as previously proposed. Aside from DarB, DdrB and Ulx, the only protein observed to be absent in P1Δ*ddrA* is DdrA itself. Thus, DdrA may be the protein that protects P1 DNA from *Eco*A-mediated endonuclease activity, although it is not clear if DarA and Hdf are also required for DdrA activity or are simply required for incorporation of DdrA into the virion.

Band densitometry was performed on SyproRuby-stained SDS-PAGE of P1 using ImageJ (Schneider *et al.*, 2012) to estimate the copy numbers of P1 antirestriction system components associated with the virion. These calculations suggest that there are ~45 DarB, ~100 DarA, ~40 DdrB and ~400 Hdf molecules in the average P1 virion (Table 2.3). The copy number of Ulx and DdrA could not be estimated due to their low mass and band intensity. These values indicate that P1 capsids contain ~35 MDa of major capsid protein and a combined mass of ~28 MDa of Dar proteins (DarB, DarA, DdrB and Hdf) per virion, indicating the Dar proteins contribute to a significant proportion of the mass of P1 heads.

Morphological defects are linked to hdf and darA

Under normal conditions, P1 is known to produce three morphological head size variants termed "big" (P1B, ~86 nm), "small" (P1S, ~65 nm) and "minute" (P1M, ~47 nm) (Walker & Anderson, 1970). The P1B variant (hereafter referred to as "normal" head size or P1N) packages the complete genome and is infectious, while P1S and P1M are defective as they can package only ~40% and ~10% of the P1 genome, respectively (Walker & Anderson, 1970). The proportion of P1S virions has been reported to vary depending upon the host, with P1 lysates induced from *E. coli* K-12 containing 20-37% P1S heads; P1M heads have been reported as less than 1% of most lysates (Walker & Anderson, 1970).

Table 2.3. Copy number of virion-associated proteins of P1 antirestriction system. **Reprinted with permission from Piya *et al*., 2017.** The copy number of constituent proteins of P1 antirestriction system was determined by densitometry on SyproRuby stained SDS-PAGE of P1 and mutants using ImageJ. The predicted copy number of tail sheath protein was determined based on the number of striations previously reported (six copies per striation, 54 striations) (Walker & Anderson, 1970) and used as an internal standard in all calculations. The molecular mass of tail sheath, DarB, and Hdf were determined from the predicted translational product of their respective genes. The molecular mass of DarA was determined based on the processing of product of darA into the mature form found in virions (Streiff *et al.*, 1987). Finally, the molecular mass of major capsid protein (MCP) and DdrB were estimated based on their migration on SDS-PAGE. The copy number of MCP was also calculated from densitometry as a quality check; there are predicted to be 955 copies of the MCP per virion based on the previously reported T=16 P1 capsid symmetry (Walker & Anderson, 1970). The copy number of MCP as determined by densitometry is ~17% lower than expected, which provides some indication of the accuracy of this analysis. Data is shown for analyses of two replicate SDS-PAGE gels.

Phage P1 proteins	Molecular Mass (kDa)	Copy number per virion			Total predicted protein mass per virion (MDa)
		Repeat 1	Repeat 2	Average	
Tail Sheath	57	324*	324*	324*	18
Major Capsid Protein	44	769	819	794	35
DarB	252	41	47	44	11
DarA	59	108	102	105	6
DdrB	50	40	34	37	2
Hdf	22	426	371	399	9

*: predicted copy number based on literature; this number was used as the basis for calculating copy number of the other proteins.

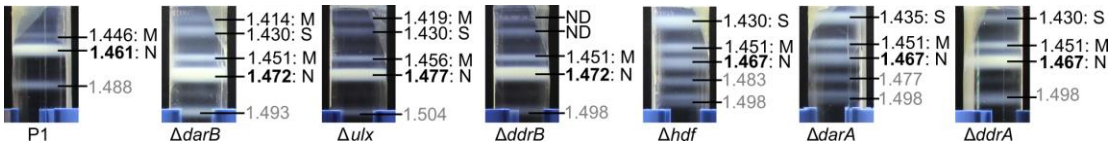


Fig. 2.7. Purification of virions using Cesium chloride isopycnic gradient centrifugation. Reprinted with permission from Piya *et al*., 2017. Parental P1 and mutants were purified by using isopycnic gradient of Cesium chloride as described earlier. The phage bands were extracted using a syringe and refractive index was measure using Abbe Refractometer. The refractive index was converted to density (g cc^{-1}) by comparison with standard CsCl density-refractive index correlation table. The corresponding phage bands indicated with bold-face densities were used for comparative SDS-PAGE analysis shown in Fig. 5. The phage bands corresponding to grey-color densities denote tail-less phage heads. After dialysis, virions from all extracted bands were imaged under TEM and head size survey was conducted as described earlier. The major population of head-size variants present in each phage bands is indicated after density, separated by a colon (:). "N" denotes normal head size virions, "S" denotes small head size virions, and "M" denotes mixed population of normal and small head size virions.

In isopycnic gradient purification of P1 and its antirestriction mutants, variations in banding patterns were observed in some of the mutant phages, particularly P1Δ*hdf* and P1Δ*darA* (Fig. 2.7). These mutants exhibited more intense bands at regions in the gradients corresponding to lighter buoyant densities than the parental P1, which suggested that the disruption of these genes also affected viral morphogenesis. The phage bands were extracted whenever possible and the head size of purified phages were measured from negative-stained TEM images and their distribution is described in Fig. 2.7.

In order to study the effects of antirestriction gene deletions on the overproduction of P1S virions, the parental and mutant phages were induced from lysogens of *E. coli* MG1655 and a head-size survey was conducted for each phage by measuring the diameters of several hundred phage heads photographed from negative-stain TEM images prepared from the phage lysates. Observed phage head diameters ranged from 52 nm to 95 nm (Fig. 2.8). In accordance with previous reports, head diameters >67 nm were classified as "normal" (i.e., P1N) and head diameters ≤67 nm were classified as small (P1S) variants. Virions with "minute" (P1M) head size were rarely observed and were excluded from classification. All phages produced heads of a bimodal size distribution, with a smaller size class of ~60 nm and a normal size class centered around ~80 nm (Fig. 2.8, Table 2.4). As shown in Fig. 2.8, the parental P1 produced virions with ~82% normal-sized heads and ~18% small heads, in agreement with previous observations of P1 (Walker & Anderson, 1970). In lysates of P1Δ*hdf* and P1Δ*darA* this relationship was essentially inverted, with P1Δ*hdf* producing ~85 % small-headed virions and P1Δ*darA* producing ~82% small-headed virions. Other phage mutants produced small heads at frequencies similar to the parental P1, ranging from 14.1% for P1Δ*ddrB* to 28.1% for P1Δ*darB*, however this variation is within the range previously reported for P1 (Walker & Anderson, 1970).

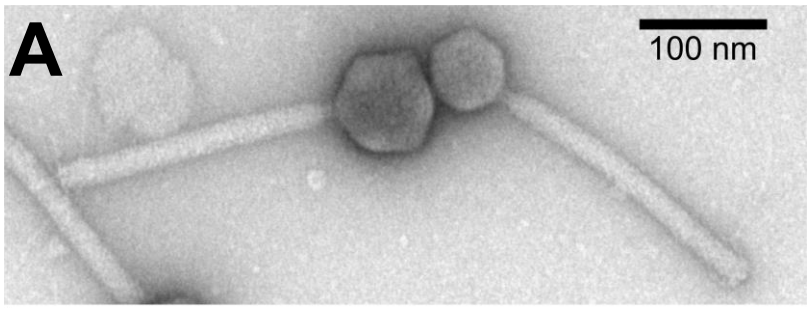


Fig. 2.8. Head size variation in parental P1 and mutants. Reprinted with permission from Piya *et al*., 2017. Lysates of phage P1 contain virions of two major head size classes. The virions of >67 nm head diameter have been classified as "normal" and those of ≤ 67 nm as "small". **Panel A:** Transmission electron micrograph (TEM) showing typical normal head (left) and small head (right) virions. **Panel B:** The diameters of parental P1 and antirestriction mutant virions were measured from TEM images. The X-axis in each histogram denotes the head size diameter in nm and the Y-axis denotes the percentage of virions of each head size present. The phage genotype (parental P1 or antirestriction mutant) and the number of virions measured are denoted in the upper right of each histogram. On the right, the percentages of virions falling into the small and normal head size classes for each phage mutant are summarized. The *darB*, *ulx*, *ddrB* and *ddrA* mutants have similar head size distributions as parental P1, while this distribution is reversed in *hdf* and *darA* mutants, which are dominated by small-headed virions.

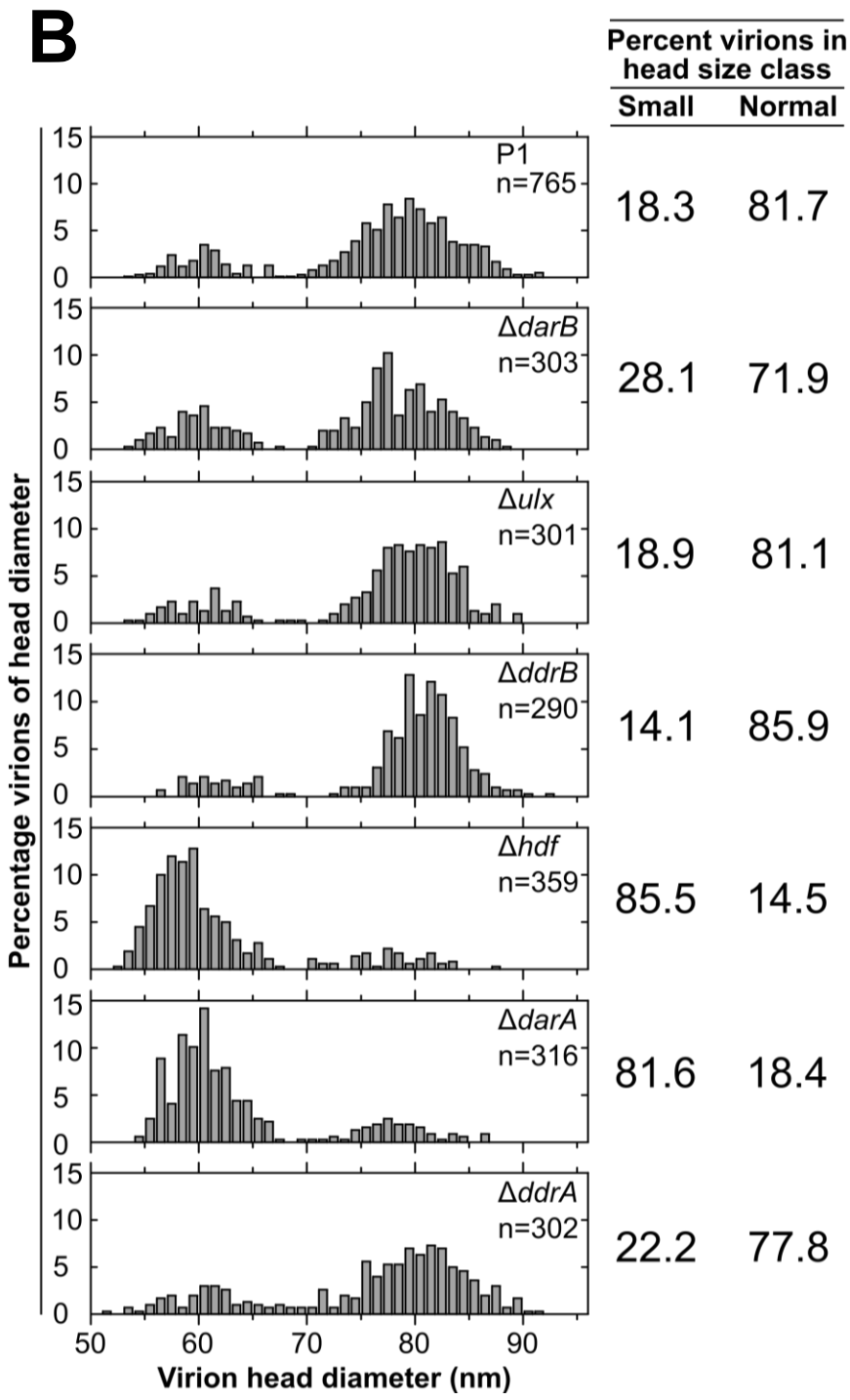


Fig. 2.8 Continued.

Table 2.4. Mutations in phage P1 antirestriction genes affect head morphogenesis. Reprinted with permission from Piya *et al*., 2017. Lysates of phage P1 and its antirestriction mutants were observed by negative stain TEM and head diameters of the phage populations were measured. Phage heads were placed into two size classes, small (≤ 67 nm) and normal (> 67 nm). P1 *hdf* and *darA* mutants exhibited high proportions of small-headed virions.

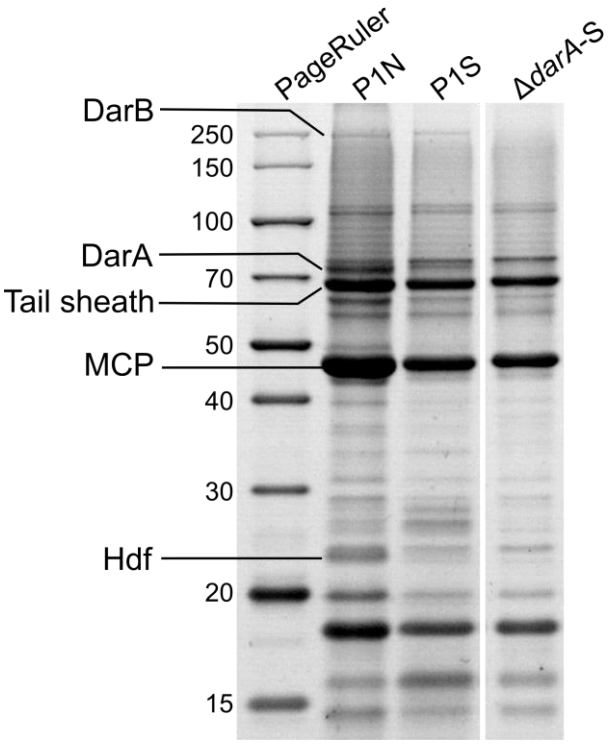
Phage strain	No. virions measured	Dimension of small head (nm) (Average ± S.D.)	Dimension of normal heads (nm) (Average ± S.D.)
P1	765	60 ± 3	79 ± 4
Δ*darB*	303	59 ± 3	78 ± 4
Δ*ulx*	301	60 ± 3	80 ± 4
Δ*ddrB*	290	61 ± 3	81 ± 3
Δ*hdf*	359	58 ± 3	77 ± 4
Δ*darA*	316	60 ± 3	78 ± 4
Δ*ddrA*	302	60 ± 4	80 ± 5

Previous work has shown that deletion of the entire *darA* operon, or transposon insertions within *lydA* and *darA*, resulted in overproduction of P1S virions in a phenotype denoted Vad$^-$ (for viral architecture determinant) (Iida *et al.*, 1998). Based on the data presented here, the Vad$^-$ phenotype can be assigned to two genes within the *darA* operon, *hdf* and *darA*. These genes are directly adjacent to each other (Fig. 2.1B) and are the earliest non-lysis genes present in the P*dar* transcript. Deletion of either gene results in the failure to incorporate any of the other virion-associated antirestriction components (Fig. 2.5), and the loss of antirestriction activity against *Eco*A, *Eco*B and *Eco*K (Fig. 2.3). The presence of the Vad$^-$ phenotype in both P1Δ*hdf* and P1Δ*darA* indicates that both Hdf and DarA play important roles in determining head size during P1 morphogenesis.

In many viruses (bacteriophages such as P22, T2, λ and eukaryotic viruses such as HSV-1 and other herpesviruses), assembly of the major capsid protein into phage heads of the correct size and shape is guided by a scaffolding protein, resulting in a spherical immature procapsid (Mateu, 2013). Studies in phages T4, P22 and λ have shown that procapsids are assembled initially without any DNA (King *et al.*, 1980). During procapsid maturation and concomitant DNA packaging, the scaffolding protein is lost from the capsid (Dokland, 1999). Absence of functional scaffolding proteins results in aberrant capsid assembly or no assembly at all (Dokland, 1999; Aksyuk & Rossmann, 2011). No major scaffold protein has been identified for P1 (Lobocka *et al.*, 2004). Disruption of *hdf* and *darA* results in an increased proportion of small P1 heads (Fig.

2.8), suggesting that Hdf and DarA play roles in directing proper capsid assembly, most likely at the time of procapsid formation. However, Hdf and DarA are still present in mature virions (Fig. 2.5) and hence do not fit the classical definition of scaffolding proteins which are present in viral proheads but absent in mature virions (Dokland, 1999). Moreover, the presence or absence of these proteins do not result in strict phenotypes: the deletion of *hdf* or *darA* still produces ~20% normal-sized phage heads, and the parental P1 containing intact *hdf* and *darA* produces ~20% aberrant capsids. Hdf and DarA may not play a direct role in the observed antirestriction activity of P1, but rather their absence prevents the incorporation of the other antirestriction system components.

Since the absence of either Hdf or DarA was found to induce greater production of small head-sized virions, we wished to determine if the small head-size progeny of parental P1 are capable of incorporating Hdf and DarA. Comparative SDS-PAGE band intensities of P1N and P1S virions suggest that small-head phages incorporate ~65% of the major capsid and ~40% of the DarB proteins relative to normal P1N virions (Fig. 2.9). These numbers are comparable to what would be expected from a reduction in head diameter from 80 nm to 60 nm, with concomitant reductions of capsid surface area by ~56% and capsid volume by ~42%. While the P1S virions are capable of incorporating both DarA and Hdf, band densitometry suggests that P1S heads incorporate only ~14% and ~12% of DarA and Hdf proteins, respectively, relative to P1N (Fig. 2.9). At this point it is difficult to determine if P1S heads are produced as a result of reduced incorporation of



Phage P1 proteins	Percent band intensities of proteins in P1S relative to P1N virions		
	Repeat 1	Repeat 2	Average
MCP	62	68	65
DarB	38	44	41
DarA	14	14	14
Hdf	16	8	12

Fig. 2.9. Comparative SDS-PAGE of P1N and P1S virions. Reprinted with permission from Piya *et al*., 2017. P1N, P1S and P1Δ*dar*A small head-size virions were extracted from isopycnic CsCl gradients, run on 12% Tris-glycine SDS-PAGE and stained with SyproRuby as described above. Band intensities in each lane were normalized to the tail sheath band and densitometry was performed on the major capsid protein (MCP), DarB, DarA and Hdf bands in ImageJ. As expected, bands corresponding to DarB, DarA, and Hdf were not observed in P1Δ*darA* small head-size virions. This analysis was repeated twice and the relative band intensities measured in P1S virions are summarized in the table below the gel image.

Hdf and DarA, or if lower amounts of Hdf and DarA are incorporated into proheads that have already committed to the smaller size.

The incorporation of the P1 antirestriction proteins into the capsid appears to be linked to the early stages of capsid morphogenesis. All of the virions run on SDS-PAGE in Fig. 2.5 were taken from CsCl bands corresponding to normal phage buoyancy (1.46-1.48 g cc^{-1}, Fig. 2.7) and electron microscopy surveys of these fractions confirmed that they contained >95% normal head size P1N virions (data not shown). Even though the P1Δ*hdf* and P1Δ*darA* virion proteomes shown in Fig. 2.5 are the minority fraction of the phage lysate containing normal-sized heads, the downstream antirestriction components DdrA, DdrB, DarB and Ulx are still absent from the virions, demonstrating that the formation of normal-sized heads is in itself not sufficient for incorporation of these antirestriction proteins. This suggests that it is Hdf and DarA themselves, and not their effects on head morphogenesis, that determine the incorporation of the other antirestriction proteins.

Conclusions

The P1 Dar system is composed of multiple proteins: Hdf, DarA, DdrA, DdrB, DarB and Ulx. The role of Hdf and DarA in antirestriction system is unique because they are also essential for the production of normal head-size virions. In the presence of Hdf and DarA, ~80% of P1 progeny are of normal head-size and ~20% are of small head-size (Fig. 2.10A). The distribution of normal and small head-size is reversed in the absence

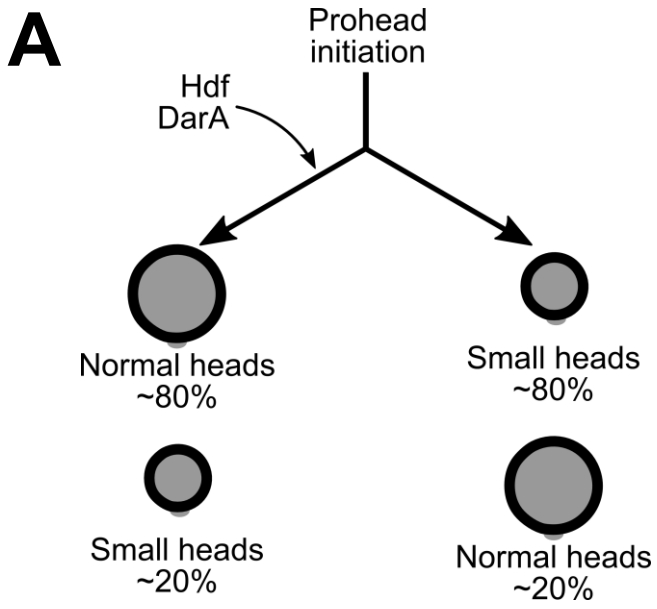


Fig. 2.10. Models of P1 head-size determination and antirestriction component incorporation. Reprinted with permission from Piya *et al*., 2017. Panel A: The parental P1 with intact *darA* and *hdf* produces ~80% normal head-size (P1N) and ~20% small head-size (P1S) progeny. In the absence of either Hdf or DarA, this ratio is inverted, producing progeny with mostly (~80%) aberrant small heads. **Panel B:** Incorporation of P1 antirestriction system components follows a distinct pathway. The pathway presumably begins with the initiation of prohead formation by the phage portal and scaffold proteins, followed by co-incorporation of Hdf and DarA into the prohead. In addition to exacerbating the head-size defect as shown in Panel A, the absence of Hdf or DarA prevents the incorporation of all downstream Dar components. The incorporation of Hdf and DarA does not confer protection against *Eco*A, *Eco*B or *Eco*K restriction. Incorporation of DdrA appears to be the next step, as *ddrA* mutants incorporate only DarA and Hdf. Incorporation of DdrA provides protection against *Eco*A restriction. The incorporation of DarB, DdrB and Ulx is difficult to separate into separate steps, but packaging of DarB into virions provides protection against *Eco*B and *Eco*K restriction. Ulx appears to be an enhancer of DarB packaging as *ulx* mutants incorporate less DarB than normal, while DdrB acts as a negative regulator of DarB activity.

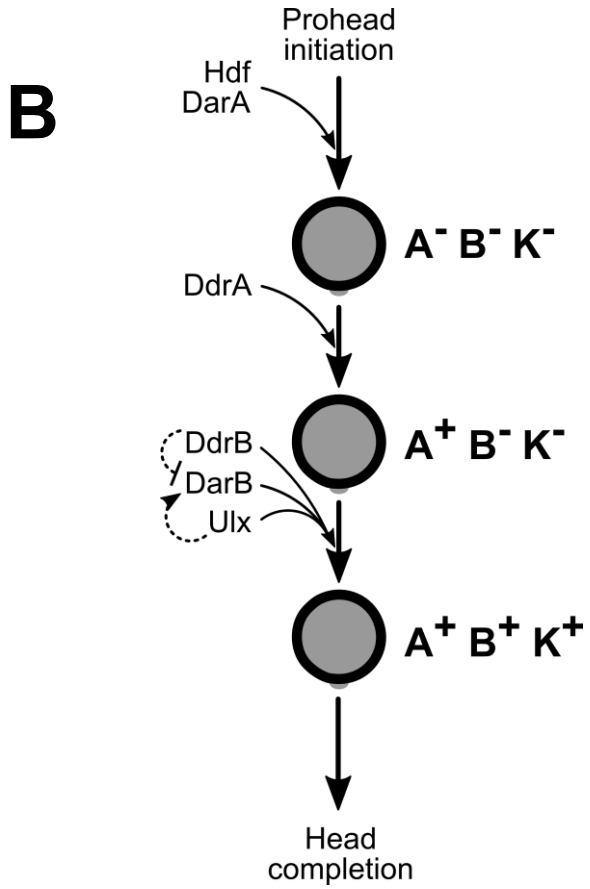


Fig. 2.10 Continued.

of either Hdf or DarA (Fig. 2.8). Because the absence of either Hdf or DarA results in a high abundance of aberrant small capsids, it is probable that these proteins play a role in the early stages of prohead formation. The proteins Hdf, DarA and DdrA are required for the protection of phage DNA from restriction by the *Eco*A Type I R-M system, and DarB is required for protection against *Eco*B and *Eco*K. Ulx enhances the antirestriction phenotype of DarB, and DdrB appears to negatively affect this activity by an unknown

mechanism (Fig. 2.10B). All of these phenotypes could be complemented *in trans.* Proteomic analysis of isogenic P1 mutants shows a clear order of incorporation for each component into the mature virion, as illustrated in (Fig. 2.10B). In this proposed model, Hdf and DarA are incorporated first in a co-dependent manner, shortly after prohead initiation. The absence of either Hdf or DarA results in failure to incorporate any of the other antirestriction proteins. Following Hdf and DarA, DdrA is incorporated, followed by DarB, DdrB and Ulx. The incorporation of these last three proteins are not strictly co-dependent: the absence of DarB results in no apparent Ulx incorporation, but absence of Ulx results only in reduced DarB incorporation. This behavior suggests that Ulx acts as a chaperone or packaging factor to increase the amount of DarB incorporated into the virion. The presence of DdrB results in reduced antirestriction activity, presumably by negatively affecting DarB but the mechanism is not known.

Given the incorporation of these proteins into the P1 virion and the *cis*-acting nature of the antirestriction phenotype, it is evident that the P1 antirestriction system must exert its activity upon infection of the host cell (Iida *et al.*, 1987). Since it is difficult to imagine how these proteins could act while still in the confines of the capsid, they must be introduced into the host cytoplasm via the phage tail upon infection. It is not a requirement that all six of the antirestriction proteins described here be translocated into the cytoplasm to perform some direct protective action. Some proteins, particularly the more upstream components such as DarA and Hdf, likely only serve to ensure the incorporation of the active components into the virion. Strictly speaking, a minimum of

only two proteins are required to account for the observed phenotypes: one that provides protection against *Eco*A, and one that protects against *Eco*B and *Eco*K. Based on bioinformatic evidence, DarB is the likely candidate to provide *Eco*B and *Eco*K protection as the predicted methyltransferase and helicase domains of DarB suggest a possible enzymatic mechanism to provide protection from host restriction. The protein or proteins responsible for the observed anti-*Eco*A activity are more difficult to assign. DdrA is one candidate, as phage lacking DdrA contain DarA and Hdf but do not express any antirestriction activity.

The translocation of phage proteins from the head into the host cytoplasm is a known feature in several *Caudovirales* phages. The capsid of the temperate *Salmonella* Typhimurium podophage P22 contains three internal proteins, gp16, gp20, and gp7, that leave the confines of the capsid to perform their function of assisting in translocation of DNA across host membranes (Jin *et al.*, 2015). Coliphage N4 is known to translocate its ~380 kDa DNA-dependent RNA polymerase from the phage head into the host cytoplasm early in the infection process (Choi *et al.*, 2008). Perhaps the closest analog to the P1 system shown in this work is the paradigm coliphage T4, whose heads contain three non-essential internal proteins, IPI, IPII and IPIII (Black & Ahmad-Zadeh, 1971). The ~10 kDa T4 IPI is ejected into the host cytoplasm during phage infection and protects the incoming phage DNA by inhibiting a glucosyl-hmC DNA-specific restriction endonuclease (Bair *et al.*, 2007). The proteins of the Dar system are assembled into the P1 capsid and at least some of them are ejected into the host

cytoplasm, where they protect P1 DNA from cleavage by Type I R-M systems. In our study of the P1 antirestriction system, we determined that the incorporation of the P1 antirestriction proteins follow a distinct pathway that involves at least six proteins. We also found that proteins involved in the assembly pathway influence head morphogenesis. Phage P1 is the first reported instance, to our knowledge, where the antirestriction system has also been implicated in capsid morphogenesis.

CHAPTER III

CHARACTERIZATION OF P1 GENES OF UNKNOWN FUNCTION AND PACKAGING OF DARB INTO P1 PROCAPSIDS

Introduction

Bacteriophages (phages) are natural predators of bacteria. The constant evolutionary arms race between bacteria and phages has led to development of diverse defense and counter-defense strategies against each other (Labrie *et al.*, 2010). *Caudovirales* phages such as P1 initiate the infection cycle by adsorption to host receptors, followed by ejection of DNA into host cytoplasm. Bacterial restriction and modification (R-M) systems provide a barrier to foreign DNA, including phage DNA, by cleaving inappropriately modified DNA. Based upon subunit stoichiometry, DNA specificity and restriction mechanisms, R-M systems are broadly categorized as type I, II and III. (Kruger & Bickle, 1983; Loenen *et al.*, 2014). Please refer to Chapter I of this dissertation for detailed information on mechanisms of action of restriction-modification systems.

To avert DNA cleavage by the bacterial R-M systems, phages have evolved with diverse antirestriction mechanisms (Tock & Dryden, 2005). Enterobacteriaceae phage P1 protects its DNA from type I R-M system mediated DNA cleavage by injecting Dar proteins into host. Earlier reports indicated that Dar system was composed of two proteins, DarA and DarB, which are encoded in two distinct operons in P1 genome (Iida

et al., 1987; Lobocka *et al.*, 2004). It was determined from plating assays that P1 *darA* mutants are severely restricted in host with type I *Eco*A, whereas P1 *darB* mutants are severely restricted in host with type I *Eco*B or *Eco*K systems (Iida *et al.*, 1987). Recent study has shown that several gene in *darA* and *darB* operons, including *darA* and *darB* themselves, encode components of P1 antirestriction system. P1 mutants deleted for *ulx* show *darB*-like restriction phenotype, whereas mutants deleted for *hdf* or *ddrA* show *darA*-like phenotype (Fig. 3.1). Moreover, Hdf and DarA, which were shown to be required for incorporation of all Dar proteins, were also found to affect capsid size determination (Piya *et al.*, 2017). Since multiple genes in two distinct operons comprise the Dar system, we hypothesized that P1 genes from other different operons might influence P1 antirestriction system. We also interested to see if any other P1 genes could affect capsid size determination.

Genetic evidence has established that DarB is required for protection of P1 DNA from type I *Eco*B and *Eco*K systems following ejection of DNA into host (Iida *et al.*, 1987; Piya *et al.*, 2017). Presence of DarB in mobile genetic elements suggest that DarB could protect these elements from restriction during their transfer across different hosts in nature (Gill *et al.*, 2011). Biological function of DarB has been studied in context of phage P1 only. However, little is known about the assembly of DarB into P1 procapsids and biochemical activity of DarB leading to protection of DNA.

P1 Dar system is encoded in two operons, separated by genes of unrelated functions, thus it is possible that other as-yet undetected genes may also play a role in this

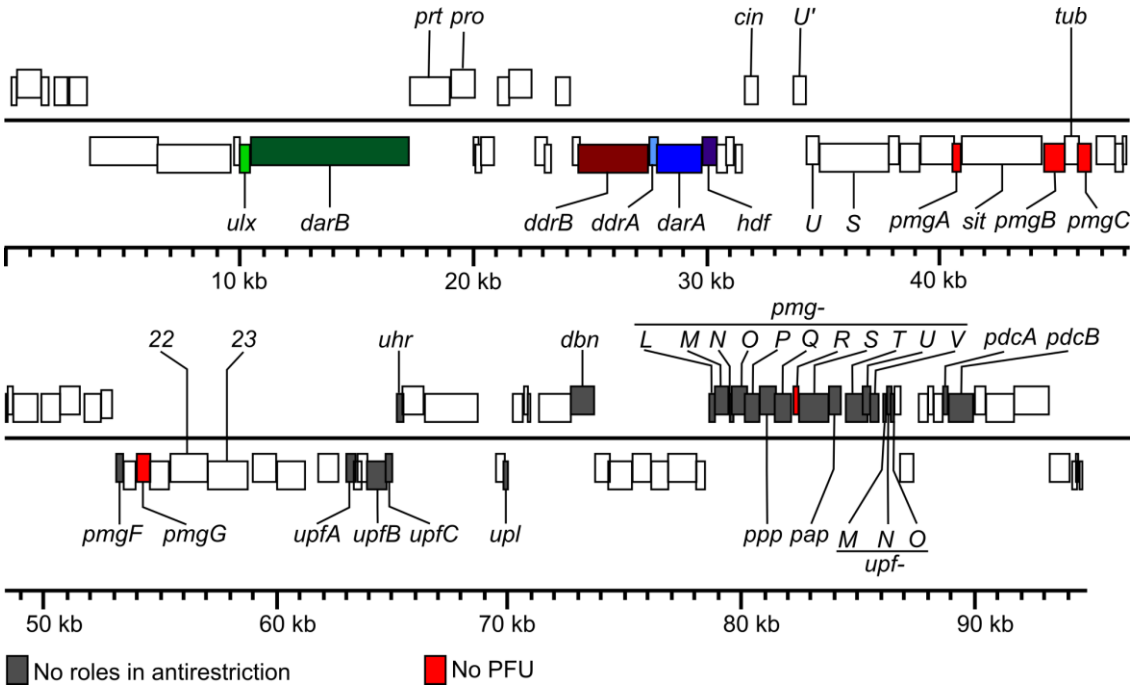


Fig. 3.1. Genome map of P1. Rectangular blocks in the figure represent P1 genes. The genes above the black line are transcribed towards right, whereas those below the black line are transcribed towards left. The genes *darB*, *ulx*, *hdf*, *darA*, *ddrA*, and *ddrB* encoding known components of P1 antirestriction are indicated. P1 genes of unknown function that were knocked out in this study are colored red or grey. P1 mutants knocked out for genes in red color did not form plaques and were not tested for plating efficiency. P1 mutants knocked out for genes in grey color formed plaques and were lysogenized into *E. coli* strain WA921. Thermally induced lysates were tested for plating efficiency in *E. coli* strains with type I *Eco*A, *Eco*B or *Eco*K system. The genes encoding major structural components are indicated for reference. The ruler denotes the nucleotide position of genes as in the published P1 genome (NC_005856.1).

antirestriction system. In order to identify other components of the P1 antirestriction system, if any, a library of P1 mutants knocked out for genes of unknown function was constructed. In order to further examine the incorporation of Dar components into mature virions, P1 mutants defective for DNA packaging was constructed. Several truncations of DarB were also generated to study mechanisms of its incorporation into P1 capsids. Moreover, we have optimized purification of DarB to elucidate reaction mechanisms *in vitro*.

Materials and methods

Bacterial strains and phages

The bacterial strains and phages used in this study are listed in Table 3.1. All phage P1 mutants used in this study are isogenic gene deletions of P1CM*clr*100, simply referred to as P1 (Rosner, 1972). Unless specified otherwise, *E. coli* strains, were cultured in LB broth or LB agar at 37 °C as described before (Piya *et al.*, 2017). Culture was grown at 30 °C to maintain P1 lysogen and shifted to 42 °C to thermally induce P1. As needed, the medium was supplemented with 10 µg mL^{-1} chloramphenicol (LB cm), 30 µg mL^{-1} kanamycin (LB kan) or 100 µg mL^{-1} ampicillin (LB amp).

Generation of single gene P1 knockout mutants

Single-gene deletions knockout library of P1 was constructed by using λ Red recombinase mediated homologous recombination method as described before (Datsenko & Wanner, 2000). In brief, P1 was lysogenized into *E. coli*

Table 3.1. Bacterial strains, phages and plasmids

Strains, phages or plasmids	Genotype or relevant characteristic	Reference/source
***E. coli* strains**		
WA2379	*leu*$^-$ *met*$^-$ *lac*$^-$ r_{A^-} m_{A^-}	(Arber & Wauters-Willems, 1970) / The Coli Genetic Stock Center, Yale University (CGSC)
W3110	F$^-$ λ$^-$ *rpoS*(*Am*) *rph-1 Inv*(*rrnD-rrnE*) r_{K^+} m_{K^+}	(Iida *et al.*, 1987; Hayashi *et al.*, 2006) / CGSC
WA921	*thr*$^-$ *leu*$^-$ *met*$^-$ *lac*$^-$ r_{K^-} m_{K^-}	(Wood, 1966; Arber & Wauters-Willems, 1970) / CGSC
WA960	*thr*$_B^+$ *leu*$^-$ *met*$^-$ *lac*$^-$ r_{B^+} m_{B^+}	(Wood, 1966; Arber & Wauters-Willems, 1970) / CGSC
BW25113(pKD46)	F$^-$ l$^-$ *rpoS*(*Am*) *rph-1 rrnB3* Δ*lacZ4787 hsdR514* Δ(*araBAD*)*567* Δ(*rhaBAD*)*568*	(Baba *et al.*, 2006) / CGSC
BW25141(pKD4)	*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ Δ*phoBR580 hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *galU95 endA*$_{BT333}$ *uidA*(Δ*MluI*)::*pir*$^+$ *recA1*	(Datsenko & Wanner, 2000) / CGSC
MG1655	*lacI*^{q1}*tonA*::Tn*10*	(Park *et al.*, 2006) / Lab stock
Phages		
P1Δ*darA*	in-frame deletion of *darA* in P1CM*clr*100	(Piya *et al.*, 2017)
P1Δ*darB*	in-frame deletion of *darB* in P1CM*clr*100	(Piya *et al.*, 2017)
P1Δ*darA*Δ*pacA*	in-frame deletion of *pacA* in PΔ*darA*	This study
P1Δ*darB*Δ*pacA*	in-frame deletion of *pacA* in PΔ*darB*	This study
P1Δ*prt*	in-frame deletion of *prt* in P1CM*clr*100	This study
P1Δ*pro*	in-frame deletion of *pro* in P1CM*clr*100	This study
P1Δ*pmgA*	in-frame deletion of *pmgA* in P1CM*clr*100	This study
P1Δ*pmgB*	in-frame deletion of *pmgB* in P1CM*clr*100	This study
P1Δ*pmgC*	in-frame deletion of *pmgC* in P1CM*clr*100	This study
P1Δ*pmgF*	in-frame deletion of *pmgF* in P1CM*clr*100	This study
P1Δ*pmgG*	in-frame deletion of *pmgG* in P1CM*clr*100	This study
P1Δ*upfA*	in-frame deletion of *upfA* in P1CM*clr*100	This study
P1Δ*upfB*	in-frame deletion of *upfB* in P1CM*clr*100	This study
P1Δ*upfC*	in-frame deletion of *upfC* in P1CM*clr*100	This study

Table 3.1. Continued.

Strains, phages or plasmids	Genotype or relevant characteristic	Reference/source
P1Δ*uhr*	in-frame deletion of *uhr* in P1CM*clr*100	This study
P1Δ*upl*	in-frame deletion of *upl* in P1CM*clr*100	This study
P1Δ*dbn*	in-frame deletion of *dbn* in P1CM*clr*100	This study
P1Δ*pmgL*	in-frame deletion of *pmgL* in P1CM*clr*100	This study
P1Δ*pmgM*	in-frame deletion of *pmgM* in P1CM*clr*100	This study
P1Δ*pmgN*	in-frame deletion of *pmgN* in P1CM*clr*100	This study
P1Δ*pmgO*	in-frame deletion of *pmgO* in P1CM*clr*100	This study
P1Δ*pmgP*	in-frame deletion of *pmgP* in P1CM*clr*100	This study
P1Δ*ppp*	in-frame deletion of *ppp* in P1CM*clr*100	This study
P1Δ*pmgQ*	in-frame deletion of *pmgQ* in P1CM*clr*100	This study
P1Δ*pmgR*	in-frame deletion of *pmgR* in P1CM*clr*100	This study
P1Δ*pmgS*	in-frame deletion of *pmgS* in P1CM*clr*100	This study
P1Δ*pap*	in-frame deletion of *pap* in P1CM*clr*100	This study
P1Δ*pmgT*	in-frame deletion of *pmgT* in P1CM*clr*100	This study
P1Δ*pmgU*	in-frame deletion of *pmgU* in P1CM*clr*100	This study
P1Δ*pmgV*	in-frame deletion of *pmgV* in P1CM*clr*100	This study
P1Δ*upfM*	in-frame deletion of *upfM* in P1CM*clr*100	This study
P1Δ*upfN*	in-frame deletion of *upfN* in P1CM*clr*100	This study
P1Δ*upfO*	in-frame deletion of *upfO* in P1CM*clr*100	This study
P1Δ*pdcA*	in-frame deletion of *pdcA* in P1CM*clr*100	This study
P1Δ*pdcB*	in-frame deletion of *pdcB* in P1CM*clr*100	This study
Plasmids		
pCP20	FLP+, λ *c*I857^{+}, λ p_R Rep^{ts}, Ap^R, Cm^R	(Cherepanov & Wackernagel, 1995) / Zeng Lab
p*darB*	P1 *darB* cloned into pBAD24	(Piya *et al.*, 2017)
p*darB* 1-1750	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 1-1750	This study
p*darB* 90-1750	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 90-1750	This study
p*darB* 90-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 90-2255	This study

Table 3.1. Continued.

Strains, phages or plasmids	Genotype or relevant characteristic	Reference/source
p*darB* 5-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 5-2255	This study
p*darB* 10-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 10-2255	This study
p*darB* 20-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 15-2255	This study
p*darB* 25-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 20-2255	This study
p*darB* 25-2255	P1 *darB* fragment cloned into pBAD24 to express truncated DarB residues 25-2255	This study
pMB838	source of gentamicin resistance cassette	Addgene / Lab stock
pBAD24g	pBAD24 with gentamycin resistance cassette	This study
p*pro*	P1 *pro* cloned into pBAD24g	This study
p*pmgA*	P1 *pmgA* cloned into pBAD24g	This study
p*pmgB*	P1 *pmgB* cloned into pBAD24g	This study
p*pmgC*	P1 *pmgC* cloned into pBAD24g	This study
p*pmgG*	P1 *pmgG* cloned into pBAD24g	This study
p*pmgR*	P1 *pmgR* cloned into pBAD24g	This study
pETDuet-1	Protein expression vector	Millipore Sigma / Lab stock
pETDuet_*darB*	P1 *darB* cloned into pETDuet-1	This study
pETDuet_2x-His_*darB*	pETDuet_*darB* engineered to introduce additional N-terminal His-tag	This study
pBAD24_2x-His_*darB*	gene fragment encoding 2x-His_DarB cloned into pBAD24	This study
pFTSKi-*tetR-mCherry*	template for PCR amplification of *tetR-mCherry*	(Shao *et al*., 2017) / Zeng Lab
pBAD24_*darB*N9	Derivative of pBAD24 constructed to express recombinant proteins with N-terminal 9 residues to DarB fused to N-terminus of any protein	This study
pBAD24_*darB*N30	Derivative of pBAD24 constructed to express recombinant proteins with N-terminal 30 residues to DarB fused to N-terminus of any protein	This study
pBAD24_*darB*N9_*tetR-mCherry*	*tetR-mCherry* cloned into pBAD24_*darB*N9	This study
pBAD24_*darB*N30_*tetR-mCherry*	*tetR-mCherry* cloned into pBAD24_*darB*N30	This study

BW25113(pKD46). P1 lysogens were grown in LB cm + amp. At OD_{550} = 0.1, 1 mM L-arabinose was added to induce Red proteins and cells were cultured up to OD_{550} 0.5-0.6. At this point, the culture was harvested, and cells were made electrocompetent. Cells were transformed with gel purified DNA (QIAquick Gel Extraction Kit), which was obtained by PCR amplification of *Eco*P1^R *FRT*-flanked *kan* gene from plasmid pKD4. The primers to amplify *kan* gene were designed to include homology regions to P1 DNA and pKD4 priming sites, as described before. The transformed cells recovered in LB at 30 °C for 2 h. After overnight incubation at 30 °C, Cm^R and kan^R colonies were selected, and mutations were confirmed by PCR and sequencing (Piya *et al.*, 2017). The primers used to amplify *kan* inserts are listed in Table 3.2.

Flp-mediated excision of kan insert and generation of double-gene knockout mutants

All single-gene deletion mutants of P1 were generated by replacing respective genes with *kan* gene flanked with *FRT* (*F*lp *r*ecombination *t*arget) sites as described above. *Saccharomyces cerevisiae* encoded Flp enzyme mediates recombination between these *FRT* sites, enabling excision of the antibiotic resistance gene (Cherepanov & Wackernagel, 1995). *E. coli* MG1655 strain lysogenized with either P1Δ*darA* or P1Δ*darB* was made electrocompetent and transformed with plasmid pCP20 as previously described (Cherepanov & Wackernagel, 1995; Datsenko & Wanner, 2000). Lysogens of either P1Δ*darA* or P1Δ*darB* mutants are Cm^R and Kan^R, and plasmid pCP20 transformants are Cm^R and Amp^R. P1Δ*darA* or P1Δ*darB* lysogens transformed with pCP20 were selected as Cm^R and Amp^R colonies. The *FLP* gene can be induced

Table 3.2. Primers and synthetic DNA fragments

Primers		
Forward primer 5'-3'	**Reverse primer 5'-3'**	**Usage**
atcggacgtcatgttacgcagcag caacgatg	atcggagctcttaggtggcggt acttgggtc	To PCR amplify gentamicin resistance cassette from pMB838
atcggagctcctgtcagaccaagt ttactcatatatactttagattg	atcggacgtcactcttccttttt tcaatattattgaag	To PCR amplify pBAD24 backbone
atactgagctcaatgaacaagcta tctatgggggtgt	atactggcgcgccgtcacggtc aacgagacgac	To clone *darB* into pETDuet-1
tgagcggataacaattcccctcta g	aatttcgattatgcggccgtgt ac	To PCR amplify synthetic DNA fragments designed to introduce additional His-tag to pETDuet-1
atcgtctagagtaaaggaggtatc gatatgggcagcagccatcac	atcgaagcttgtcacggtcaac gagacgac	To clone *darB* fused with 2x His-tag from pETDuet_*darB*-12His into pBAD24
atcgtctagagtaataaaggaggt atcgagatgaacaagctatctatg ggggtg	atcgaagcttttatgagcgatc gccgataacaag	To clone *darB* into pBAD24 to express residues 1 - 1750
atcgtctagagtaataaaggaggt atcgagatggaatattacacacca aagccgatcg	atcgaagcttttatgagcgatc gccgataacaag	To clone *darB* into pBAD24 to express residues 90 - 1750
atcgtctagagtaataaaggaggt atcgagatggaatattacacacca aagccgatcg	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 90 - 2255
atcgtctagagtaataaaggaggt atcgagatgtctatgggggtgttt cgctg	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 5 - 2255
atcgtctagagtaataaaggaggt atcgagatgcgctgttcaagtgtc agcg	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 10 - 2255
atcgtctagagtaataaaggaggt atcgagatgagcgaaatattgaaa tacattagggc	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 15 - 2255
atcgtctagagtaataaaggaggt atcgagatgtacattagggcaata acatctcaccg	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 20 - 2255
atcgtctagagtaataaaggaggt atcgagatgacatctcaccgagcg cc	atcgaagcttttatgcgtattg ttggatgacggc	To clone *darB* into pBAD24 to express residues 25 - 2255
tttcgttaccctcgcgtcaaaaag agtttttacgaaaggaagcataag tggtgtaggctggagctgcttc	acgtgggtccgttacgcaactt ctggccattaatcatcatcccc gtacagatgggaattagccatg gtcc	To knockout *pacA* from P1 genome
tccgcacgtatctgattgattg	gtcttcccaaacaacacatcag c	Sequencing primers to verify *pacA* deletion in P1 genome

Table 3.2. Continued.

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
aaattccgttcaaacacgatgtga attattctaattaaggtgcaatct tggtgtaggctggagctgcttc	tcagtgaccgttttcaaaacat cagtcattatcgtttccctctt taaagaatgggaattagccatg gtcc	To knockout *prt* from P1 genome
ccgtatttaatcggcgctcg	ggcgaccgtgaccgtaatag	Sequencing primers to verify *prt* deletion in P1 genome
aaatccttgaatcggtctttaaag agggaaacgataatgactgatgtt ttgtgtaggctggagctgcttc	tttggtggttatttaaacggat tgattgaattattaaacgtgat gatgctatgggaattagccatg gtcc	To knockout *pro* from P1 genome
ttacctgttcagcgatgtgc	gctattgtatgcagggcttc	Sequencing primers to verify *pro* deletion in P1 genome
tcaattgatgatccgttgatgaat gactacgcgagagttgattgatgg ccgtgtaggctggagctgcttc	ctgtctttagtcggtgttgttt tactcatagcaccacgtcctgt gtgataatgggaattagccatg gtcc	To knockout *pmgA* from P1 genome
tgaaacgaagcggctgtttg	gatataaagccttcagccagc	Sequencing primers to verify *pmgA* deletion in P1 genome
gcttcccttctgggctttagtttt tctctttgaattaaggagcaagga tggtgtaggctggagctgcttc	ggaaaaacgcataggattcagt gagcctttcagcgtcctgaatc ggtataatgggaattagccatg gtcc	To knockout *pmgB* from P1 genome
cttccggcttaaccaaactg	tcccgaatacctgcaagctc	Sequencing primers to verify *pmgB* deletion in P1 genome
gtagaatcggttaacacaccagat tctacgaggtttcaatgacaccac gagtgtaggctggagctgcttc	aagcaacattaagccccatatc agccccctcacttcaacatgga tgagaaatgggaattagccatg gtcc	To knockout *pmgC* from P1 genome
aacacaacgcccattaaagg	tcaacttcgacagagaactgc	Sequencing primers to verify *pmgC* deletion in P1 genome
ttttttgtgcttaaatcatgtcaa tatagcgaaattttgagcatatta tggtgtaggctggagctgcttc	gataattccgggagagcggttg ggtttttcatgcttttctgcg cagagaatgggaattagccatg gtcc	To knockout *pmgF* from P1 genome
ttagtaccgaagatgtgactgctg	agctgaacccgacacatattgc	Sequencing primers to verify *pmgF* deletion in P1 genome
ggaactccggcctttaacttgaat ggctcctatagcttatggggttta cagtgtaggctggagctgcttc	agtgttattgtgccccatataa aatccttttactggaacgcccg aacaatatgggaattagccatg gtcc	To knockout *pmgG* from P1 genome
aggttgttcgtaacggtgag	ttcaggaatttgcgacgtgc	Sequencing primers to verify *pmgG* deletion in P1 genome

Table 3.2. Continued

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
gttaatgattacaaccgagctatt agcggtaactaaaagggattttta tggtgtaggctggagctgcttc	ccaacgtagccggttggaaggg agtcgtattattctacgctttc gatgagatgggaattagccatg gtcc	To knockout *upfA* from P1 genome
aatttaacgccgacgccgac	tttccctccagcacacatcg	Sequencing primers to verify *upfA* deletion in P1 genome
gagacggcctagttcaggtaagtt aggagatatcatgctggaaaaaga ctgtgtaggctggagctgcttc	tgaagaaaaattatcaatgaag tcctttgttactgtgccgcttt gtttaaatgggaattagccatg gtcc	To knockout *upfB* from P1 genome
tcccgtagagagcgaaccac	ggcttgctttagctgatgacg	Sequencing primers to verify *ufpB* deletion in P1 genome
aatctaagttaaaataacgaaaat cagagcaaatcattggtgatgacg tggtgtaggctggagctgcttc	cgtctccggctgcaacggtaga gccgcatgaaacaggatcacca acgcatatgggaattagccatg gtcc	To knockout *upfC* from P1 genome
tgcatctgacgactccaaac	gctggtaatagtgtcgggatag	Sequencing primers to verify *upfC* deletion in P1 genome
accattcagccatcgcccttcaat gggcatttgtttggagtcgtcaga tggtgtaggctggagctgcttc	atagtaatgtttttccagtgct taaacatgttgtaacccttgaa tatcaaatgggaattagccatg gtcc	To knockout *uhr* from P1 genome
atacttgccacgtcatcacc	tttgtttcgcgctttgcttc	Sequencing primers to verify *uhr* deletion in P1 genome
ttcaccgcaacggaaagagcattc ctggtggacctgtagattgggata tggtgtaggctggagctgcttc	tggtgaatgcacaggctgatgt gccgcaactacagtagtgcgcg ctttgcatgggaattagccatg gtcc	To knockout *upl* from P1 genome
gggaaataaggttgcggtgc	cgcaggaacgaccaataaacg	Sequencing primers to verify *upl* deletion in P1 genome
cggtttgataactatgccggtgct gactggcaagaggattattaatgc aagtgtaggctggagctgcttc	ggcttttctgttatgacgggtt caatttttatccgttaccgcg cgacggatgggaattagccatg gtcc	To knockout *dbn* from P1 genome
aggacgcagacctgattatg	gcattcatctcttccgcagg	Sequencing primers to verify *dbn* deletion in P1 genome
aaatgtatcattctgcccttaagt aggttcttcacgaggaaacaaaat tggtgtaggctggagctgcttc	aaactcttcctcagtaaagcgc aacataatcagtcaaatcctgc cggtcgatgggaattagccatg gtcc	To knockout *pmgL* from P1 genome
gtgctttgggttgtagtgattg	agtgctcgcaatttccgtcg	Sequencing primers to verify *pmgL* deletion in P1 genome

Table 3.2. Continued.

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
gcaaaacagcgtaaatatggcatg cgaccggcaggatttgactgatta tggtgtaggctggagctgcttc	agcgcacatgcactaattaatt tattattttaagcagcatacaa ccacttatgggaattagccatg gtcc	To knockout *pmgM* from P1 genome
cagaaagcaggcacaggtcg	aggcgagcattcatttcagc	Sequencing primers to verify *pmgM* deletion in P1 genome
catgtgcgctctttgtggtagtgc actttaacatcgggagaataatcg tggtgtaggctggagctgcttc	agctgcaaattcagcctccagg cgagcattcatttcagcgatta cagggtatgggaattagccatg gtcc	To knockout *pmgN* from P1 genome
gtgcagaacaagcgattacc	ggaatctccctcccggtatg	Sequencing primers to verify *pmgN* deletion in P1 genome
gctgaatttgcagctgagaatgaa cataccacccagggcgactaggac tggtgtaggctggagctgcttc	ctgcctctaaatttattcttgg gaaacattgaattgcactgctc ctacttatgggaattagccatg gtcc	To knockout *pmgO* from P1 genome
cgataagtggttgtatgctgc	gaagcacagaacgccatctc	Sequencing primers to verify *pmgO* deletion in P1 genome
gcggagtgccgattatagcggtca atggggccatcgactggcttaacc gcgtgtaggctggagctgcttc	gtgggtgctatcatgtattctt cagccattctttaagagtcatc tgcggaatgggaattagccatg gtcc	To knockout *pmgP* from P1 genome
cggctatgactgttctctgc	gcaatagttccagacattcgac	Sequencing primers to verify *pmgP* deletion in P1 genome
ggtattccgcagatgactcttaaa gaatggctgaagaatacatgatag cagtgtaggctggagctgcttc	ccagagcgaagaactaaagcaa tcttcatgctgcaccatcacct ttcactatgggaattagccatg gtcc	To knockout *ppp* from P1 genome
tgaacctggctttccatatcg	tatccggtcggaacaactcg	Sequencing primers to verify *ppp* deletion in P1 genome
ggaaacctgacattgatgaaagtg aaaggtgatggtgcagcatgaaga ttgtgtaggctggagctgcttc	tttaaaaatatggaattattag agcaatattattctgattctcg ctcaaaatgggaattagccatg gtcc	To knockout *pmgQ* from P1 genome
caccgggcaagaaactatcg	gcgatacctggtgatgacatg	Sequencing primers to verify *pmgQ* deletion in P1 genome
atgtacactcatcacgttttttat tagagcaatctacaaggtgcacta tggtgtaggctggagctgcttc	ttgattggttttgtgctgaagt gatcatttcaaagttccgtatt agcttgatgggaattagccatg gtcc	To knockout *pmgR* from P1 genome
ggccgatgttgtggaacctg	gcatctccgctgactgaacc	Sequencing primers to verify *pmgR* deletion in P1 genome

Table 3.2. Continued.

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
atacttaggaaaaatgaccgaagc acaagctaatacggaactttgaaa tggtgtaggctggagctgcttc	caccagcattaaaagtgacact gtaactatcagcgaacgtaaat agtgccatgggaattagccatg gtcc	To knockout *pmgS* from P1 genome
ggctgatcgcctttgttacg	atgcagcctgtcttccgttg	Sequencing primers to verify *pmgS* deletion in P1 genome
ggtggagtgcgcccaccagcattt ttttcgtccaatgaggagggcatt tggtgtaggctggagctgcttc	gttttaaaaatcaagatttatt agagcaattattgttgatgaag aagcgcatgggaattagccatg gtcc	To knockout *pap* from P1 genome
cgaattgggtttagaggaaacagc	tttcacttcgcctttcatgc	Sequencing primers to verify *pap* deletion in P1 genome
tctctatcacttttggcggcatcg tcgcgcatgggtaaggggggatgca tggtgtaggctggagctgcttc	aatgtccggatagttgccagcc tcgtaacccagggattagcact ccagctatgggaattagccatg gtcc	To knockout *pmgT* from P1 genome
acgttaatggtcagggagtatg	cgtgcgtctcatctgatctc	Sequencing primers to verify *pmgT* deletion in P1 genome
ttattacgccagatcatcataaac aagccgagaaaagtgacctagaaa gggtgtaggctggagctgcttc	aggtgtttgatgtgatctgcaa ctctcatacttcaccctcgctt gtatcgatgggaattagccatg gtcc	To knockout *pmgU* from P1 genome
ggcatcacgcattctgttgg	cccggttacgtctgttcttg	Sequencing primers to verify *pmgU* deletion in P1 genome
ttgccgttggtggatgctggcgat acaagcgagggtgaagtatgagag ttgtgtaggctggagctgcttc	ttctgtatgtgctggtgggtac ctgtagttcagctttcgttggc atttaaatgggaattagccatg gtcc	To knockout *pmgV* from P1 genome
cacccagcaattacgagaaag	cgcctgttattagccagattc	Sequencing primers to verify *pmgV* deletion in P1 genome
gatgaggataatagccagaatctg gctaataacaggcgcatctaaaaa tggtgtaggctggagctgcttc	tgccgtaagctcacgttaacga ctttctttcaccgaatccaact atataaatgggaattagccatg gtcc	To knockout *upfM* from P1 genome
gtacccaccagcacatacag	ttatctgccttgccgtgc	Sequencing primers to verify *upfM* deletion in P1 genome
aaacccaacccttatatagttgga ttcggtgaaagaaagtcgttaacg tggtgtaggctggagctgcttc	acaaagttatgcacttgcaaga gggccattttctaaatattgtg atgtttatgggaattagccatg gtcc	To knockout *upfN* from P1 genome
tggctaataacaggcgcatc	ccttatcccgttcttcctgac	Sequencing primers to verify *upfN* deletion in P1 genome

Table 3.2. Continued.

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
gaaacatcacaatatttagaaaatggccctcttgcaagtgcataacttggtgtaggctggagctgcttc	tcccatccagaataattgagtaacgactattatttaaccagcaagtaacatgggaattagccatggtcc	To knockout *upfO* from P1 genome
tggattcggtgaaagaaagtcg	gcagattctctggttgctcg	Sequencing primers to verify *upfO* deletion in P1 genome
tctgtcgccgacacgttacgtagattgtatggttctgcggagtagattaagtgtaggctggagctgcttc	acatcacacctttaatcactgattgggctttatctgctgcccggcattctatgggaattagccatggtcc	To knockout *pdcA* from P1 genome
aagtctccgccacctacctg	ccgtcttccatgtgctcgattaac	Sequencing primers to verify *pdcA* deletion in P1 genome
gaatgccgggcagcagataaagcccaatcagtgattaaaggtgtgatgtggtgtaggctggagctgcttc	taaagaaatagcaatacattagagcaattttatctaacactcgacgaatgatgggaattagccatggtcc	To knockout *pdcB* from P1 genome
gtcccgtgaagcgttaaagatg	agaagagatttagtgcggatcatg	Sequencing primers to verify *pdcB* deletion in P1 genome
tctagagtaaaggaggatcgatcttggcagacaataaaatcacgctatc	atcgaaagctttcagtcattatcgtttccctctttaaag	To clone *prt* into pBAD24g
atcgatctagagtaaaggaggatcgatcatgactgatgttttgaaaacggtcac	atcgaaagcttttatttaaacggattgattgaattattaaacgtgatg	To clone *pro* into pBAD24g
ggggatcctctagaataaaggaggtcagtacatggccaataataacgaaattgatcc	aaaacagccaagctttcatagcaccacgtcctg	To clone *pmgA* into pBAD24g
ggggatcctctagaataaaggaggatgctaatgcttttaccccttttccc	aaacagccaagctttcagcgtcctgaatcgg	To clone *pmgB* into pBAD24g
ggggatcctctagaataaaggaggatgctaatgacaccacgacaattactc	aaacagccaagctttcacttcaacatggatgagaaaag	To clone *pmgC* into pBAD24g
ggggatcctctagaataaaggaggatgctaatggctcctatagcttatggg	aaacagccaagctttactggaacgcccgaa	To clone *pmgG* into pBAD24g
ggggatcctctagaataaaggaggatgctaatgtggccattccgacg	aaacagccaagctttcaaagttccgtattagcttgtg	To clone *pmgR* into pBAD24g

Table 3.2. Continued

Forward primer 5'-3'	Reverse primer 5'-3'	Usage
actctctactgtttctccatacccg	tgtatcaggctgaaaatcttctctcatc	To PCR amplify synthetic DNA fragments designed to introduce N-terminal DarB residues into pBAD24
atcgagaattcgtgtctagattagataaaagtaaagtgattaacagcg	atcgaggtaccttacttgtacagctcgtccatgc	To clone *tetR-mCherry* into pBAD24_*darB*N9 and pBAD24_*darB*N30
Synthetic DNA fragments (5' - 3')		
aattgtgagcggataacaattcccctctagaaataattttgtttaacttaagaaggagatataccatgggcagcagccatcaccatcatcaccacggcagcagccatcaccatcatcaccacagccaggatccgaattcgagctcggcgcgcctgcaggtcgacaagcttgcggccgcataatgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatcgaaattaatacgactcactata		Synthetic DNA fragment to introduce additional His-tag to pETDuet-1
gacgctttttatcgcaactctctactgtttctccatacccgttttttgggctagcaggaggtaatacaccatgaacaagctatctatgggggtgttgaattcaccatggtacccggggatcctctagagtcgacctgcaggcatgcaagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatc		Synthetic DNA fragment to introduce 9 N-terminal DarB residues into pBAD24
gacgctttttatcgcaactctctactgtttctccatacccgttttttgggctagcaggaggtaatacaccatgaacaagctatctatgggggtgttttcgctgttcaagtgtcagcgaaatattgaaatacattagggcaataacatctcaccgagcgccggaattcaccatggtacccggggatcctctagagtcgacctgcaggcatgcaagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatc		Synthetic DNA fragment to introduce 30 N-terminal DarB residues into pBAD24

from pCP20 by thermal induction since the gene is under control of λ *c*I857 repressor (Cherepanov & Wackernagel, 1995). However, thermal induction of Flp was not desired in our system as this would induce lysogenized P1 mutants as well. It was reported that basal level of Flp induction was sufficient to excise *FRT*-flanked regions, hence this strategy was applied (Cherepanov & Wackernagel, 1995). Lysogens of P1 mutants, transformed with pCP20 were grown overnight in LB amp + cm at 30 °C. The overnight culture was streaked on LB amp + cm plates to get isolated colonies, from which Kan^S colonies were selected from patch plating. To ensure that the selected Kan^S colonies were lysogens of P1 mutants, virions production by thermal induction was tested. The induced P1 mutants were lysogenized into *E. coli* MG1655 for maintenance of the strains, as previously described and selected as Cm^R colonies (Piya *et al.*, 2017). The excision of *kan* gene was further verified by sequencing.

In order to knock out additional genes, P1 mutants were lysogenized into *E. coli* BW25113(pKD46) and selected as Cm^R and Amp^R colonies; additional gene was disrupted with *FRT*-flanked *kan* gene, as described above.

Efficiency of plating (EOP) assay

All phages used in EOP assays were lysogenized into *E. coli* WA921 and phages were induced, as previously described (Piya *et al.*, 2017). EOP assays were conducted following previously established protocol, with minor modifications (Piya *et al.*, 2017).

Phages were diluted 10-fold in SM buffer and 10 μL from each dilution was spotted on the soft agar overlay of the host strains.

Plasmid construction

The plasmid pBAD24 was engineered to provide gentamicin resistance as described below. The gentamicin resistance cassette was PCR amplified from plasmid pMB838 with primers designed to have *Aat*II and *Sac*I restriction sites on 5' end of forward and reverse primers, respectively. Similarly, the backbone of plasmid pBAD24 was PCR amplified with primers designed to have *Sac*I and *Aat*II restriction sites on 5' end of forward and reverse primers, respectively. All PCR reactions were conducted using Phusion Hi-Fidelity PCR Master Mix (New England Biolabs), following the manufacturer's recommended protocol. The gentamicin resistance cassette was ligated into pBAD24 backbone using standard molecular biology techniques. P1 genes *prt*, *pro*, *pmgA*, *pmgB*, *pmgC*, *pmgG* and *pmgR* were cloned into pBAD24g (*g* for *g*entamicin resistance). These genes were PCR amplified using primers with *Xba*I and *Hind*III restriction sites on 5' end and were ligated into respective restriction sites in pBAD24g following standard molecular biology techniques.

P1 *darB* was cloned into pETDuet-1 (Millipore Sigma) for expression and subsequent purification of recombinant DarB by immobilized metal affinity chromatography (IMAC). The plasmid pETDuet_*darB* was further engineered to express DarB with 2x N-terminal His-tag. Synthetic DNA (Integrated DNA Technologies) with regions of

pETDuet-1 vector from *lac* operator (downstream of T7 transcription start-1) to T7 promoter-2 was obtained, with some modifications. An additional His-tag was introduced downstream of the native His-tag, separated by a linker Gly-Ser-Ser. The introduced His-tag was followed by Ser-Gln to maintain similarity with the native reading frame of pETDuet-1 backbone. Other regions of the pETDuet-1 backbone was not altered. The PCR amplified synthetic DNA fragment was ligated into *Xba*I and *Sac*I sites in pETDuet_*darB* following standard molecular biology techniques. To test for the function of recombinant DarB *in vivo,* DNA fragment encoding DarB with N-terminal 2x His-tag was amplified from pETDuet_2x-His_*darB* with primers designed to have *Xba*I and *Hind*III restriction sites on 5' end and ligated into pBAD24 following standard molecular biology techniques.

To determine capsid targeting sequence in DarB, primers were designed with homology to different regions of P1 *darB* so that several N- and C-terminal residues would be truncated in the expressed mutant DarB. For constructing plasmid expressing DarB with N-terminal truncations, the start codon ATG was included in the forward primers, whereas for C-terminal truncations, the stop codon TAA was included in the reverse primers. The 5' end of forward and reverse primers were designed to include *Xba*I and *Hind*III restriction sites. The amplified PCR product was ligated into *Xba*I and *Hind*III sites in pBAD24 following standard molecular biology techniques.

In order to determine if the selected N-terminal residues of DarB were sufficient to target foreign proteins to P1 capsids, two derivatives of plasmid pBAD24 were constructed. Synthetic DNA fragments that encompassed regions of plasmid pBAD24 from 50 bp upstream of *Nhe*I restriction site to 50 bp downstream of *Hind*III restriction site, with some modifications, were ordered. The RBS was modified to "aggaggt", followed by 8 arbitrary nucleotides. The nucleotides encoding the N-terminal 9 or 30 residues of DarB (including the first residue Met) were added downstream of the arbitrary nucleotides. The pBAD24 multiple cloning sites from *Eco*RI to *Hind*III were added downstream of *darB* specific nucleotides. Both synthetic DNA fragments were PCR amplified and ligated into *Nhe*I and *Hind*III restriction sites of pBAD24 using standard molecular biology techniques. The two pBAD24 derivatives, pBAD24_*darB*N9 and pBAD24_*darB*N30 can be used to clone any genes to express recombinant proteins with selected DarB residues fused to the N-terminus. In this study, *tetR-mCherry* was PCR amplified with primers designed to have *Eco*RI and *Kpn*I restriction sites on the 5' end and ligated into respective restriction sites of pBAD24_*darB*N9 and pBAD24_*darB*N30. Primers and synthetic DNA used in this study are listed in Table 3.2.

Determination of capsid targeting sequence of DarB

Plasmids expressing proteins of interest were transformed into electrocompetent *E. coli* strain WA921 and transformed cells were selected as Amp^R colonies after overnight incubation at 37 °C. The overnight culture of Amp^R colonies were lysogenized with

P1Δ*darB* and lysogens were selected on LB amp + cam + kan, as previously described (Piya *et al.*, 2017). Phage lysates were produced from these P1Δ*darB* lysogens as previously described, with minor modifications (Piya *et al.*, 2017). Lysogenic strains were grown at 30 °C in LB amp + cam + kan to OD_{550} 0.4-0.6 and P1Δ*darB* was thermally induced at 42 °C. 1 mM L-arabinose (Sigma-Aldrich) was also added at the time of thermal induction. Samples were collected 30 min post-induction and analyzed by SDS-PAGE to check the expression of proteins from the plasmids. The lysates were harvested as described previously (Piya *et al.*, 2017). The phages were purified by CsCl isopycnic centrifugation and the packaging of proteins into capsids were analyzed by SDS-PAGE of virions.

Purification of virions by CsCl isopycnic centrifugation

Phages were thermally induced in 500 mL LB supplemented with appropriate antibiotics. Phages were concentrated 100-fold by centrifugation and purified by CsCl isopycnic centrifugation as described (Piya *et al.*, 2017).

Size exclusion chromatography of P1 procapsids

P1 mutants disrupted for DNA packaging (P1Δ*pacA*) were thermally induced from lysogens and lysates were harvested as described before. Protein components were precipitated from lysates by addition of ammonium sulfate to a final saturation of 60% (Simpson, 2006). The pellet was resuspended in SM buffer, passed through 0.22 μm filter (Millipore), followed by size exclusion chromatography on Superdex 200 Increase

100/300 GL (GE Healthcare). The void volume fractions, comprising P1 proheads and tails, were collected and analyzed by SDS-PAGE.

IMAC protein purification

E. coli BL21(DE3) cells were transformed with plasmid pETDuet_2x-His_*darB* as previously described (Green & Sambrook, 2012; Piya *et al.*, 2017) with minor modifications. BL21(DE3) cells were concentrated 10x while making them electrocompetent and were transformed with ~2 ng of pETDuet_*darB*-12His. Cells were recovered in 1mL LB for 1 h at 37 °C and plated to LB amp overnight at 37 °C. Single amp^R colony was selected and grown overnight on LB amp. A fresh culture was started in LB amp by 100-fold dilution of the overnight culture. At ~ OD_{550} 0.6-0.8, culture was placed on ice for ~15 mins, 1mM IPTG was added to induce protein synthesis, followed by overnight incubation at 16 °C. Culture was harvested by centrifugation at 5000 x g for 20 mins and washed with cold LB. After additional centrifugation at 5000 x g for 20 mins, cell pellets were stored in -80 °C until further processing.

Cell pellet from 1 L IPTG-induced BL21(DE3) culture was thawed in ice and resuspended in ~30 mL of sonication buffer [50 mM Tris-Cl pH 8.0, 500 mM NaCl, 15% (v/v) glycerol, 10 mM imidazole], followed by addition of 1 mg mL^{-1} chicken egg white lysozyme (Sigma-Aldrich), 100 μg mL^{-1} bovine pancreas DNase I (Sigma-Aldrich), SIGMA*FAST* protease inhibitor (Sigma-Aldrich), 1mM $MgSO_4$ and 0.01% Triton X-100 (Sigma-Aldrich). The resuspended cells were sonicated to lyse the cells,

followed by centrifugation for 15 min at 10000 x g at 4°C. The supernatant fraction that contained soluble recombinant DarB was mixed with 5 mL bed volume of pre-equilibrated Ni-NTA agarose resin (Qiagen) in a shaking platform for ~16 h at 4 °C. DarB bound to Ni-NTA agarose resin was purified by gravity-flow chromatography. Non-specific proteins bound to resin was eluted with IMAC wash buffer (50 mM Sodium phosphate pH 7.4, 500 mM NaCl) with gradient concentration of imidazole (20 mM, 50 mM, 60 mM, 70 mM, 80 mM, 100 mM). The target protein was eluted with elution buffer (50 mM Sodium phosphate pH 7.5, 500 mM NaCl, 500 mM imidazole). The elution fraction was concentrated, and buffer exchanged (50 mM Sodium phosphate pH 7.4, 150 mM NaCl) in 50 kDa molecular weight cutoff centrifugal filters (Amicon). The concentrated protein was stored in 20% glycerol (v/v) at -20°C until further use (Green & Sambrook, 2012). Samples were collected at each step for SDS-PAGE analysis.

SDS-PAGE

Phage samples for SDS-PAGE analysis were prepared as previously described (Piya *et al.*, 2017). For other protein samples, DNase I treatment was omitted. The gels were stained with either SYPRO Ruby (Thermo Scientific) or Coomassie (GE Healthcare).

Transmission electron microscopy

Samples for electron microscopy were stained with 2% uranyl acetate and imaged in a JEOL 1200 EX transmission microscope under 100 kV accelerating voltage in the

Microscopy and Imaging Center, Texas A&M University, College Station (Valentine *et al.*, 1968). Crude lysates of P1 procapsids were concentrated 10-fold in 50 kDa MWCO centrifugal filters (Amicon) for imaging.

Results and discussion

P1 genes of unknown function are not involved in antirestriction system

The first reported antirestriction phenotype of phage P1 demonstrated that two P1 genes *darA* and *darB,* comprised the P1 antirestriction system (Iida *et al.*, 1987). It has been shown that *darA* and *darB* are present in two distinct operons in P1 genome (Lobocka *et al.*, 2004). Recent work has shown that along with *darA* and *darB*, other genes present in the operons encoding these two genes, *hdf*, *ddrA*, *ddrB* and *ulx* are associated with the P1 antirestriction system (Fig. 3.1) (Piya *et al.*, 2017). Moreover, *hdf* and *darA* have been depicted to be involved in determination of head size during virion morphogenesis (Piya *et al.*, 2017).

Since the genes encoding P1 antirestriction system were present in two different operons in P1 genome, it was hypothesized that genes elsewhere in the P1 genome could be involved in providing protection against type I R-M systems. The identity of the antirestriction protein protecting P1 DNA from the type I *Eco*A system is still not clear. Moreover, proteins Hdf and DarA from the P1 antirestriction system were also found to affect capsid morphogenesis (Piya *et al.*, 2017). P1 capsid morphogenesis has not been well studied and the proteins forming the scaffolding core have not been identified

(Lobocka *et al.*, 2004; Piya *et al.*, 2017). Despite the role of Hdf and DarA in determining normal P1 capsid size, they do not fit the classical definition of scaffolding proteins, as both Hdf and DarA are still present in the mature virions (Piya *et al.*, 2017). We attempted to study the role of other P1 genes in antirestriction system and virion morphogenesis. There are several genes in P1 genomes for which any definitive function has not been assigned. The genes that were suggested to have unknown function or putative morphogenetic functions were of interest to us (Lobocka *et al.*, 2004). Isogenic deletions of 29 such genes in P1 genome were made and the role of each gene was interrogated in relation to P1 antirestriction system and virion morphogenesis (Fig. 3.1).

Successful isogenic deletions of these 29 genes in P1 lysogens suggest that these genes do not affect stability of P1 lysogens. P1 mutants deleted for genes *pmgA, pmgB, pmgC, pmgG* or *pmgR* did not form plaques on lawns of *E. coli* host, suggesting these genes are essential for lifecycle of P1. The essentiality of these five genes has not been verified by *in trans* complementation.

The remaining 24 gene knockout mutants of P1 were able to form plaques on lawns of *E. coli* host. These P1 mutants were lysogenized into *E. coli* WA921 strain and phage lysogens were induced to check for restriction phenotype. It has been established that the EOP of P1Δ*ddrA* is severely reduced in cells with type I *Eco*A, *Eco*B or *Eco*K systems as virions P1Δ*ddrA* mutants are unable to incorporate antirestriction proteins in capsids during virion morphogenesis (Piya *et al.*, 2017). The EOP of all P1 mutants were

compared to the EOP of P1Δ*ddrA* to study roles played by those genes in antirestriction. In this study, the EOP of P1Δ*ddrA* was reduced by ~10^{-3}, 10^{-4} or 10^{-4} in *E. coli* strains with type I *Eco*A, *Eco*B or *Eco*K systems respectively. The EOP of all other phage mutants appeared ~1 in *E. coli* strain with type I *Eco*A system and ~10^{-1} - 10^{-2} in *E. coli* strains with type I *Eco*B or *Eco*K systems (Fig. 3.2). This suggests that the antirestriction system of these mutants were not compromised and these phage mutants were able to employ antirestriction system to protect DNA from type I R-M system mediated DNA cleavage following infection of the respective host.

Dar proteins are packaged into P1 capsids before DNA

The Dar antirestriction system of P1 is different from other studied antirestriction systems as Dar proteins act strictly *in cis* (Iida *et al.*, 1987). These Dar proteins are incorporated into capsids during virion morphogenesis and at least some proteins must be delivered into the host cytoplasm during phage infection, where they protect DNA from host restriction (Iida *et al.*, 1987; Piya *et al.*, 2017). Recent work has revealed that Dar antirestriction proteins are incorporated into P1 capsids in a specific order during virion morphogenesis (Piya *et al.*, 2017). The assembly pathway of most phages in *Caudovirales* order bear similarity in that assembly starts with prohead, which is eventually filled with genetic material (Aksyuk & Rossmann, 2011). Phage P1 is not unique in packaging proteins into the head which are later transferred to the host cytoplasm on infection. Phage T4 also has several internal proteins packaged into procapsids during capsid morphogenesis. These internal proteins of T4 are incorporated

into scaffolding core, followed by procapsid shell formation and DNA packaging (Hong & Black, 1993; Leiman *et al.*, 2003). In double-stranded DNA phages, DNA is packaged

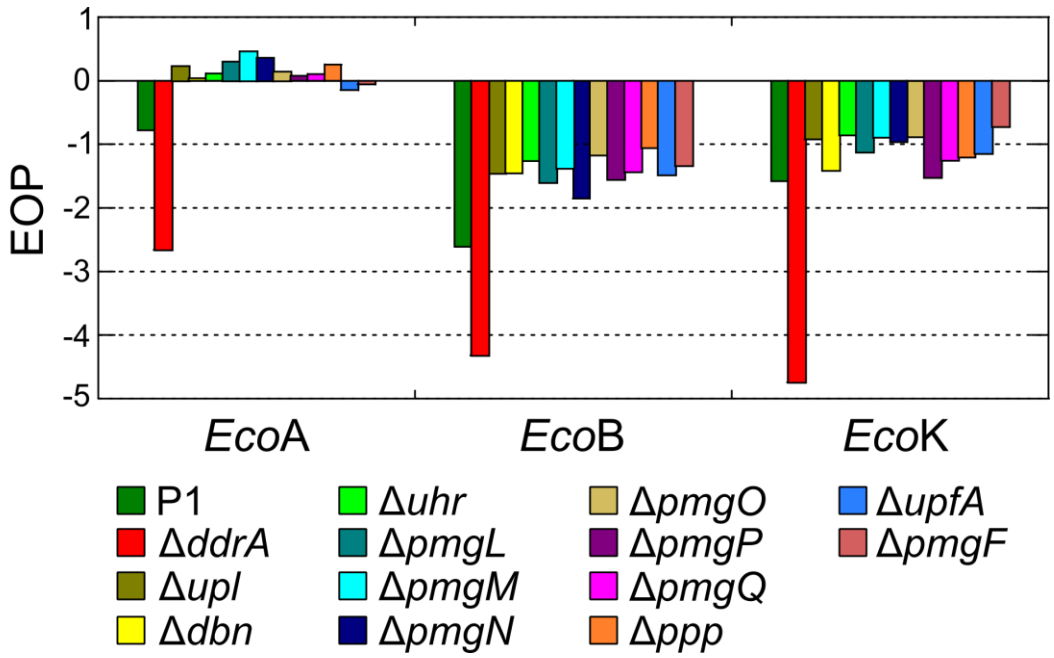


Fig. 3.2. P1 genes of unknown function are not required for antirestriction. Phage P1 or mutants were lysogenized into restriction-modification deficient *E. coli* strain WA921. The lysogens were thermally induced and tested for plating efficiency in *E. coli* strains with type I *Eco*A, *Eco*B or *Eco*K R-M systems. The EOP of parental P1, with functional antirestriction system is ~10^{-1}, ~10^{-3} or 10^{-2} in strains with type I *Eco*A, *Eco*B or *Eco*K systems respectively, whereas the EOP of P1Δ*ddrA*, with disrupted antirestriction system, is ~10^{-3}, ~10^{-4} or ~10^{-4}-10^{-5} in strains with type I *Eco*A, *Eco*B or *Eco*K systems respectively. The EOP of all other P1 mutants generated in this study appeared normal compared to the EOP of P1Δ*ddrA*, suggesting that these genes do not play any roles in protecting P1 DNA from the type I R-M systems tested in this study.

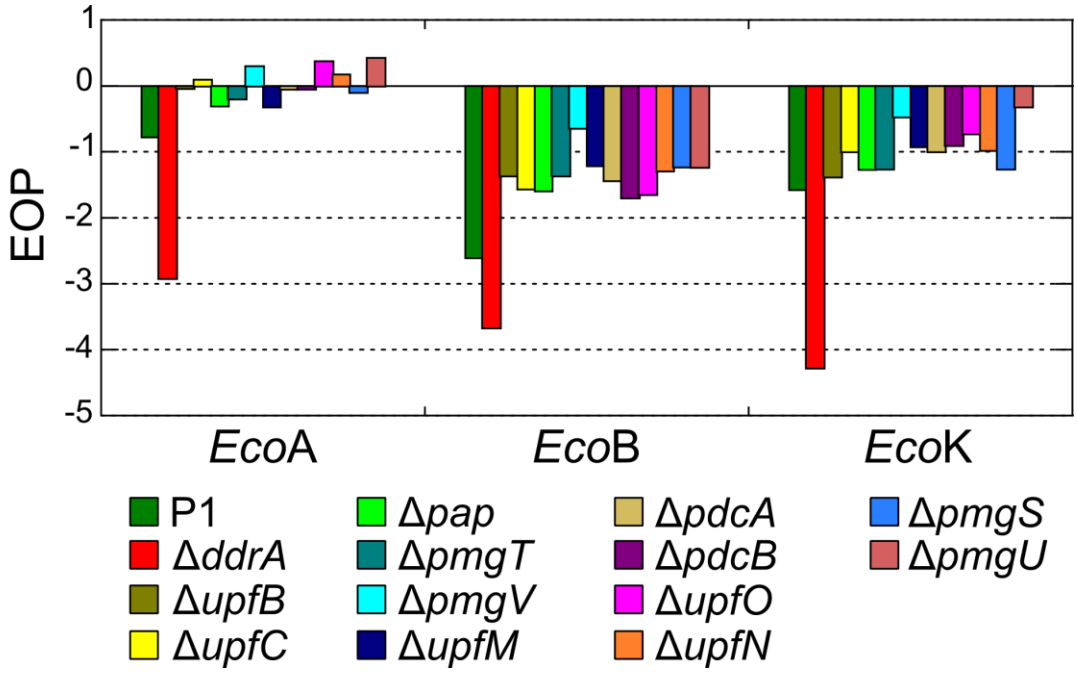


Fig. 3.2. Continued.

into procapsids by a packaging complex, composed of small (TerS) and large (TerL) subunits (Aksyuk & Rossmann, 2011). The DNA packaging complex of P1, which is composed of proteins encoded by *pacA* and *pacB*, cleave P1 DNA concatamers at *pac* sites located within the coding region of *pacA* (Fig. 3.3A) (Skorupski *et al.*, 1992). In the absence of either PacA or PacB, DNA cannot be packaged into P1 procapsids and virion morphogenesis is stalled at the procapsid stage (Skorupski *et al.*, 1992).

In order to determine if the antirestriction proteins are incorporated into P1 procapsids before DNA, *pacA* was deleted from P1 prophage. These Δ*pacA* prophages were then induced and procapsids were purified by ammonium sulfate precipitation followed by size exclusion chromatography as described above in materials and methods. These methods are suitable for purification of macromolecular complexes, hence both procapsids and tails are purified in these fractions. These procapsid/tail fractions as well as CsCl purified intact parental P1 virions were then analyzed by SDS-PAGE (Fig. 3.3B). Previously identified virion-associated Dar proteins, DarB, DarA, DdrB and Hdf, have been labelled. Other proteins, Ulx and DdrA, which are also associated with the Dar antirestriction system are not seen in this SDS-PAGE because of their small size and low copy number. It has been established that DarB can be incorporated into virions efficiently only in the presence of all other Dar proteins. Thus, the presence of DarB is indicative of the presence of other Dar proteins in the capsids. In the lanes of both CsCl

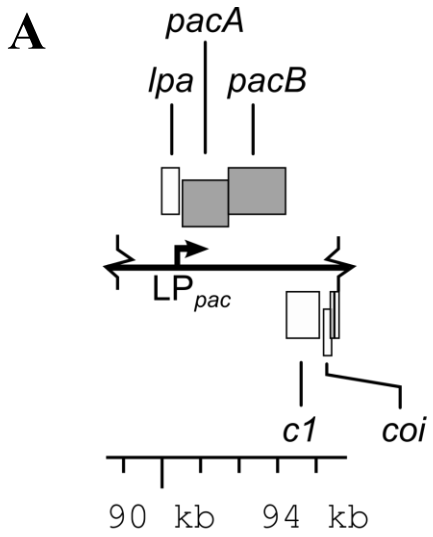


Fig. 3.3. P1 antirestriction components are incorporated into P1 procapsids before DNA. Panel A: The image represents a partial genomic map of P1 showing genes *pacA* and *pacB*, required for DNA packaging during virion morphogenesis. Both *pacA* and *pacB* are transcribed rightwards. The ruler at the bottom represents the nucleotide position as in the published P1 genome (NC_005856.1). The gene *pacA* was deleted from P1 genome in this study. **Panel B:** Large macromolecular structures present in thermally induced lysates of P1Δ*pacA* and P1Δ*darA*_Δ*pacA* were concentrated by ammonium sulfate precipitation and fractionated by size exclusion chromatography. The void volume from size exclusion chromatography, consisting of P1 procapsids and tails, was collected and analyzed by 12% Tris-glycine SDS-PAGE. P1 virions, purified by cesium chloride isopycnic centrifugation, were also run in SDS-PAGE for comparison. The protein bands corresponding to the components of P1 antirestriction system seen in the SDS-PAGE have been annotated. Presence of proteins bands corresponding to DarB, DarA, DdrB and Hdf in the lane of P1Δ*pacA* indicates that these antirestriction components are packaged into P1 procapsids before DNA. Other protein bands appearing in the procapsid lanes are denoted by arrows.

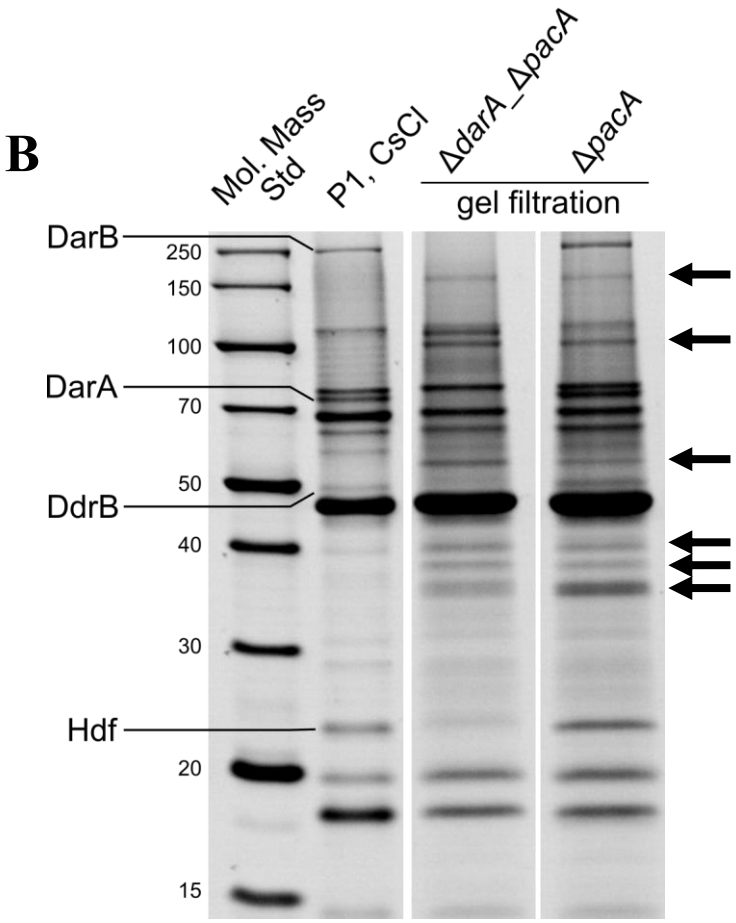


Fig. 3.3. Continued.

purified intact P1 virions and P1Δ*pacA* fractions, the band corresponding to DarB is visible. However, it is possible that this visible protein band of DarB in P1Δ*pacA* lane is not associated with the procapsids, but co-purified from the lysates at the applied ammonium sulfate concentrations (Wingfield, 2001). In order to rule out this possibility, procapsid/tail components of P1Δ*darA*_Δ*pacA* were also purified from crude lysates and were analyzed simultaneously (Fig. 3.3B). P1Δ*darA* virions are unable to package DarB, thus it is expected that P1Δ*darA*_Δ*pacA* procapsids also cannot package DarB (Piya *et al.*, 2017). However, the *darB* gene is intact in the genome of the P1Δ*darA* prophage, thus DarB is synthesized normally when lysogens of P1Δ*darA* are thermally induced, despite DarB not being incorporated into capsids (Fig. 3.4). A protein band corresponding to DarB is not visible in the lane of P1Δ*darA*_Δ*pacA*, indicating that DarB proteins are not incorporated into procapsids, and also are not purified along with procapsid/tail fractions by this technique despite being present in lysates. As mentioned above, in the absence of PacA, virion morphogenesis is stalled at the procapsid stage as the DNA packaging event is disrupted. Hence, comparative analysis of protein bands of P1Δ*pacA* and P1Δ*darA*_Δ*pacA* mutants suggest that DarB and thus other antirestriction proteins are packaged into procapsids before DNA.

Electron densities in transmission electron micrographs of P1 procapsids suggest presence of other internal proteins

In negative-stained transmission electron micrographs (TEM) of P1Δ*pacA* lysates, empty procapsid structures were visible (Fig. 3.5). It was interesting to observe internal

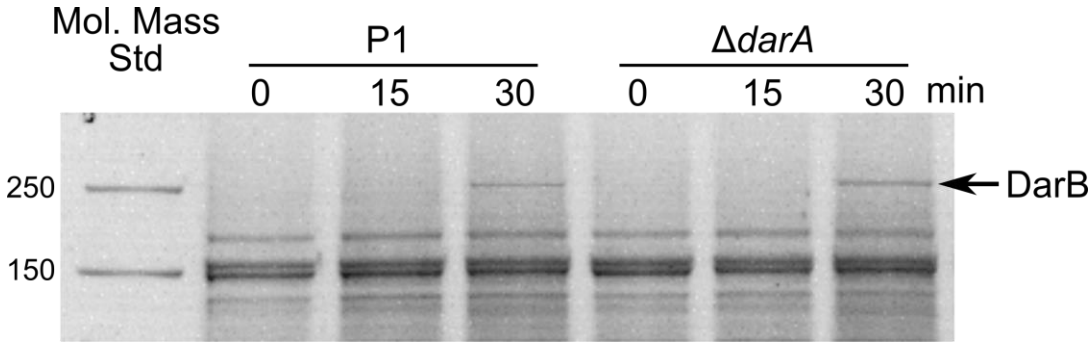


Fig. 3.4. DarB is synthesized in P1Δ*darA* mutants. E. coli MG1655 lysogens of P1 and P1Δ*darA* were cultured in 30 °C up to OD_{550} 0.4-0.5 and thermally induced at 42 °C. Samples were collected before and at 15, and 30 min post-thermal induction. 0.3 OD equivalent of samples were loaded in 8% Tris-glycine SDS-PAGE and gel was stained with SyproRuby.

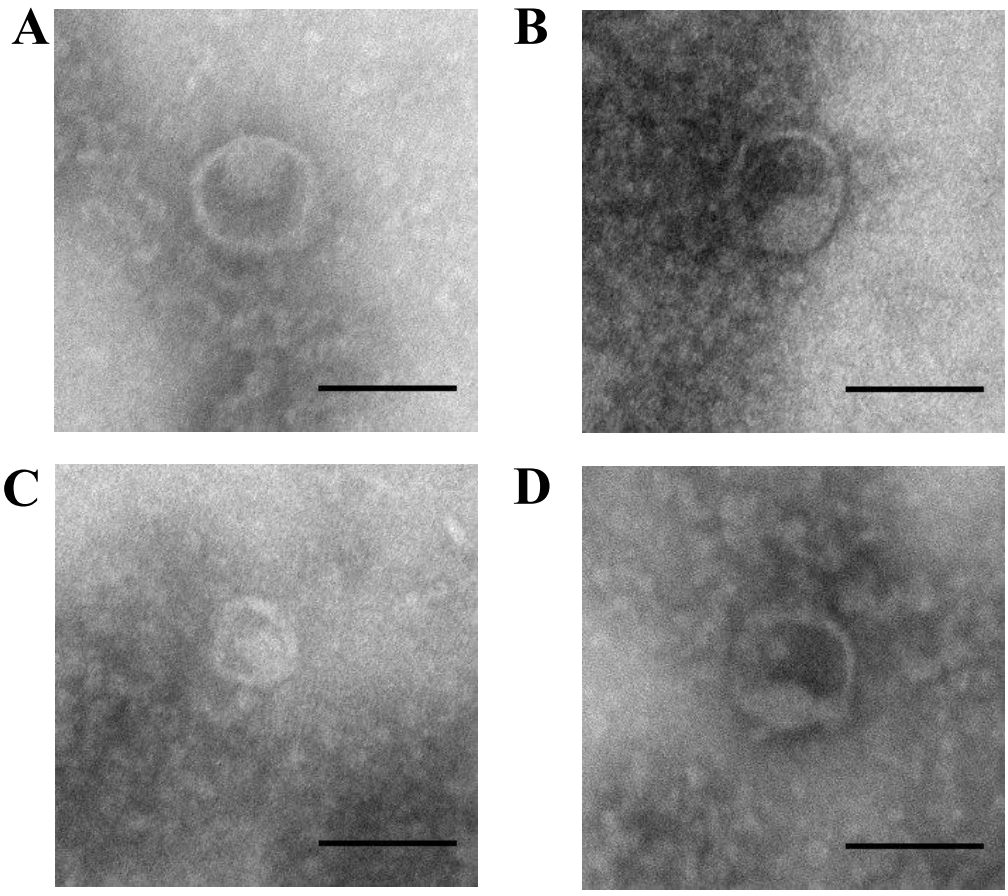


Fig. 3.5. P1 procapsids show electron densities suggesting presence of protein contents. Images shown are representative negative-stained transmission electron micrographs of procapsids of P1 (A), P1Δ*darA* (B and C) and P1Δ*darB* (D), generated by deletion of DNA packaging component *pacA* from P1 genome. Since these mutants are defective in DNA packaging, the presence of electron density can be corroborated to presence of proteins. Presence of electron density in P1Δ*darA* (B and C) and P1Δ*darB* (D) indicate these densities correspond to proteins other than Dar proteins. These electron densities are also seen in both normal (B) and small (C) size class procapsids of P1Δ*darA.* The scale bar represents 100 nm.

electron densities inside these protein shells. In a complete virion, this electron density inside capsid suggests packaged DNA (Chapter II, Fig 2.8). However, P1Δ*pacA* is defective in DNA packaging, thus it is extremely unlikely that this density is from the packaged DNA. It has been demonstrated that P1 Dar proteins are localized into P1 capsids and contribute significantly to the mass of the virion (Iida *et al.*, 1998; Piya *et al.*, 2017). Hence, we speculated that these densities correspond to the internal Dar proteins. However, the electron densities were still conspicuous in the negative-stained micrographs of both P1Δ*darB*_Δ*pacA* (Fig. 3.5D) and P1Δ*darA*_Δ*pacA* (Fig. 3.5B and 3.5C). P1Δ*darB* virions are missing DarB and Ulx, whereas P1Δ*darA* virions are missing all known proteins of P1 Dar system (Piya *et al.*, 2017). Thus, it can be concluded, from SDS-PAGE and TEM analysis of P1Δ*darA*_Δ*pacA* procapsids, that the electron densities visible in P1 procapsids do not correspond to the antirestriction proteins. Negative-stained TEM images of phage P22 and T4 procapsids also show similar electron density inside the capsid shells that correspond to scaffolding protein core (King *et al.*, 1973; Traub & Maeder, 1984; Black & Rao, 2012). Thus, it is possible that the density observed in P1 procapsids also correspond to the scaffolding core, which might be removed during DNA packaging (Fig. 3.5).

Comparative SDS-PAGE analysis protein bands of P1Δ*pacA* and P1Δ*darA*_Δ*pacA* show that the lanes of these procapsids contain protein bands (marked with arrows) that are absent in the lanes of CsCl purified intact P1 virions (Fig. 3.3B). Procapsids of other phages such as P22 and T4 contain scaffolding proteins that are either recycled or

cleaved and are not present in mature virions (Dokland, 1999). These protein bands in P1 procapsid lanes could be the constituents of the P1 scaffolding core. Because of the limitations of the technique applied in this study to obtain procapsids, it is also possible that some protein bands in the procapsids lanes of Fig. 3.3B are contaminating host proteins, thus the identity of these proteins need to be confirmed by mass spectrometry.

N-terminal residues of DarB provides signal for capsid targeting

Phages such as T4 and P22 incorporate internal proteins into capsids during early stages of capsid morphogenesis (Hong & Black, 1993; Jin *et al.*, 2015). The capsid-associated phage T4 internal protein III (IPIII) is directed to the assembly core by specific N-terminal residues, designated as capsid targeting sequence (CTS) (Showe & Black, 1973; Mullaney & Black, 1996). Since DarB is also packaged into the P1 capsid, it is possible that DarB has a CTS that directs nascent DarB to P1 procapsids. To determine if DarB is directed to capsid assembly by any specific residues, mutant DarB truncated at the N- and/or C-terminus were expressed from a plasmid-based system. Full-length DarB consists of 2255 residues, with a mass of ~250 kDa. Initially, three versions of truncations were constructed: a C-terminal truncation, with DarB containing residues 1-1750 ($DarB_t$ 1-1750), an N-terminal truncation, with DarB containing residues 90-2255 ($DarB_t$ 90-2255) and both N- and C-terminal truncations, with DarB containing residues 90-1750 ($DarB_t$ 90-1750). The truncated DarB was induced in P1Δ*darB* lysogen and the virions were tested for restoration of antirestriction phenotype and for incorporation of

truncated DarB. Two conserved domains of DarB were kept intact while constructing all truncated versions.

EOP assays shows that none of these truncated DarB restored the antirestriction phenotype of P1Δ*darB* mutant (Fig. 3.6A). In order to determine if the truncated DarB was expressed, whole cell samples of induced lysogens were analyzed via SDS-PAGE (Fig. 3.7A). DarB expressed from parental P1 is not visible because of the sensitivity of the staining procedure. However, the presence of DarB in P1virions is evident from the normal EOP of P1 in strains with type I *Eco*B or *Eco*K (data not shown). Protein bands corresponding to expressed protein were visible for all truncations suggesting that DarB truncations were expressed at a higher level from plasmid-based system compared to *wt* DarB expressed from its native locus in P1.

To determine if any of these DarB truncations were packaged into virions, P1Δ*darB* virions complemented *in trans* with DarB truncations were purified by CsCl isopycnic gradient centrifugation and analyzed by SDS-PAGE. Of the three truncations, protein bands corresponding to only $DarB_t$ 1-1750 was visible in the SDS-PAGE of virions. Protein bands corresponding to two other truncations, $DarB_t$ 90-1750 and $DarB_t$ 90-2255, were not visible. This suggests the N-terminal ~89 residues of DarB are required for targeting DarB into capsid during capsid morphogenesis (Fig. 3.8A). However, since $DarB_t$ 1-1750 cannot restore the antirestriction phenotype of P1Δ*darB* virions, it is

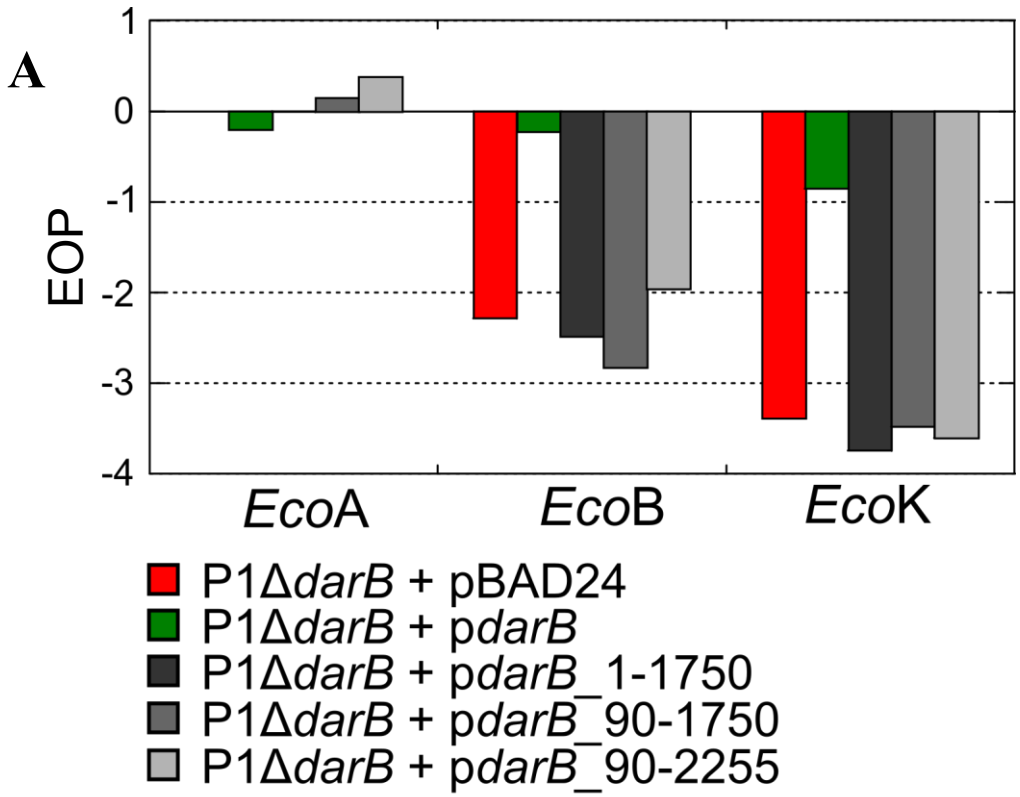


Fig. 3.6. Plating efficiency of P1Δ*darB* complemented with truncated DarB. In order to test for packaging signal in DarB, several N- and C-terminal truncations of DarB were constructed and tested for their ability to complement restriction phenotype. The empty vector pBAD24 or pBAD24 cloned with *darB* fragments were transformed into *E. coli* strain WA921 and lysogenized with P1Δ*darB*. The phage lysates were obtained as described in materials and methods and were tested for restoration of plating efficiency on *E. coli* strains with type I *Eco*A, *Eco*B or *Eco*K R-M systems. The EOP data shown have been normalized to P1 induced from strain WA921 transformed with plasmid pBAD24. **Panel A:** The relative EOP of P1Δ*darB* was reduced by ~10^{-2} and 10^{-3} in *Eco*B and *Eco*K strains respectively (red bars). The EOP was restored to normal levels by full length DarB (green bars). None of the truncated DarB was able to restore the EOP (grey bars). **Panel B:** As described in panel A, the EOP of P1Δ*darB* was reduced in strains with *Eco*B and *Eco*K strains (red bar) and the EOP was restored to normal levels by full length DarB (dark green bar). N-terminal truncation of DarB containing residues 5-2255 was able to restore plating deficiency in *Eco*B and *Eco*K strains (light green bar). None of the other truncations restored plating deficiency in *Eco*B and *Eco*K strains (grey bars). In both panels A and B, the plating efficiency was not affected in *Eco*A strain. The data shown are averages of two biological replicates.

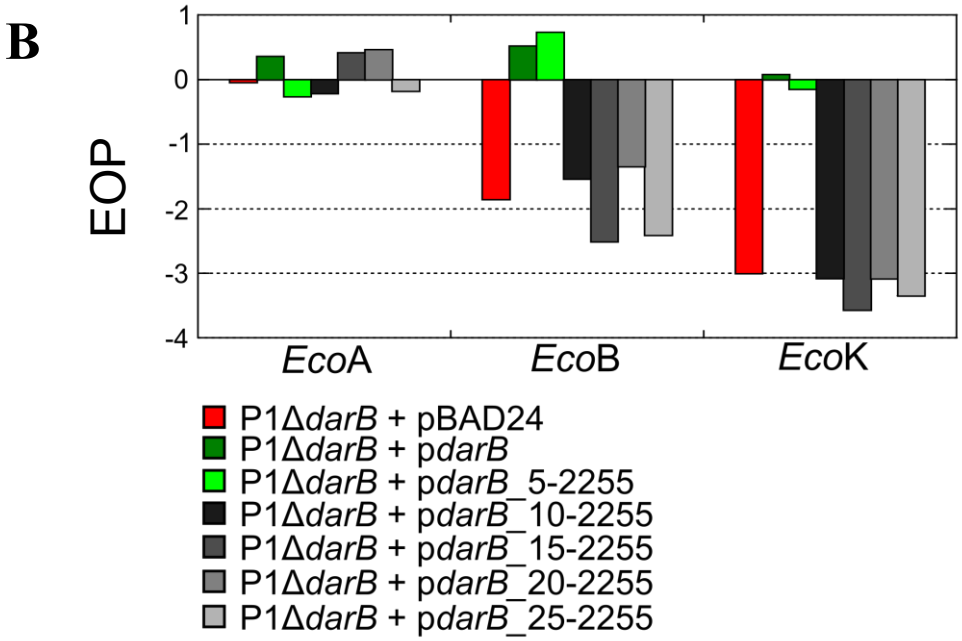


Fig. 3.6. Continued.

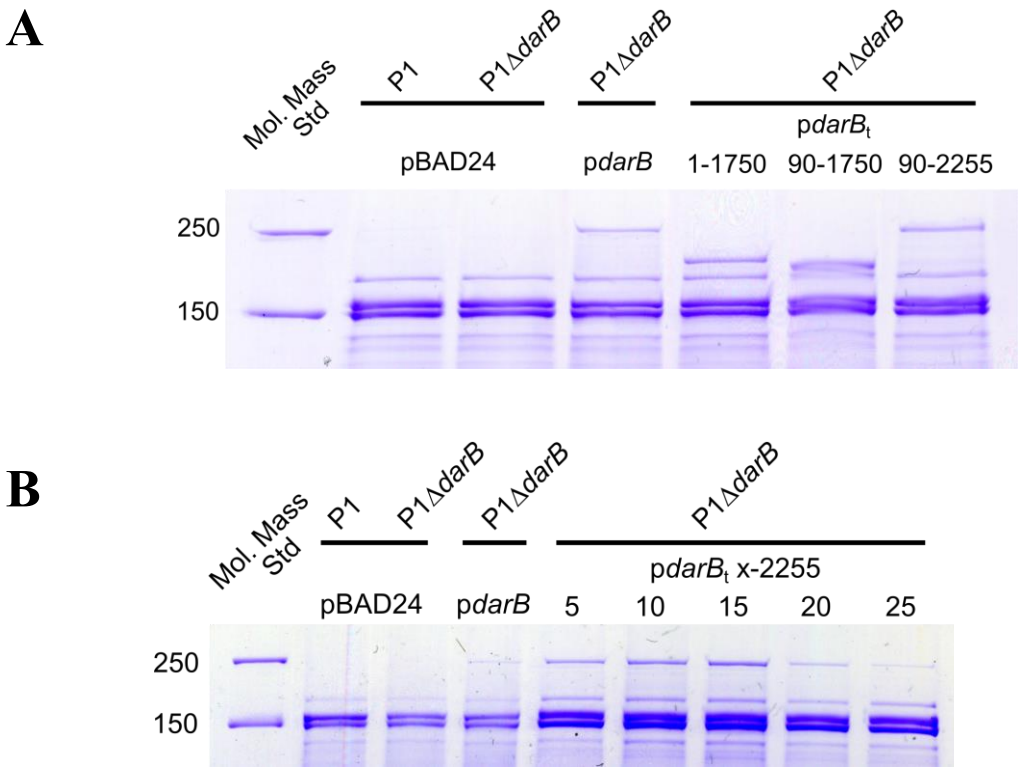


Fig. 3.7. Truncated DarB can be expressed. For both panels A and B, *E. coli* strain WA921 was transformed with empty plasmid pBAD24 or plasmid encoding full length or truncated DarB as denoted. Transformed cells were lysogenized with P1 or P1Δ*darB* as indicated. Phages and protein production from plasmids were induced simultaneously, as described before. Samples were collected ~30 min post-induction and analyzed by SDS-PAGE and Coomassie staining to check for protein induction from plasmids. Both full-length and truncated versions of DarB are induced from the plasmids, as evident from presence of corresponding protein bands.

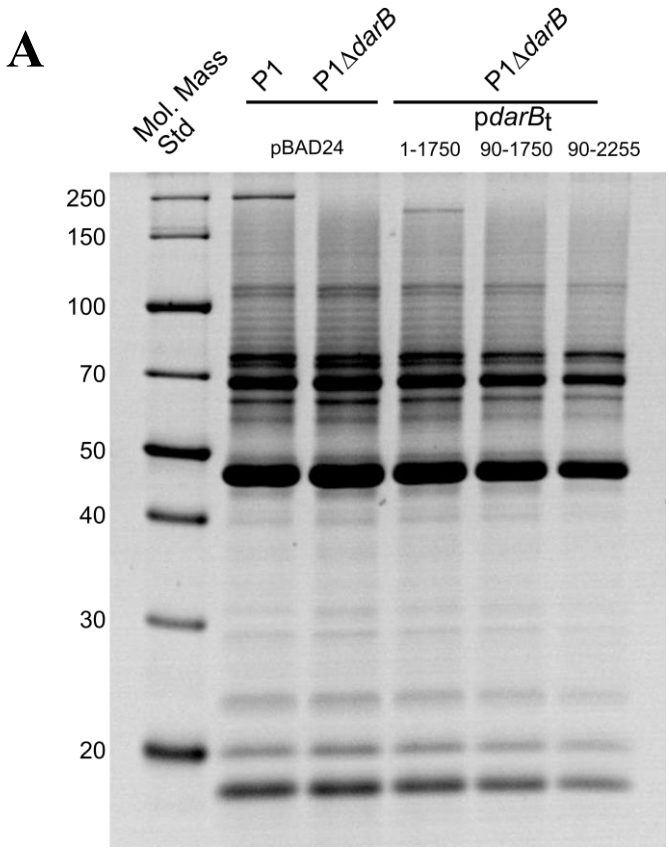


Fig. 3.8. N-terminal residues of DarB provide signal for capsid targeting. For both panels A and B, phages P1 or P1Δ*darB* were lysogenized into *E. coli* strain WA921 containing respective plasmid. Phages and proteins from plasmids were induced simultaneously as described before. Phages were purified by cesium chloride isopycnic gradient centrifugation and analyzed by SDS-PAGE, followed by SyproRuby staining. **Panel A:** $DarB_t$ 1-1750 is incorporated into P1Δ*darB* virions. **Panel B:** $DarB_t$ 5-2255 is packaged into P1Δ*darB* virions. All other truncations of DarB are not packaged into P1Δ*darB* virions, suggesting few N-terminal DarB residues provide signal for capsid targeting.

B

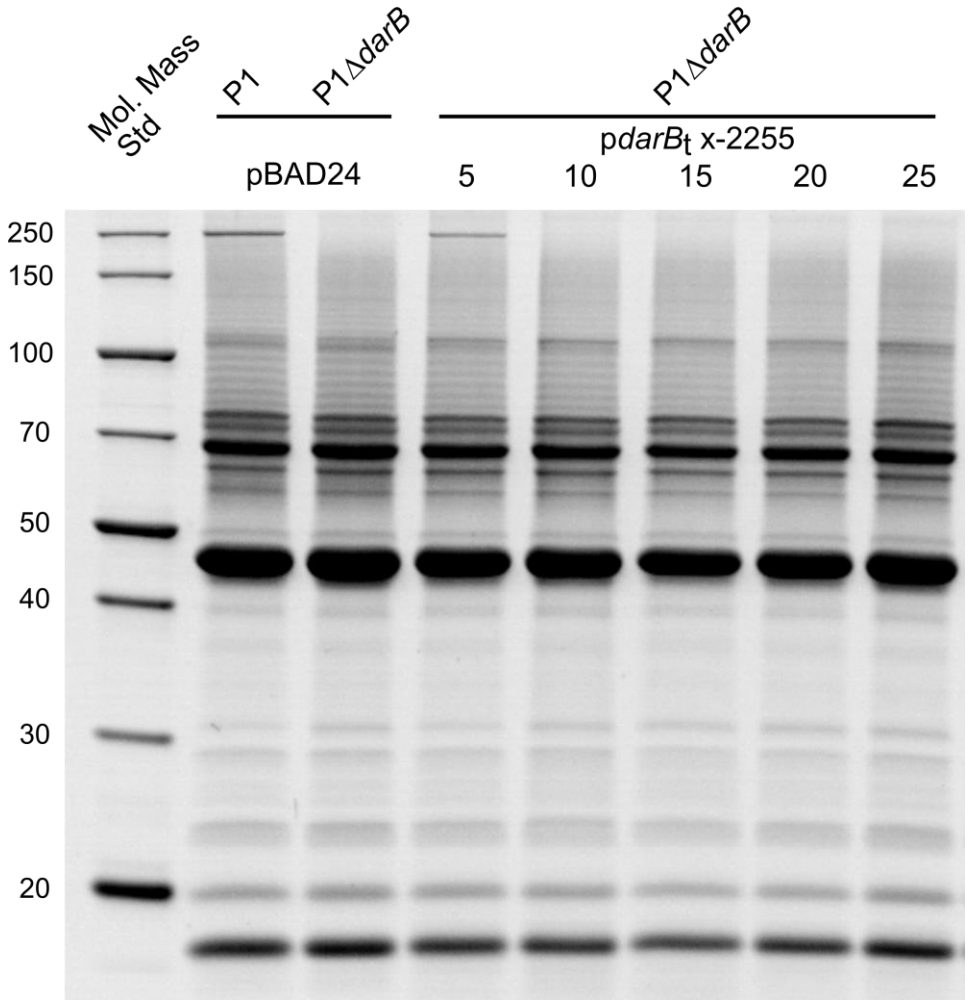


Fig. 3.8. Continued.

uncertain if $DarB_t$ 1-1750 can be ejected into host cells during infection. Moreover, $DarB_t$ 1-1750 is missing ~500 C-terminal residues. Even though there is no identifiable conserved domain within these C-terminal 500 residues of DarB, these 500 residues might be necessary for folding of DarB into native conformation or for establishing other protein-protein or protein-DNA interactions to perform its functions.

Once we established that N-terminal residues of DarB are required for capsid localization, we sought to determine the specific residues required for this purpose. Several constructs were produced to express truncated DarB translated from specific N-terminal residues: 5 ($DarB_t$ 5-2255), 10 ($DarB_t$ 10-2255), 15 ($DarB_t$ 15-2255), 20 ($DarB_t$ 20-2255) and 25 ($DarB_t$ 25-2255). Of these truncations, only $DarB_t$ 5-2255 was able to complement the antirestriction phenotype of P1Δ*darB* (Fig. 3.6B). SDS-PAGE of the induced lysogens showed that other four truncated DarB proteins were expressed from plasmids as well (Fig. 3.7B). P1Δ*darB* virions complemented with these series of truncated DarB were purified as described above and analyzed by SDS-PAGE. Protein band corresponding to $DarB_t$ 5-2255 was visible in SDS-PAGE, which corelates to why $DarB_t$ 5-2255 was able to complement antirestriction phenotype of P1Δ*darB*. Neither of the truncated DarB starting with residues 10, 15, 20, and 25 were packaged into the virions (Fig. 3.8B). This suggests that N-terminal amino acid sequences of DarB spanning residues 5 to 9 are crucial for targeting DarB into procapsids during capsid morphogenesis.

Purification of DarB

Genetic evidence suggests that DarB protects P1 DNA from host type I R-M system mediated DNA cleavage, following infection (Iida *et al.*, 1987; Piya *et al.*, 2017). However, biochemical mechanisms of DNA protection have not been elucidated. More importantly, it is not known if DarB needs any interacting partners to perform its biological role. Bioinformatic analysis shows that P1 DarB consists of a methyltransferase and a DExH helicase domain (Gill *et al.*, 2011). Based on the presence of conserved domains, biochemical activity of DarB can be predicted, however this must be confirmed *in vitro*. Since much of this future work will depend on the ability to obtain purified DarB protein, we sought to purify DarB by using affinity chromatography. Optimization of purification steps would also facilitate the study of the interacting partners of DarB by conducting pull-down assays (Green & Sambrook, 2012).

Initial attempts were made to purify DarB by cloning P1 *darB* into expression vector pETDuet-1 in-frame with the native His-tag. However, eluted DarB was not of high purity. Other available tags, glutathione S-transferase (GST) (GE Healthcare), maltose-binding protein (MBP) (New England Biolabs) and chitin binding domain (CBD) (New England Biolabs) were explored by fusing them to the N-terminus of DarB. Fusion of these tags to DarB produced varying results: inefficient binding of recombinant DarB to purification resin (GST-tag), toxicity (MBP-tag) and no protein expression (CBD-tag). Because of the relatively large size of DarB (~250 kDa), we reasoned that a single His-

tag did not provide sufficient binding affinity to the recombinant DarB. Hence, an additional His-tag was introduced into the backbone of pETDuet-1, downstream of the native His-tag, resulting in the production of recombinant DarB with 2x His-tag in its N-terminus (2x-His_DarB). The recombinant DarB has the following residues on its N-terminus: MGSSHHHHHHGSSHHHHHHSQDPNSSS (the underlined residues indicate His-tag)

Affinity purification of recombinant 2x-His_DarB yielded protein with significant purity as determined by SDS-PAGE analysis (Fig. 3.9). Even though His-tags are smaller compared to other affinity tags, they can still affect biochemical properties of recombinant proteins (Booth *et al.*, 2018). Thus, we sought to determine if 2x-His_DarB can complement the restriction phenotype of P1Δ*darB*. The gene encoding 2x-His_DarB was PCR amplified and cloned into pBAD24 vector to test for function of 2x-His_DarB *in vivo*. EOP assays suggest that 2x-His_DarB can restore the restriction phenotype of P1Δ*darB* (Fig. 3.10). Thus, it can be inferred that 2x His-tag fused to N-terminus of DarB does not affect the biochemical activity of DarB significantly, at least *in vivo*.

Conclusions

In Chapter 2 of this dissertation, it was demonstrated that P1 antirestriction is comprised of multiple components, in addition to the established *darA* and *darB* (Iida *et al.*, 1987; Piya *et al.*, 2017). Moreover, the role of *hdf* and *darA* in head size determination during virion morphogenesis suggest that P1 antirestriction is also linked to capsid

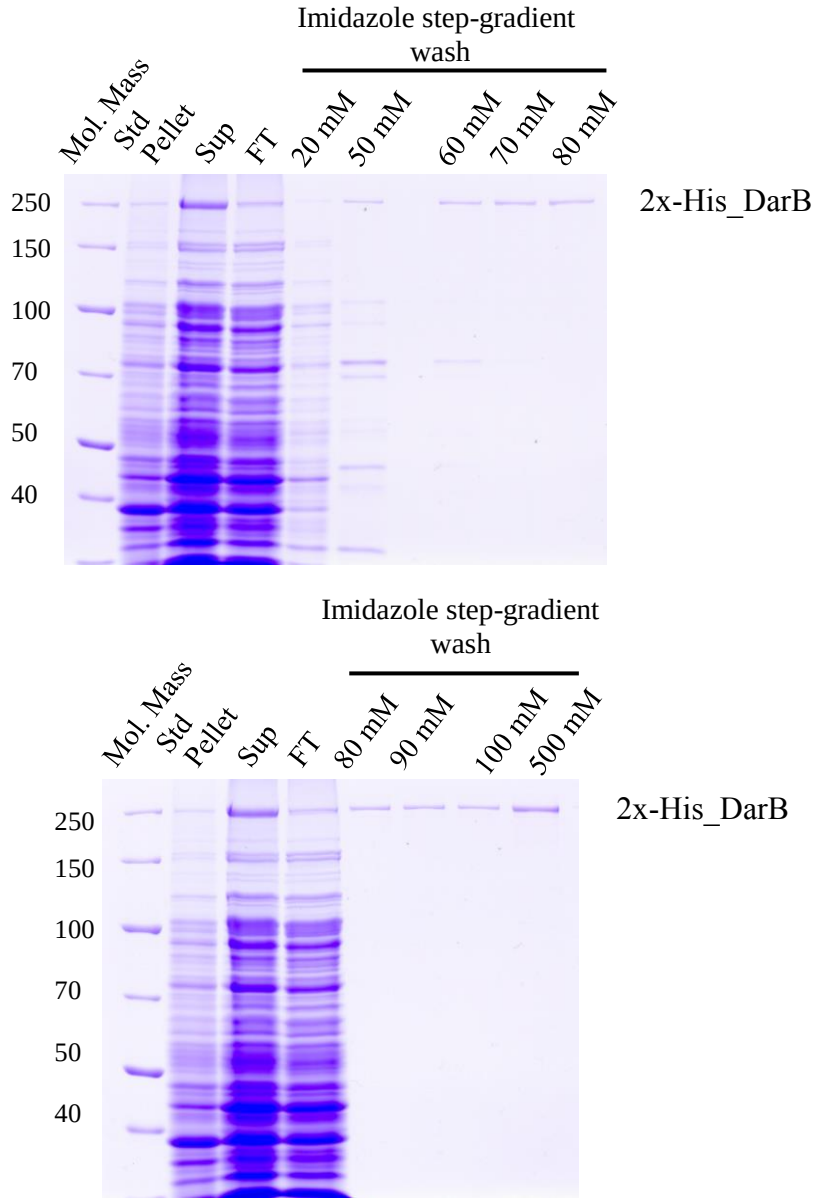


Fig. 3.9. IMAC purification of P1 DarB. *E. coli* BL21(DE3) strain was transformed with the expression plasmid pETDuet_2x-His_*darB* and protein production was induced by addition of 1 mM IPTG. BL21(DE3) cells expressing 2x His_DarB were sonicated and the supernatant (Sup) was applied to Ni-NTA agarose resin. Flow-through (FT) fraction was collected and non-specific proteins bound to the resin was washed with buffer containing increasing gradient of imidazole. 2x His_DarB bound to Ni-NTA resin was eluted with buffer containing 500 mM imidazole. All fractions were collected and analyzed by SDS-PAGE, followed by Coomassie staining.

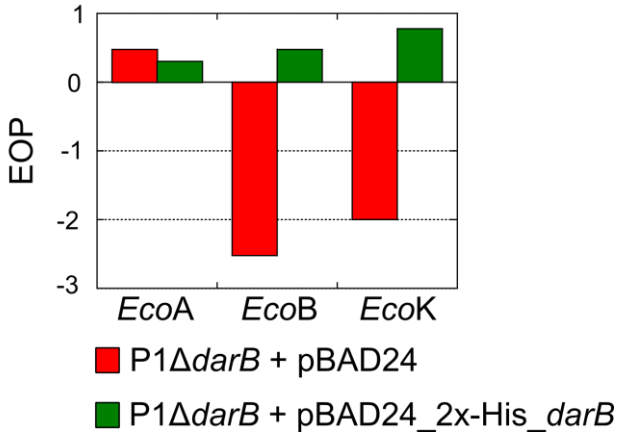


Fig. 3.10. DarB with 2x His-tag fused to N-terminus is functional. Phages for complementation were induced as described before. The EOP of P1Δ*darB* is reduced in *E. coli* strains with *Eco*B or *Eco*K systems, as demonstrated before (red bar). DarB with 2x His-tag fused to its N-terminus can restore the plating defect of P1Δ*darB* in strains with *Eco*B or *Eco*K systems suggesting that the N-terminal fusion of 2x His-tag does not disrupt function of DarB (green bar). The data shown are averages of two biological repeats.

morphogenesis. As the components of the P1 Dar system were distributed in two operons in P1 genome (Piya *et al.*, 2017), we sought to interrogate genes in other operons for their role in antirestriction and/or capsid morphogenesis. We targeted P1 genes of unknown function for this study. Isogenic deletion knockouts of these genes were created in P1 lysogens and induced virions were tested for restriction phenotype in *E. coli* hosts with type I *Eco*A, *Eco*B or *Eco*K systems by conducting plating assays (Fig. 3.1). The efficiency of plating (EOP) of all of these P1 mutants were comparable to the EOP of parental P1, which suggests that the antirestriction system is intact in these mutants (Fig. 3.2). Thus, these P1 genes do not appear to contribute to the activity of the P1 antirestriction system.

Lysates of knockout mutants of five P1 genes *pmgA, pmgB, pmgC, pmgG* and *pmgR* did not form plaques on lawns of their *E. coli* host (Fig. 3.1), indicating that these five genes are essential for plaque formation under laboratory conditions. At this stage, it is too early to ascribe any functions to these genes as plaque non-forming phenotype of these mutants still needs to be complemented. Genes *pmgA, pmgB, pmgC* and *pmgG* are found in vicinity of other genes encoding tail components. Although it is tempting to speculate that these genes encode components of tail assembly based on their location in P1 genome, it might be misleading to do so as genes encoding components of same pathway are not organized together in P1 genome. After complementing the phenotype pertaining to those genes, lysates of mutant virions may be analyzed by transmission electron microscopy to ascertain potential effects of these genes on virion morphogenesis.

Virions such as P22 and T4 incorporate several internal proteins into procapsids during earlier stages of capsid morphogenesis, followed by DNA packaging (Bazinet & King, 1988; Arisaka, 2005; Aksyuk & Rossmann, 2011). By disrupting DNA packaging in P1, virion morphogenesis was stalled at formation of procapsids and tail and both procapsids and tails were purified in same fraction. Comparative SDS-PAGE analysis of the procapsids/tail fractions of P1Δ*pacA* and P1Δ*darA*_Δ*pacA* suggest that P1 antirestriction components are packaged before DNA (Fig. 3.3).

P1 Dar proteins are incorporated in capsids and at least one of these, DarB, is injected into the new host during infection. As phage DNA ejected into host are packaged in the inner space of capsids, it can be extrapolated that these Dar proteins are also packaged in the same space. It was demonstrated in Chapter 2 of this dissertation that P1 Dar proteins occupy a significant mass in the capsid (Piya *et al.*, 2017), which suggests that not having Dar proteins inside of P1 procapsids will leave more space for DNA to be packaged. Virions of P1Δ*darA* cannot package any of the Dar proteins, thus internal space in P1Δ*darA* procapsids is likely to be greater than those of P1 procapsids. Since P1 packages DNA into procapsids by headful mechanism (Bachi & Arber, 1977; Lobocka *et al.*, 2004), virions of P1Δ*darA* should be able to package more DNA than P1, which can be tested by pulse-field gel electrophoresis of genomic DNA (Lingohr *et al.*, 2009).

The internal volume of P1 procapsids, as seen in negative-stained EM images, has an asymmetrically distributed electron density, suggestive of proteins because P1 procapsids constructed in this study are unable to package DNA (Fig. 3.5). Procapsids harvested at similar stages of virion morphogenesis in other phages such as P22 and T4 also show similar electron densities, which are attributed to the scaffolding core. However, this scaffolding core of both P22 and T4 occupy most of the space inside of procapsid, compared to partial space in P1 (King *et al.*, 1973; Traub & Maeder, 1984). This suggest that the scaffolding core of P1 is either proteolytically processed and the fragments are removed before the start of DNA packaging or there is a novel mechanism associated with head morphogenesis. Identification of proteolytic processing enzyme and scaffolding proteins associated with P1 head morphogenesis, if there is any, would assist in studying P1 capsid morphogenesis.

Comparative SDS-PAGE analysis of CsCl purified mature P1 virions and P1 procapsids/tail fractions show few proteins bands that are only present in P1 procapsids/tail fractions. These new protein bands present in the lanes of procapsids/tail fractions are suggestive of constituents of scaffolding core, as this core is processed and removed during later stages of capsid morphogenesis (Dokland, 1999). However, due to the limitations of current purification methods, P1 procapsids and tail components are purified in same fractions. Moreover, these fractions may contain contaminating host proteins. After optimization of procapsids purification, it can be determined from SDS-

PAGE if these proteins bands are still present in the lanes of procapsids and can be identified by mass spectrometry.

P1 antirestriction protein DarB is incorporated into capsid in a definite pathway involving other capsid-associated proteins as well (Piya *et al.*, 2017). By analyzing incorporation of DarB mutants, truncated for several residues at N- or C- terminus, into P1Δ*darB* virions, we have determined that N-terminal residues ranging from 5^{th} to 9^{th} are essential for packaging of DarB into P1 capsids (Fig. 3.8B). Residues that are crucial for targeting proteins into capsids have also been found in other phages. Phage T4 capsids contain hundreds of copies of internal proteins (IPI, IPII and IPIII) (Black & Ahmad-Zadeh, 1971). It has been determined that IPIII is incorporated into T4 procapsids with a capsid targeting sequence (CTS) present in N-terminus of this protein (Mullaney & Black, 1996). Moreover, the CTS can also direct other foreign proteins to be packaged into T4 capsids and these foreign proteins are further injected into new host during T4 infection (Hong & Black, 1993; Mullaney & Black, 1996). Although DarB N-terminal residues 5^{th}-9^{th} seem to be required for packaging of DarB into P1 virions, it is not evident if those residues are sufficient. Fusion of foreign proteins to the respective N-terminal residues of DarB could tell if these particular residues are sufficient for packaging any protein into P1 capsids. Moreover, fusion of fluorescent proteins could be advantageous as injection of these fusion proteins into new host following P1 infection can be tested using microscopy.

Although it has been established that DarB protects P1 DNA from type I R-M system mediated DNA cleavage following ejection of DNA into host cytoplasm (Iida *et al.*, 1987; Piya *et al.*, 2017), biochemical mechanism behind this protection has not been determined. Because of the presence of conserved methyltransferase and helicase domains in DarB, it may be hypothesized that DarB methylates and/or translocates along DNA following ejection of both DarB and DNA into new host (Lobocka *et al.*, 2004; Gill *et al.*, 2011). In order to determine the activity of DarB *in vitro*, we sought to purify DarB using affinity chromatography. By fusion of 2x His-tag to N-terminus of P1 DarB, recombinant DarB could be purified, by immobilized metal affinity chromatography, to sufficient purity to conduct enzyme assays (Fig. 3.9). Moreover, we have also shown that 2x His-tag fusion to DarB does not abolish its biological activity (Fig. 3.10). *In vitro* enzyme assays can now be conducted to determine the biochemical mechanism of how DarB protects DNA from type I R-M system mediated restriction.

Many proteins can perform biological functions only when they are associated with other protein subunits forming a complex. Thus, it cannot be ruled out yet that DarB could only function in presence of other protein subunits. In the type I R-M system complex, DNA specificity is provided by HsdS subunit (Murray, 2000). Since DarB has been shown to protect DNA against other type I R-M systems from *Salmonella* as well, it is plausible that DarB might interact with HsdS subunit of type I R-M complex (Iida *et al.*, 1987). Moreover, other proteins from Dar system might be co-injected with DarB during phage infection. Since affinity tag-based purification of DarB has been optimized,

same strategy can be applied to set up pull-down assays to determine interacting partners of DarB. Setting up *in vitro* enzyme assays for DarB may be worthwhile, only after discovering interacting partners of DarB, if any.

CHAPTER IV

GENOME-WIDE SCREENS REVEAL ESCHERICHIA COLI GENES REQUIRED FOR GROWTH OF T1-LIKE PHAGE LL5 AND RV5-LIKE PHAGE LL12

Introduction

Escherichia coli was discovered by the German microbiologist Theodor Escherich, in the study of infant gut microbes (Escherich, 1988). Because of several traits such as easily culturable, short doubling time, *E. coli* has been used as a model organism for understanding basic biology as well as for biotechnology applications. *E. coli* is a member of Enterobacteriaceae, and is related to other pathogens such as *Salmonella*, *Klebsiella*, *Serratia* and *Yersinia*. *E. coli* is a facultative aerobe, and along with other obligate anaerobes, forms a part of commensal gut flora (Blount, 2015). These commensal *E. coli* strains do not cause disease in human host, barring immunocompromised individuals. However, there are several strains of *E. coli* that have acquired virulence factors and cause diseases with symptoms, ranging from mild discomfort to life-threatening. Depending upon disease etiology, these strains have been categorized into several pathotypes: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), and diffusely adherent *E. coli* (DAEC) (Kaper *et al.*, 2004).

This study aims to study phage-host interactions between *E. coli* and two virulent phages, LL5 and LL12, which were isolated against ETEC host strains. ETEC strains can be distinguished from other *E. coli* pathotypes by the presence of two toxins: heat-labile enterotoxins (LTs) and heat-stable enterotoxins (HTs). These strains might express only one or both of those toxins (Kaper *et al.*, 2004). These enterotoxins induce traveler's diarrhea (TD), characterized by watery diarrhea, for which the symptoms could range from mild to severe (Kaper *et al.*, 2004). TD is one of the most common illnesses contracted by people from developed countries travelling to less developed countries of the world. TD is also accompanied by other symptoms such as nausea, vomiting, abdominal pain, fever or blood in stool (Taylor *et al.*, 2017). The causative agent of TD varies in different geographical regions; *Campylobacter* spp. is the primary one in Southeast Asia whereas ETEC is the major one in the Latin America, Africa, south Asia and the Middle East (Tribble, 2017). Besides, other bacterial pathogens such as EAEC, *Shigella* spp., non-typhoidal *Salmonella* spp., and viruses and parasites have also been reported to cause TD (Taylor *et al.*, 2017; Tribble, 2017). TD has been, so far, successfully treated with antibiotics, but the global increase in the emergence of the antibiotic resistance in bacteria warrants evaluation of alternative treatment approaches (Tribble, 2017).

Bacteriophages (phages) are the natural predators of bacteria. Due to emergence of multidrug resistant pathogens and limitation in the discovery of antibiotics, there has been a renewed interest in phage therapy in western medicine. Phage therapy is the

application of phages to treat pathogenic bacteria (Young & Gill, 2015). Phage therapy could be a solution to increasing antibiotic resistance incidence in the causative agents of TD. Phage infection cycle starts with adsorption of phages to host receptors, ejection of phage genetic material into host, replication of genetic material and production of phage structural components, followed by lysis of host cells to liberate progeny phages to surroundings. As bacteria are constantly preyed upon by these phages, bacteria are evolved with diverse defense mechanisms to thwart phage infection (Labrie *et al.*, 2010). Hence, for phage therapy to succeed, the necessity of thorough understanding of phage-host interactions cannot be overstated.

Phage replication needs to be very robust because phage infection cycle is short. Phages may not carry all genes required for necessary functions because there is a limitation in their genome size. Therefore, to have an efficient replication cycle, phages utilize various host functions. For instance, phages need quick and abundant chaperone power so that the assembly of virions can be completed before host lysis (Georgopoulos, 2006). Same holds true for eukaryotic viruses as well. Before the advent of modern genetic tools, phage-host interactions were studied using classic genetic tools. Despite these classic approaches being powerful, it does have its limitations as these techniques are labor-intensive and lengthy. With the availability of modern genetic resources, the large-scale host-phage interactions can be studied efficiently. By applying these modern genetic approaches, a number of genome-wide screens have been conducted to study host factors required for viral replication in organisms such as HIV (Brass *et al.*, 2008),

Influenza virus (Hao *et al.*, 2008), phage λ (Maynard *et al.*, 2010) and phage T7 (Qimron *et al.*, 2006).

We conducted forward-genetics screen of phages, LL5 and LL12 against the Keio collection to characterize any host function that is required for efficient phage replication. The Keio collection is a library of single-gene deletions of all non-essential genes in *E. coli* K-12 strain BW25113 (Baba *et al.*, 2006). This screen will help in understanding host-virus interaction (Maynard *et al.*, 2010) by revealing host factors involved in infection and replication of LL5 and LL12. The phages will not be able to infect and propagate efficiently in a host devoid of certain crucial functions, resulting in these mutant hosts outgrowing the WT host, upon simultaneous infection of the phages. Identification of host factors required for phage propagation is crucial for the development of therapeutic phages because the bacterial host can develop resistance to phages by mutating the host factors, resulting in absolute protection from phage infection and propagation. If the knowledge about host factors is available, then the phage genome can be engineered to include those factors so that there is reduced dependency in the host genome encoded factors (Qimron *et al.*, 2006). More importantly, if the host receptors can be characterized for individual phages, phage cocktails, composed of phages targeting different host receptors, can be prepared, which will improve the efficacy of phage therapy. Understanding host-virus interaction is very crucial for the application of phage therapy because it will facilitate the engineering of

robust phages (Qimron *et al.*, 2006). The results of this screen are discussed in terms of host factors required for infection and propagation of phages LL5 and LL12.

Materials and methods

Bacterial strains and plasmids

The Keio collection was purchased from Thermo Scientific (Baba *et al.*, 2006). Strains from the ASKA library used for complementation were purchased from National BioResource Project (NIG, Japan) (Kitagawa *et al.*, 2005). The parental *E. coli* strain BW25113 was obtained from Ry Young (Texas A&M University, College Station, TX). *E. coli* strains from the Keio collection and their transductants were cultured in LB (Lennox) broth [10 g L^{-1} Bacto tryptone (BD), 5 g L^{-1} Bacto yeast extract (BD), 5 g L^{-1} NaCl (Avantor)] or LB agar [LB broth amended with 15 g L^{-1} Bacto agar (BD)] at 37 °C amended with 30 μg mL^{-1} kanamycin (LB kan) and strains containing plasmids from the ASKA library were maintained on LB amended with 10 μg mL^{-1} chloramphenicol (LB cm). Plasmid DNA from ASKA library strains was extracted using a QIAprep Spin Miniprep Kit (Qiagen). In complementation experiments with the ASKA plasmids, LB plates and top agar were supplemented with 0.05 - 0.1 mM IPTG to induce protein expression (Kitagawa *et al.*, 2005).

Phage isolation and culture

The phages LL5 and LL12 were isolated against clinical isolates of enterotoxigenic *E. coli* (ETEC) obtained from John Deaton (Deerland Enzymes, Kennesaw, GA). Phages

were isolated by the enrichment method (Gill *et al.*, 2012) from filter-sterilized (0.22 μm) wastewater influent collected in College Station, TX in 2011. Both phages were subsequently cultured using *E. coli* strain DH5α as host. Phage lysates were prepared by the confluent plate lysis method (Adams, 1959) using LB (Miller) bottom plates (10 g L^{-1} Bacto tryptone, 5 g L^{-1} Bacto yeast extract, 10 g L^{-1} NaCl, 15 g L^{-1} Bacto agar) and top agar consisting of 10 g L^{-1} tryptone, 10 g L^{-1} NaCl, 5 g L^{-1} Bacto agar. Phages were harvested and stored as filter-sterilized (0.22 μm) lysates in lambda diluent (25 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$, 0.01% w/v gelatin) at 4 ºC (Green & Sambrook, 2012).

Plaque assays were conducted using both spot titer and full-plate titration methods (Adams, 1959). For spot titers, phages were diluted ten-folds and 10 μL of each dilution was spotted on solidified lawns of 4 ml top agar inoculated with 100 μL of a fresh overnight host culture prepared as described above. For full-plate titers, 100 μL of phages diluted ten-folds were mixed with 100 μL of host culture in 4 ml of molten top agar and poured over LB plates as described above. Plaques were enumerated after 16-18 h incubation at 37 °C. The efficiency of plating (EOP) was calculated as the ratio of the number of plaques appearing on the lawn of a test strain to the number of plaques on the reference strain.

Phage genome sequencing and annotation

Phage DNA was purified from high-titer lysates by a modified Wizard DNA purification kit (Promega) as previously described (Summer, 2009). Phage LL5 was sequenced by 454 pyrosequencing at the Emory GRA Genome Center (Emory University, GA); trimmed FLX Titanium sequence reads were assembled into a single contig at 19.9-fold coverage using Newbler 2.5.3 (454 Life Sciences) at default settings. Phage LL12 was sequenced by Illumina TruSeq as unpaired 100-base reads; reads were quality-controlled by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and assembled with Velvet 1.1 (Zerbino, 2010) into a single contig at 28.3-fold coverage. Assembled phage contigs were confirmed to be complete by PCR using primers facing off each end of the contig and sequencing of the resulting products. Structural annotation was conducted using Glimmer3 (Delcher *et al.*, 1999) and MetaGeneAnnotator (Noguchi *et al.*, 2008) with tRNAs predicted by ARAGORN (Laslett & Canback, 2004) or tRNAscan-SE (Lowe & Chan, 2016) and gene functions predicted by InterProScan (Jones *et al.*, 2014) or Conserved Domain Database (Marchler-Bauer *et al.*, 2017), TMHMM (http://www.cbs.dtu.dk/services/TMHMM), BLASTp (Camacho *et al.*, 2009) and HHpred (Soding *et al.*, 2005). Phage genome annotation was conducted using the Phage Galaxy instance hosted by the Center for Phage Technology at Texas A&M University (cpt.tamu.edu). The annotated phage genomes were deposited in NCBI Genbank under accession no. MH491968 (LL5) and MH491969 (LL12).

Transmission electron microscopy

Phages were stained with 2% uranyl acetate and imaged in a JEOL 1200EX transmission electron microscope (TEM) under 100 kV accelerating voltage at the Texas A&M University Microscopy and Imaging Center, as previously described (Valentine *et al.*, 1968; Piya *et al.*, 2017). The size parameters of phages were measured electronically using ImageJ (Schneider *et al.*, 2012).

Screening and confirmation of phage-insensitive mutants

In order to optimize the input phage concentrations and incubation times, phages LL5 and LL12 stocks were diluted ten-folds in fresh LB and 160 µL of each dilution was aliquoted into 96-well sterile transparent polystyrene flat-bottom plates (Greiner Bio-one). The plates were then inoculated with the Keio parental strain BW25115 using a 96-pin replicator (Phenix) and incubated at 37 °C for 6, 8, 10 and 18 hrs. The optical density (OD) at 550 nm was measured in a Tecan M200 plate reader at each time interval and the average OD was analyzed to determine the lowest phage concentration that inhibited bacterial growth.

The Keio collection consists of 90, 96-well plates containing two independently-generated sets of 3,985 single-gene knockouts in the *E. coli* BW25113 background (Baba *et al.*, 2006). The Keio strains were replicated into 96 well sterile polypropylene U-bottom microplates (Greiner Bio-one) containing LB kan + 8% glycerol using sterile plastic 96-pin replicators (Phenix). Plates were incubated at 37 ºC overnight and stored

frozen at -80 ºC. These plates were used as the working stocks for the following screens. The odd- and even-numbered plates have identical gene deletion mutants created by independent experiments (Baba *et al.*, 2006), and only the odd-numbered 45 plates were used for the initial screen. Initial screens were conducted in 96-well sterile transparent polystyrene flat-bottom plates (Greiner Bio-one). Phages LL5 and LL12 were diluted in fresh LB to obtain working stocks of 10^6 PFU/mL for LL5 and 10^3 PFU/mL for LL12. 160 µL of the phage working stocks were aliquoted into all wells, Keio strains were inoculated into the phage lysates from the 96-well working stocks with 96-pin replicators, and the plates were incubated for 8 hours at 37 ˚C. The OD_{550} was measured and the wells with OD_{550} higher than the pre-determined cutoff values (0.2 for phage LL5 and 0.11 for phage LL12) were scored as positive for growth.

The positive mutants obtained from the first screen were verified by repeating the assay with the same strains and their corresponding mutant strains from the even-numbered Keio collection plates, side-by-side with eight replicates per assay. Mutants that returned mean OD_{550} above the designated cutoff in either the even- or odd-numbered set were retained for further characterization by measurement of phage efficiency of plating (EOP) by spot assays on soft agar lawns (Adams, 1959) as described above. EOP was calculated as the number of plaques observed on the mutant strain divided by the number of plaques observed on the parental *E. coli* strain BW25113. Mutants with EOPs of less than 10^{-2} were confirmed by enumerating plaques on full plates. When possible, mutant alleles were moved into the parental BW25113 background by P1

transduction using the kanamycin resistance cassette as the selectable marker (Thomason *et al.*, 2007). All gene disruptions were confirmed by PCR using primers flanking the predicted insert followed by sequencing of the PCR product to confirm disruption of the gene. All mutants were complemented by transforming the original Keio mutant or its P1 transductant with a plasmid expressing the corresponding gene under control of the *lac* promoter; all complementing plasmids were obtained from the ASKA collection (NIG, Japan).

Results and Discussion

Isolation and characterization of phages LL5 and LL12

Phages LL5 and LL12 were isolated from municipal wastewater in College Station, TX by enrichment against enterotoxigenic *E. coli* (ETEC) clinical isolates. Shortly after isolation, both phages were determined to plate efficiently on *E. coli* K-12 strains including MG1655, DH5α and the Keio parental strain BW25113 and these phages were subsequently propagated on *E. coli* DH5α for the remainder of the study.

Phages LL5 and LL12 have distinct morphology as observed by transmission electron microscopy (Fig. 4.1). LL5 is a siphophage with a head diameter of 61 nm (± 2 nm) and a flexible tail 156 nm (± 9.9 nm) in length. Phage LL12 is a large myophage with head diameter of 86.4 nm (± 2.4 nm) and a tail 112.3 nm (± 4.1 nm) in length with a pronounced baseplate.

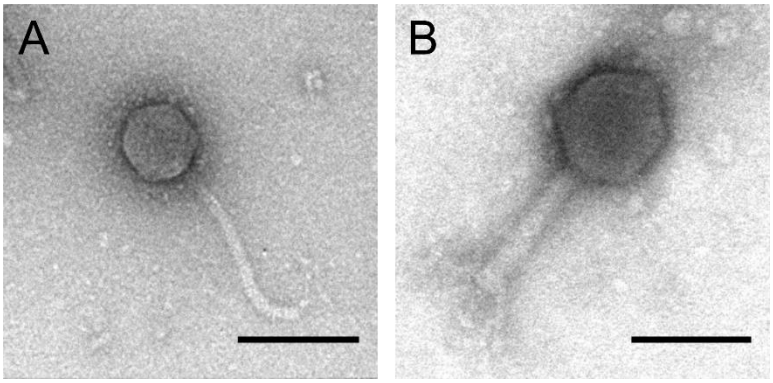


Fig. 4.1. Transmission electron micrographs of phages LL5 and LL12. The grids for imaging were prepared as described in materials and methods. **Panel A:** Phage LL5 has a capsid diameter of 61 nm (±2 nm) and a tail length of 156 nm (±9.9 nm). **Panel B:** Phage LL12 has a capsid diameter of 86 nm (±2.4 nm) and a tail length of 112 nm (±4.1 nm). These size parameters are an average of ten measurements and the error represents standard deviation. The scale bar represents 100 nm.

LL5 has a genome of 49,788 bp with 88 predicted protein-coding genes and no tRNA genes. The genome produced a circular assembly and was reopened be syntenic with other T1-like phages in the NCBI database such as TLS (NC_009540) and T1 (NC_005833). Thirty-three LL5 encoded proteins could be assigned putative functions (Table 4.1, Fig. 4.2A). Genes responsible for different stages of phage infection cycle have been identified in LL5, including a DNA primase/helicase (gp58), ATP-dependent helicase (gp60) and helicase (gp75). Structural proteins including the portal protein (gp31), major capsid protein (gp36), minor tail proteins (gp41, gp42, gp47, gp48), tail tube protein (gp43), tape measure protein (gp46), and tail fiber proteins (gp51, gp57) were identified. The small and large terminase subunits were identified as gp29 and gp30, respectively. Like its T1-like relatives, LL5 encodes a canonical lysis cassette composed of a holin (gp70), endolysin (gp71) and unimolecular spanin (gp72).

Phage LL5 is closely related to the T1-like coliphage TLS (NC_009540) (German & Misra, 2001), with 96% sequence identity over 90% query coverage of the LL5 genome based on BLASTn analysis. As would be expected given this close similarity, the LL5 genome is syntenic with TLS, with 75 LL5 proteins having homologs in the TLS genome detectable by BLASTp with an E value of less than 10^{-5}. Phage LL5 encodes two predicted tail fiber proteins, gp51 and gp57, in a genomic arrangement similar to that found in phage TLS and T1. LL5 gp51 is closely related to predicted tail fibers in other phages including gp51 of TLS (also called TspJ, YP_001285540, 98% identity), gp33 of phage T1 (FibA, YP_003912, 67% identity) and the central tail fiber protein J of

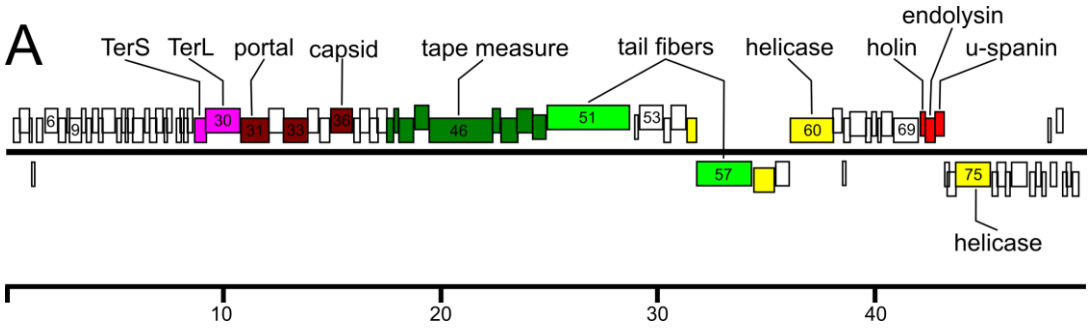


Fig. 4.2. Genome maps of phages LL5 and LL12. The predicted genes of phages LL5 (panel A) and LL12 (panel B) are represented as boxes, which are numbered to match locus number. The genes present in positive and negative DNA strands of phage genomes are separated by a line; the boxes above the line denote genes in the positive strand, whereas those below the line denote genes in the negative strand. Some of the predicted gene products playing important roles in phage infection cycle are indicated. The colored boxes represent genes predicted to perform similar functions. The scale bar represents the genomic loci in kilobases.

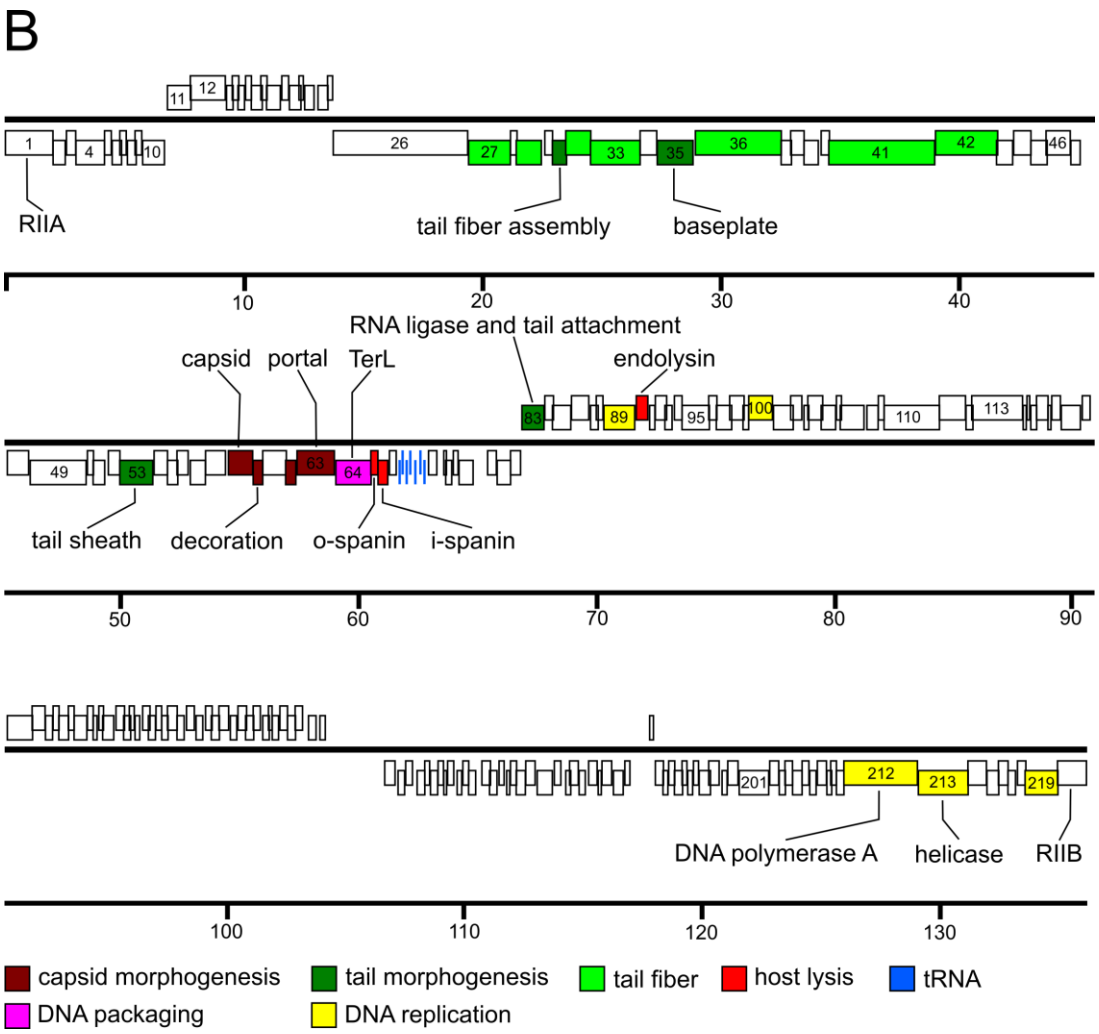


Fig. 4.2. Continued.

Table 4.1. Proteins encoded by genome of phage LL5

Gene	Start	End	Strand	Gene product
CPT_LL5_01	365	652	+	Conserved hypothetical protein
CPT_LL5_02	664	1107	+	Conserved hypothetical protein
CPT_LL5_03	1104	1211	+	Conserved hypothetical protein
CPT_LL5_04	1224	1346	-	Conserved hypothetical protein
CPT_LL5_05	1415	1678	+	Conserved hypothetical protein
CPT_LL5_06	1830	2396	+	Conserved hypothetical protein
CPT_LL5_07	2456	2770	+	Conserved hypothetical protein
CPT_LL5_08	2843	2950	+	Conserved hypothetical protein
CPT_LL5_09	2954	3496	+	Polynucleotide kinase/phosphatase
CPT_LL5_10	3493	3669	+	Conserved hypothetical protein
CPT_LL5_11	3764	3973	+	Hypothetical protein
CPT_LL5_12	4047	4301	+	Conserved hypothetical protein
CPT_LL5_13	4298	4480	+	Conserved hypothetical protein
CPT_LL5_14	4473	5048	+	Conserved hypothetical protein
CPT_LL5_15	5140	5340	+	Conserved hypothetical protein
CPT_LL5_16	5342	5506	+	Conserved hypothetical protein
CPT_LL5_17	5508	5633	+	Conserved hypothetical protein
CPT_LL5_18	5626	5871	+	Conserved hypothetical protein
CPT_LL5_19	5883	6353	+	Putative GntR-family transcriptional regulator
CPT_LL5_20	6427	6630	+	Conserved hypothetical protein
CPT_LL5_21	6634	6939	+	Conserved hypothetical protein
CPT_LL5_22	6936	7253	+	Conserved hypothetical protein
CPT_LL5_23	7320	7448	+	Conserved hypothetical protein
CPT_LL5_24	7450	7677	+	Conserved hypothetical protein
CPT_LL5_25	7860	8069	+	Conserved hypothetical protein
CPT_LL5_26	8059	8220	+	Conserved hypothetical protein
CPT_LL5_27	8204	8386	+	Conserved hypothetical protein
CPT_LL5_28	8386	8616	+	Conserved hypothetical protein
CPT_LL5_29	8704	9228	+	Terminase small subunit
CPT_LL5_30	9240	10811	+	Terminase large subunit

Table 4.1. Continued.

Gene	Start	End	Strand	Gene product
CPT_LL5_31	10865	12151	+	Portal protein
CPT_LL5_32	12156	12827	+	Conserved hypothetical protein
CPT_LL5_33	12824	13933	+	putative scaffold or prohead protease
CPT_LL5_34	13946	14425	+	Conserved hypothetical protein
CPT_LL5_35	14469	14909	+	Conserved hypothetical protein
CPT_LL5_36	14999	15973	+	Major capsid protein
CPT_LL5_37	16035	16307	+	Conserved hypothetical protein
CPT_LL5_38	16354	16773	+	Conserved hypothetical protein
CPT_LL5_39	16770	17141	+	Conserved hypothetical protein
CPT_LL5_40	17134	17574	+	Conserved hypothetical protein
CPT_LL5_41	17564	17857	+	Minor tail protein
CPT_LL5_42	17884	18087	+	Minor tail protein
CPT_LL5_43	18102	18764	+	tail tube protein
CPT_LL5_44	18842	19156	+	tapemeasure chaperone protein
CPT_LL5_45	18842	19473	+	tapemeasure chaperone protein frameshift product
CPT_LL5_46	19511	22420	+	tail tape measure protein
CPT_LL5_47	22420	22767	+	Minor tail protein
CPT_LL5_48	22835	23593	+	Minor tail protein
CPT_LL5_49	23590	24312	+	Tail tip assembly protein
CPT_LL5_50	24305	24904	+	Tail assembly protein
CPT_LL5_51	24986	28762	+	Tail fiber protein
CPT_LL5_52	29018	29161	+	Hypothetical protein
CPT_LL5_53	29224	30279	+	Exodeoxyribonuclease VIII
CPT_LL5_54	30357	30647	+	Conserved hypothetical protein
CPT_LL5_55	30694	31356	+	Recombinase
CPT_LL5_56	31395	31820	+	Single-stranded DNA binding protein
CPT_LL5_57	31853	34363	-	Tail fiber protein
CPT_LL5_58	34495	35421	-	DNA primase/helicase
CPT_LL5_59	35479	36087	-	Putative transcriptional regulator
CPT_LL5_60	36176	38149	+	ATP-dependent helicase

Table 4.1. Continued.

Gene	Start	End	Strand	Gene product
CPT_LL5_61	38152	38559	+	VRR-NUC domain protein
CPT_LL5_62	38549	38683	-	Conserved hypothetical protein
CPT_LL5_63	38631	38909	+	Conserved hypothetical protein
CPT_LL5_64	38911	39651	+	Dam methylase
CPT_LL5_65	39653	39883	+	Conserved hypothetical protein
CPT_LL5_66	39922	40119	+	Conserved hypothetical protein
CPT_LL5_67	40201	40341	+	Conserved hypothetical protein
CPT_LL5_68	40344	40823	+	HNH endonuclease
CPT_LL5_69	40907	42022	+	Conserved hypothetical protein
CPT_LL5_70	42152	42358	+	Putative holin
CPT_LL5_71	42358	42798	+	glycoside hydrolase endolysin
CPT_LL5_72	42845	43249	+	unimolecular spanin protein
CPT_LL5_73	43266	43460	-	Conserved hypothetical protein
CPT_LL5_74	43384	43776	-	Conserved hypothetical protein
CPT_LL5_75	43779	45359	-	Helicase
CPT_LL5_76	45427	45675	-	Conserved hypothetical protein
CPT_LL5_77	45672	46061	-	Conserved hypothetical protein
CPT_LL5_78	46063	46257	-	Hypothetical protein
CPT_LL5_79	46318	47013	-	Site-specific DNA methylase
CPT_LL5_80	47209	47448	-	Conserved hypothetical protein
CPT_LL5_81	47454	47675	-	Conserved hypothetical protein
CPT_LL5_82	47734	47901	-	Conserved hypothetical protein
CPT_LL5_83	48003	48134	+	Hypothetical protein
CPT_LL5_84	48131	48400	-	Conserved hypothetical protein
CPT_LL5_85	48428	48700	+	Hypothetical protein
CPT_LL5_86	48684	48878	-	Conserved hypothetical protein
CPT_LL5_87	48875	49078	-	Conserved hypothetical protein
CPT_LL5_88	49150	49428	-	Conserved hypothetical protein

phage lambda (NP_040600, 23% identity). Approximately 3 kb downstream and on the opposite strand from gp51, gp57 is similar to other T1-like tail fibers only in its N-terminal domain, with 46% identity to T1 FibB (YP_003919) from residues 1-290. The C-terminal domain of gp57 is more closely related to tail fiber proteins found in T5-like phages such as DT57C and DT571/2 (Golomidova *et al.*, 2015). LL5 gp57 is 57% identical with a C-proximal region spanning residues 515 to 830 of the 1,076-residue DT57C LtfA protein (YP_009149889), which is within the host specificity region of this protein (Golomidova *et al.*, 2016).

The LL12 genome was determined to be 136,026 bp in length and encodes 213 predicted protein-coding genes and 7 tRNAs. The genome produced a circular assembly and was reopened at a point between the genes encoding T4 RIIA and RIIB homologs, to retain its general synteny with other RIIAB-encoding myophages. Analysis of the raw Illumina reads by PhageTerm (Garneau *et al.*, 2017) suggests the presence of a non-permuted terminal redundancy of 459 bp spanning bases 104,966 - 105,424 in the genome as presented here. This predicted terminal repeat is located in a non-coding region of DNA between two convergent transcripts and corresponds to the location of non-permuted terminal repeats observed in phage phi92 (Schwarzer *et al.*, 2012). Fifty LL12-encoded proteins could be assigned putative functions, as shown in Table 4.2 and Fig. 4.2B. Major components for head morphogenesis including capsid protein (gp59), prohead protease (gp62), portal protein (gp63), and the large terminase subunit (TerL, gp64) were identified. The components for tail morphogenesis including baseplate

Table 4.2. Proteins encoded by genome of phage LL12

Gene	Start	End	Strand	Gene product
CPT_LL12_001	1	1983	-	rIIA protector from prophage-induced early lysis
CPT_LL12_002	1980	2480	-	Hypothetical conserved protein
CPT_LL12_003	2530	2910	-	Hypothetical conserved protein
CPT_LL12_004	2920	4104	-	MoxR ATPase
CPT_LL12_005	4104	4373	-	Hypothetical conserved protein
CPT_LL12_006	4418	4798	-	Hypothetical conserved protein
CPT_LL12_007	4801	5058	-	Hypothetical conserved protein
CPT_LL12_008	5058	5414	-	Hypothetical conserved protein
CPT_LL12_009	5428	5700	-	Hypothetical conserved protein
CPT_LL12_010	5703	6602	-	Putative alpha 1-3 fucosyltransferase
CPT_LL12_011	6745	7722	+	Anti-sigma factor
CPT_LL12_012	7756	9216	+	Putative metallopeptidase
CPT_LL12_013	9232	9495	+	Hypothetical conserved protein
CPT_LL12_014	9495	9758	+	Hypothetical conserved protein
CPT_LL12_015	9758	10021	+	Hypothetical conserved protein
CPT_LL12_016	10023	10256	+	Hypothetical conserved protein
CPT_LL12_017	10272	10721	+	Hypothetical conserved protein
CPT_LL12_018	10718	10963	+	Hypothetical conserved protein
CPT_LL12_019	10966	11532	+	Hypothetical conserved protein
CPT_LL12_020	11588	11863	+	Hypothetical conserved protein
CPT_LL12_021	11866	12339	+	Hypothetical conserved protein
CPT_LL12_022	12336	12506	+	Hypothetical conserved protein
CPT_LL12_023	12537	12917	+	Hypothetical conserved protein
CPT_LL12_024	13081	13488	+	Hypothetical conserved protein
CPT_LL12_025	13492	13692	+	Hypothetical conserved protein
CPT_LL12_026	13740	19376	-	Hypothetical conserved protein
CPT_LL12_027	19425	21170	-	Tail fiber protein
CPT_LL12_028	21180	21446	-	Hypothetical conserved protein
CPT_LL12_029	21458	22501	-	Tail fiber protein
CPT_LL12_030	22591	22905	-	Hypothetical conserved protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_031	22917	23501	-	Tail fiber assembly protein
CPT_LL12_032	23516	24556	-	Tail fiber protein
CPT_LL12_033	24569	26641	-	Tail fiber protein
CPT_LL12_034	26641	27327	-	Hypothetical conserved protein
CPT_LL12_035	27339	28829	-	Baseplate protein
CPT_LL12_036	28935	32534	-	Tail fiber protein
CPT_LL12_037	32534	32959	-	Hypothetical conserved protein
CPT_LL12_038	32959	33510	-	Hypothetical conserved protein
CPT_LL12_039	33512	34102	-	Hypothetical conserved protein
CPT_LL12_040	34200	34532	-	Hypothetical conserved protein
CPT_LL12_041	34532	38971	-	Tail fiber protein
CPT_LL12_042	39007	41604	-	Tail fiber protein
CPT_LL12_043	41606	42274	-	Hypothetical conserved protein
CPT_LL12_044	42284	43000	-	Hypothetical conserved protein
CPT_LL12_045	43013	43690	-	Hypothetical conserved protein
CPT_LL12_046	43690	44694	-	Hypothetical conserved protein
CPT_LL12_047	44694	45074	-	Hypothetical conserved protein
CPT_LL12_048	45086	45970	-	Hypothetical conserved protein
CPT_LL12_049	46075	48414	-	Hypothetical conserved protein
CPT_LL12_050	48469	48717	-	Hypothetical conserved protein
CPT_LL12_051	48729	49202	-	Hypothetical conserved protein
CPT_LL12_052	49366	49839	-	Hypothetical conserved protein
CPT_LL12_053	49850	51226	-	Tail sheath protein
CPT_LL12_054	51299	51850	-	Hypothetical conserved protein
CPT_LL12_055	51850	52275	-	Hypothetical conserved protein
CPT_LL12_056	52291	52749	-	Hypothetical conserved protein
CPT_LL12_057	52803	53423	-	Hypothetical conserved protein
CPT_LL12_058	53483	54307	-	Hypothetical conserved protein
CPT_LL12_059	54392	55399	-	Major capsid protein
CPT_LL12_060	55448	55837	-	Head stabilization/decoration protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_061	55858	56835	-	Hypothetical conserved protein
CPT_LL12_062	56835	57254	-	Prohead protease
CPT_LL12_063	57322	58878	-	Portal protein
CPT_LL12_064	58981	60847	-	Terminase large subunit
CPT_LL12_065	60847	61134	-	O-spanin
CPT_LL12_066	61131	61529	-	I-spanin
CPT_LL12_067	61635	61904	-	Hypothetical conserved protein
CPT_LL12_075	62861	63199	-	Conserved hypothetical protein
CPT_LL12_076	63451	63561	-	Conserved hypothetical protein
CPT_LL12_077	63546	63788	-	Conserved hypothetical protein
CPT_LL12_078	63865	64053	-	Conserved hypothetical protein
CPT_LL12_079	64064	64642	-	Conserved hypothetical protein
CPT_LL12_080	65303	65653	-	Putative transcriptional regulator
CPT_LL12_081	65701	66255	-	Phosphoesterase
CPT_LL12_082	66263	66673	-	ATP-binding protein
CPT_LL12_083	66792	67709	+	RNA ligase and tail attachment protein
CPT_LL12_084	67706	68056	+	Hypothetical conserved protein
CPT_LL12_085	68065	68826	+	Putative Sir2-like protein
CPT_LL12_086	68840	69553	+	Putative Sir2-like protein
CPT_LL12_087	69616	69924	+	Conserved hypothetical protein
CPT_LL12_088	69911	70180	+	Conserved hypothetical protein
CPT_LL12_089	70180	71463	+	DNA ligase
CPT_LL12_090	71587	72057	+	Endolysin

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_091	72121	72327	+	Conserved hypothetical protein
CPT_LL12_092	72337	72801	+	HNH endonuclease
CPT_LL12_093	72794	73084	+	Conserved hypothetical protein
CPT_LL12_094	73144	73452	+	Conserved hypothetical protein
CPT_LL12_095	73445	74584	+	Exodeoxyribonuclease
CPT_LL12_096	74584	74916	+	Conserved hypothetical protein
CPT_LL12_097	74913	75542	+	Hypothetical conserved protein
CPT_LL12_098	75494	76099	+	EndoVII packaging and recombination endonuclease
CPT_LL12_099	76038	76259	+	Conserved hypothetical protein
CPT_LL12_100	76272	77255	+	Putative DNA polymerase/exonuclease
CPT_LL12_101	77301	78119	+	Putative DNA N6-adenine methyltransferase
CPT_LL12_102	78082	78540	+	Hypothetical conserved protein
CPT_LL12_103	78568	78759	+	Hypothetical conserved protein
CPT_LL12_104	78759	79373	+	Hypothetical conserved protein
CPT_LL12_105	79367	79960	+	Hypothetical conserved protein
CPT_LL12_106	79962	80162	+	Hypothetical conserved protein
CPT_LL12_107	80162	81145	+	Thymidylate synthase
CPT_LL12_108	81244	81705	+	Hypothetical conserved protein
CPT_LL12_109	81716	81946	+	Hypothetical conserved protein
CPT_LL12_110	81943	84258	+	Ribonucleoside triphosphate reductase alpha chain
CPT_LL12_111	84298	85386	+	Ribonucleoside diphosphate reductase beta chain
CPT_LL12_112	85390	85668	+	Glutaredoxin 1
CPT_LL12_113	85665	87788	+	Anaerobic ribonucleoside triphosphate reductase
CPT_LL12_114	87852	87968	+	Hypothetical conserved protein
CPT_LL12_115	87984	88088	+	Hypothetical protein
CPT_LL12_116	88120	88356	+	Hypothetical conserved protein
CPT_LL12_117	88353	88826	+	Anaerobic ribonucleoside triphosphate reductase activating protein
CPT_LL12_118	88883	88996	+	Hypothetical conserved protein
CPT_LL12_119	88993	89343	+	Hypothetical conserved protein
CPT_LL12_120	89382	90170	+	PhoH-like protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_121	90269	90580	+	Hypothetical conserved protein
CPT_LL12_122	90640	91689	+	Clp ATP-dependent protease subunit
CPT_LL12_123	91741	92274	+	DNA methyltransferase
CPT_LL12_124	92271	92555	+	Hypothetical conserved protein
CPT_LL12_125	92590	92829	+	Hypothetical conserved protein
CPT_LL12_126	92826	93224	+	Hypothetical conserved protein
CPT_LL12_127	93264	93497	+	Hypothetical conserved protein
CPT_LL12_128	93503	94027	+	Hypothetical conserved protein
CPT_LL12_129	94058	94294	+	Hypothetical conserved protein
CPT_LL12_130	94278	94430	+	Hypothetical conserved protein
CPT_LL12_131	94514	94699	+	Hypothetical conserved protein
CPT_LL12_132	94702	95157	+	Hypothetical conserved protein
CPT_LL12_133	95218	95565	+	Hypothetical conserved protein
CPT_LL12_134	95562	95849	+	Hypothetical conserved protein
CPT_LL12_135	95846	96040	+	Hypothetical conserved protein
CPT_LL12_136	96060	96245	+	Hypothetical conserved protein
CPT_LL12_137	96341	96637	+	Hypothetical conserved protein
CPT_LL12_138	96640	96921	+	Hypothetical conserved protein
CPT_LL12_139	96921	97145	+	Hypothetical conserved protein
CPT_LL12_140	97135	97389	+	Hypothetical conserved protein
CPT_LL12_141	97402	97800	+	Hypothetical conserved protein
CPT_LL12_142	97818	98108	+	Hypothetical conserved protein
CPT_LL12_143	98191	98508	+	Sigma 54 modulation factor
CPT_LL12_144	98646	98915	+	Hypothetical conserved protein
CPT_LL12_145	99007	99270	+	Hypothetical conserved protein
CPT_LL12_146	99275	99607	+	Hypothetical conserved protein
CPT_LL12_147	99607	100002	+	Hypothetical conserved protein
CPT_LL12_148	100082	100339	+	Hypothetical conserved protein
CPT_LL12_149	100340	100690	+	Hypothetical conserved protein
CPT_LL12_150	100690	101019	+	Hypothetical conserved protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_151	101019	101306	+	Hypothetical conserved protein
CPT_LL12_152	101396	101659	+	Hypothetical conserved protein
CPT_LL12_153	101688	101828	+	Hypothetical conserved protein
CPT_LL12_154	101815	102066	+	Hypothetical conserved protein
CPT_LL12_155	102095	102400	+	Hypothetical conserved protein
CPT_LL12_156	102453	102779	+	Hypothetical conserved protein
CPT_LL12_157	102788	103087	+	Hypothetical conserved protein
CPT_LL12_158	103355	103660	+	Hypothetical conserved protein
CPT_LL12_159	103782	103997	+	Hypothetical conserved protein
CPT_LL12_160	106517	106936	-	Hypothetical conserved protein
CPT_LL12_161	107082	107345	-	Hypothetical conserved protein
CPT_LL12_162	107406	107678	-	Hypothetical conserved protein
CPT_LL12_163	107897	108208	-	Hypothetical conserved protein
CPT_LL12_164	108205	108396	-	Hypothetical conserved protein
CPT_LL12_165	108413	108703	-	Hypothetical conserved protein
CPT_LL12_166	108795	109010	-	Hypothetical conserved protein
CPT_LL12_167	109020	109124	-	Hypothetical conserved protein
CPT_LL12_168	109202	109447	-	Hypothetical conserved protein
CPT_LL12_169	109540	109719	-	Hypothetical conserved protein
CPT_LL12_170	109795	110028	-	Hypothetical conserved protein
CPT_LL12_171	110083	110367	-	Hypothetical conserved protein
CPT_LL12_172	110570	110920	-	Hypothetical conserved protein
CPT_LL12_173	110938	111237	-	Hypothetical conserved protein
CPT_LL12_174	111331	111531	-	Hypothetical conserved protein
CPT_LL12_175	111679	111798	-	Hypothetical conserved protein
CPT_LL12_176	111819	112058	-	Hypothetical conserved protein
CPT_LL12_177	112018	112389	-	Hypothetical conserved protein
CPT_LL12_178	112454	112834	-	Hypothetical conserved protein
CPT_LL12_179	112923	113546	-	Hypothetical conserved protein
CPT_LL12_180	113622	113882	-	Hypothetical conserved protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_181	114102	114299	-	Hypothetical conserved protein
CPT_LL12_182	114289	114642	-	Hypothetical conserved protein
CPT_LL12_183	114717	114962	-	Hypothetical conserved protein
CPT_LL12_184	115052	115393	-	Hypothetical conserved protein
CPT_LL12_185	115473	115592	-	Hypothetical conserved protein
CPT_LL12_186	115673	116068	-	Hypothetical conserved protein
CPT_LL12_187	116151	116549	-	Hypothetical conserved protein
CPT_LL12_188	116614	116811	-	Hypothetical conserved protein
CPT_LL12_189	117664	117810	+	Hypothetical conserved protein
CPT_LL12_190	117913	118242	-	Hypothetical conserved protein
CPT_LL12_191	118239	118433	-	Hypothetical conserved protein
CPT_LL12_192	118426	118671	-	Hypothetical conserved protein
CPT_LL12_193	118681	119013	-	Hypothetical conserved protein
CPT_LL12_194	119023	119214	-	Hypothetical conserved protein
CPT_LL12_195	119278	119472	-	Hypothetical conserved protein
CPT_LL12_196	119469	119669	-	Hypothetical conserved protein
CPT_LL12_197	119723	120181	-	Hypothetical conserved protein
CPT_LL12_198	120171	120596	-	Hypothetical conserved protein
CPT_LL12_199	120665	120862	-	Hypothetical conserved protein
CPT_LL12_200	120958	121374	-	Hypothetical conserved protein
CPT_LL12_201	121437	122663	-	Hypothetical conserved protein
CPT_LL12_202	122663	123028	-	Hypothetical conserved protein
CPT_LL12_203	123112	123357	-	Hypothetical conserved protein
CPT_LL12_204	123375	123677	-	Hypothetical conserved protein
CPT_LL12_205	123667	124056	-	Hypothetical conserved protein
CPT_LL12_206	124056	124328	-	Hypothetical conserved protein
CPT_LL12_207	124392	124646	-	Hypothetical conserved protein
CPT_LL12_208	124723	125094	-	Hypothetical conserved protein
CPT_LL12_209	125105	125299	-	Hypothetical conserved protein
CPT_LL12_210	125296	125523	-	Hypothetical conserved protein

Table 4.2. Continued

Gene	Start	End	Strand	Gene product
CPT_LL12_211	125520	125825	-	Hypothetical conserved protein
CPT_LL12_212	125854	128907	-	DNA polymerase A
CPT_LL12_213	128966	131044	-	DNA replicative helicase/primase
CPT_LL12_214	131034	131798	-	DNA cytosine methyltransferase
CPT_LL12_215	131850	132332	-	Hypothetical conserved protein
CPT_LL12_216	132345	132740	-	Hypothetical conserved protein
CPT_LL12_217	132740	133042	-	Hypothetical conserved protein
CPT_LL12_218	133103	133444	-	Hypothetical conserved protein
CPT_LL12_219	133441	134811	-	Helicase
CPT_LL12_220	134811	136016	-	rIIB-like protein

(gp35), tail sheath protein (gp53), and multiple predicted tail fiber proteins (gp27, gp29, gp32, gp33, gp36, gp41, gp42) were identified. DNA replication proteins such as DNA polymerase (gp212), DNA replicative helicase/primase (gp213), and a helicase (gp219) were also identified. The genes encoding the large terminase subunit (gp64) and DNA polymerase (gp212) are disrupted by predicted intron sequences. These introns appear to be relatively short (~275-325 bp) and do not contain any significant protein-coding ORFs. The boundaries of these introns were determined based on protein sequence similarity to homologous proteins found in other phages that were not disrupted by introns (AKU44155 in the case of gp64, and AKU44295 for gp212). Like other large myophages, genes responsible for phage lysis are distributed across the LL12 genome rather than co-localized to a contiguous cassette; the phage endolysin (gp90), i-spanin (gp66) and o-spanin (gp65) were identifiable but the phage holin could not be positively identified. Based on analysis of predicted protein sequences by BLASTp with a E value cutoff of 10^{-5}, LL12 is most closely related to other V5-like myophages, including rV5 (NC_011041) which shares 206 proteins with LL12, and ΦAPCEc02 (KR698074) which shares 204 proteins. LL12 is also more distantly related to the *E. coli* phage phi92 (NC_023693), with 48 common proteins detectable by BLASTp.

Like the related phages rV5, phi92 and ΦAPCEc02, phage LL12 encodes an extensive set of predicted tail fibers: gp27, gp29, gp32, gp33, gp36, gp41 and gp42 (Fig. 4.2B, Table 4.2). All seven of these LL12 tail fibers are similar to the tail fibers in found in rV5 and ΦAPCEc02, with protein identities ranging from 43%-100% (Table 4.3). Six of

Table 4.3. Analysis of LL12 tail fibers

LL12 tail fibers locus tag	BlastP hits				% of identical matches	E-value	% identity (Dice)	alignment quality
	Phage	Phage Accession #	Protein Accession #	Gene product				
CPT_LL12_027	rV5	NC_011041.1	YP_002003530.1	gp028	99.0	0	0.99	
	APCEc02	KR698074.1	AKO61946.1		98.0	0	0.98	
	phi92	NC_023693.1	YP_009012483.1	gp151	47.0	7E-23	0.08	weak, partial
	phi92	NC_023693.1	YP_009012482.1	gp150	43.9	2E-22	0.07	weak, partial
CPT_LL12_029	APCEc02	KR698074.1	AKO61944.1		99.1	0	0.99	
	rV5	NC_011041.1	YP_002003532.1	gp30	98.9	0	0.99	
	rV5	NC_011041.1	YP_002003535.1	gp33	43.0	2E-90	0.44	
	APCEc02	KR698074.1	AKO61941.1		43.0	2E-90	0.44	
	phi92	NC_023693.1	YP_009012479.1	gp147	23.5	6E-18	0.25	full length
CPT_LL12_032	APCEc02	KR698074.1	AKO61941.1		99.4	0	0.99	
	rV5	NC_011041.1	YP_002003535.1	gp33	99.1	0	0.99	
	APCEc02	KR698074.1	AKO61944.1		43.3	6E-91	0.44	
	rV5	NC_011041.1	YP_002003532.1	gp30	42.7	1E-88	0.44	
	phi92	NC_023693.1	YP_009012479.1	gp147	27.0	9E-29	0.30	N- and C-terminus only
CPT_LL12_033	rV5	NC_011041.1	YP_002003536.1	gp34	99.9	0	1.00	
	APCEc02	KR698074.1	AKO61940.1		99.7	0	1.00	
	phi92	NC_023693.1	YP_009012474.1	gp142	42.2	0	0.41	good, almost full length
CPT_LL12_036	APCEc02	KR698074.1	AKO61937.1		99.0	0	0.99	
	rV5	NC_011041.1	YP_002003539.1	gp37	83.6	0	0.86	

Table 4.3. Continued

LL12 tail fibers locus tag	BlastP hits			gene product	% of identical matches	E-value	% identity (Dice)	alignment quality
	Phage	Phage Accession #	Protein Accession #					
CPT_LL12_041	APCEc02	KR698074.1	AKO61932.1		95.9	0	0.96	
	rV5	NC_011041.1	YP_002003543.1	gp41	91.5	0	0.70	
	rV5	NC_011041.1	YP_002003545.1	gp43	38.4	3E-15	0.04	
	APCEc02	KR698074.1	AKO61930.1		38.4	3E-15	0.04	
	phi92	NC_023693.1	YP_009012483.1	gp151	37.7	3E-07	0.02	very weak, patchy
CPT_LL12_042	rV5	NC_011041.1	YP_002003544.1	gp42	99.4	0	0.99	
	APCEc02	KR698074.1	AKO61931.1		99.2	0	0.99	
	phi92	NC_023693.1	YP_009012473.1	gp141	34.95	2E-173	0.36	good, almost full length

these seven proteins are also detectable in the more distantly-related phage phi92, with three of these, gp29, gp33 and gp42 producing alignments to nearly the full-length phi92 proteins 147, 142 and 141, respectively (Table 4.3). CryoEM reconstructions of phi92 have indicated that this phage possesses multiple sets of tail fibers that are mounted to the baseplate in downward, sideward, and upward orientations (Schwarzer *et al.*, 2012). These multiple tail fibers may contribute to a broadened host range in this phage and its relatives (Schwarzer *et al.*, 2012). The electron density of the downward-facing tail fiber was assigned to gp143, which is not conserved in LL12 (Schwarzer *et al.*, 2012). LL12 gp27 shows weak similarity to the N-terminus of phi92 gp150, which is predicted to form downward-facing tail spikes in cryoEM reconstructions (Schwarzer *et al.*, 2012). LL12 gp41 also possesses similarity to rV5 gp41 (Table 4.3), however LL12 gp41 is missing the C-terminal chaperone of endosialidase domain (pfam13884) of rV5 gp41 spanning residues 1151-1200.

Host range determination for phages LL5 and LL12

Infection by STEC strains can result in watery or bloody diarrhea, hemolytic uremic syndrome, microangiopathic hemolytic anemia and thrombocytopenia (Gyles, 2007). *E. coli* strains belonging to several pathotypes tend to be clonal and are grouped as serotypes based on the O-antigens (lipopolysaccharide) and H-antigens (flagella) (Kaper *et al.*, 2004). As phage LL12 bears similarity to phages rV5 and ΦAPCEc02, both of which infect STEC serotype O157:H7 (Kropinski *et al.*, 2013; Dalmasso *et al.*, 2016),

we sought to determine if phages LL5 and LL12 are also able to infect STEC representatives.

Phages LL5 and LL12 were spotted on soft agar overlays of STEC strains and EOP compared to the Keio parental strain BW25113. Phage LL5 was unable to form plaques on any of the tested STEC strains, and phage LL12 exhibited EOPs of close to 1 on STEC strains with serotypes O157:H7, O145:NM, O121:H19, O111 and O121:H19, demonstrating a relatively broad host range among STEC serotypes for this phage (Table 4.4).

Development and optimization of screening assay

Multiplicity of Infection (MOI) is the ratio of the number of the phages to host cells in a culture. The purpose of the phage Keio screen was to identify host genes required for the phage to successfully infect, replicate within, and lyse their host cells. To determine this, it was imperative to optimize MOI for each phage as excessively high MOI's could result in bacterial growth inhibition if the phage were able to infect the cells but still not produce progeny, while too low MOI's could result in false positive results (Georgopoulos, 2006). Initially, the lowest input phage concentration required to control growth of parental BW25113 strain after 8 hr incubation at 37 °C was determined. A log higher phage concentration was applied in this screen so as to minimize false positives. The number of bacterial cells inoculated by the 96-pin replicator were determined by

Table 4.4. Host range of phages LL5 and LL12. Phage LL5 and LL12 were tested for their ability to infect Shiga toxin-producing *Escherichia coli* (STEC) by spotting serially diluted phages on the soft agar lawns of respective STEC isolates. The efficiency of plating (EOP) is relative to the number of plaques formed on the Keio collection parental *E. coli* strain BW25113. Cells marked with "-" indicate an EOP of less than 10^{-7} (insensitive to phage). The data is the average of two biological replicates.

STEC serotype	Isolate ID	Phage LL5 EOP	Phage LL12 EOP	LPS Core types[b]
Not STEC	BW25113	1.0	1.0	K-12
O157:H7[a]	USDA-FSIS 380-94	-	0.8	R3
O104:H21	ATCC BAA-178	-	-	
O145:NM[a]	83-75	-	0.7	R1, K-12
O26:H11[a]	H30	-	-	R3
O111:H-[a]	JBI-95	-	-	R3
O121:H19	ATCC BAA-2219	-	0.7	
O146	ATCC BAA-2217	-	1.0	
O103:H11	ATCC BAA-2215	-	-	
O145:Nonmotile	ATCC BAA-2192	-	-	R1, K-12
O26:H11	ATCC BAA-2196	-	-	R3
O45:H2	ATCC BAA-2193	-	-	
O103:H2[a]	CDC 90-3128	-	-	R3
O121:H19[a]	CDC 97-3068	-	0.6	
O45:H2[a]	CDC 96-3285	-	-	

[a]Sources of these isolates are described in (Kirsch *et al.*, 2014).
[b]LPS core types information obtained from (Amor *et al.*, 2000).

viable counts. Based upon the cells inoculated and PFU of phages used, the initial MOI of LL5 and LL12 in this screen was calculated to be 1.0 and 0.001, respectively.

Determination of genes required for phage propagation

The Keio collection consists of a total of 3,985 individual gene knockout mutants in the *E. coli* K-12 strain BW25113. Each gene knockout is represented twice in the collection (the results of two independent experiments) (Baba *et al.*, 2006), thus the total collection contains 7,970 mutants, with each independent gene knockout mutants represented with even and odd numbers. Phages LL5 and LL12 were screened against the entire odd-numbered series of 3,985 Keio single-gene knockouts as described above. *E. coli* mutants that were unable to support phage growth, as indicated by their growth to an OD_{550} of at least 0.2 or 0.11 at 8 h in the presence of phage LL5 or LL12, respectively, were considered positive hits in this initial screen. Using this selection criteria, 37 knockout mutants (21 mutants for each phage) were selected for further investigation (Tables 4.5, 4.6, 4.7). For each of these initial hits, the screening experiment was repeated using the same odd-numbered mutant and its even-numbered counterpart from the collection. From this second experiment, 11/21 mutants identified against LL5 and 9/21 mutants identified against LL12 were found to produce the same phenotype in at least one of the paired knockouts, and these were retained for further study (Tables 4.5, 4.6, 4.7).

Table 4.5. Summary of hits in Keio screening. Initial screening of the Keio collection against both phages LL5 and LL12 gave 21 hits. These initial hits were verified using both of the independently generated Keio mutants using the same screening procedure. For phages LL5 and LL12, 10 and 9 strains, respectively, were determined to be true positives, which were tested for plating efficiency of the respective phages. Phage LL5 had a reduced efficiency of plating (EOP) of at least ~10^{-2} , as determined by spot titer assay, in 9 strains. Only 6 of these strains showed similar plating defect in full plate assay. Only 4 of these strains could be P1 transduced and phage LL5 had a similar plating defect, as compared to the original Keio mutants, in 3 of the P1 transductants. Phage LL12 had a reduced EOP of at least ~10^{-2} in 4 strains, as determined by spot titer. All 4 of these strains showed similar plating defect in full plate assay. Only 1 of these strains could be P1 transduced and phage LL12 had similar plating defect in the P1 transductant, as compared to the original Keio strain. Whenever applicable, gene complementation to restore the plating defect was conducted on the P1 transduced strains. When the kanamycin resistance marker could not be transduced, gene complementation was assayed on the original Keio mutants.

Phages	Initial hits	Verified hits	Strains with plating defect		P1 transduced strains	P1 transduced strain with plating defect
			Spot titer	Full plate titer		
LL5	21	10	8	4	4	3
LL12	21	9	3	3	1	1

Table 4.6. Results of initial (untargeted) screening and targeted re-screening of phage LL5 against the Keio *E. coli* knockout collection. In the initial screen, all mutants yielding a positive result were screened a second time against both independently-generated gene knockouts present in the Keio collection, denoted as the representatives from the even- and odd-numbered plate sets. Mutants with a positive result from either set were then tested for their efficiency of plating (EOP) by both spot titer and full-plate titration methods. Mutants exhibiting a significant EOP defect (less than ~0.05) were used for further study. The presence of the appropriate gene deletion was confirmed by PCR and sequencing. One knockout from each even/odd pair was selected for P1 transduction of the *kan*-marked deletion into the parental *E. coli* strain BW25113 background, re-tested for EOP defects and complemented *in trans*. Selected gene knockouts from the collection that were not identified in the initial screen were targeted for re-screening (bottom panel). Mutants were cultured from the Keio collection, the presence of the appropriate gene deletion confirmed by PCR, and the EOP determined in the knockout and its complemented counterpart. Blank cells denote that data was not collected, usually because the desired EOP defects were not observed.

Initial screen of Keio collection

Gene deletion	Results of first screen	Results of second screen		EOP (spot titer)		EOP (plate titer)		Keio mutant used for further work	Deletion confirmed by PCR	Transducible by P1*vir*	EOP (plate titer) in P1 transductant	EOP (plate titer) in complemented strain
		Odd	Even	Odd	Even	Odd	Even					
nuoM	+	+	+	0.06	0.04	0.3	0.3					
yehQ	+	-	-									
ycdB	+	+	-	0.4	0.2							
ydcR	+	+	-	1	1							
ygcN	+	-	-									

Table 4.6. Continued.

Gene deletion	Results of first screen	Results of second screen		EOP (spot titer)		EOP (plate titer)		Keio mutant used for further work	Deletion confirmed by PCR	Transducible by P1*vir*	EOP (plate titer) in P1 transductant	EOP (plate titer) in complemented strain
		Odd	Even	Odd	Even	Odd	Even					
rof	+	-	-									
crcB	+	-	-									
ydjO	+	-	-									
holD	+	-	-									
aroC	+	-	-									
ydgL	+	-	-									
idi	+	-	-									
rfaP	+	+	+	$< 8 \times 10^{-8}$	1.0	$< 7.5 \times 10^{-8}$	1.1	Odd	Confirmed	Yes	$< 7.5 \times 10^{-8}$	2.2
rfaG	+	+	-	1.1	0.3				Confirmed			
rfaF	+	+	+	0.3	0.03	1.3	0.3					
rfaQ	+	+	+	0.01	0.4	0.04	0.04	Odd	Confirmed	Yes	0.3	
rfaH	+	-	+	0.003	0.01	0.6	0.2					
rfaY	+	-	-									
secB	+	+	+	0.01	0.03	0.05	0.02	Odd	Confirmed	Yes	0.06	0.2
ppiB	+	+	-	0.003	0.3	0.04	0.4	Odd	Confirmed	Yes	0.09	1.5
ylaC	+	+	-	0.07	0.03	0.4	0.8					

Table 4.6 Continued.
Targeted re-screening

Gene deletion	Deletion confirmed by PCR	Keio mutant used for further work	EOP (spot titer)	EOP (plate titer)	EOP (plate titer) in complemented strain
lpcA	Confirmed	Even	$< 8 \times 10^{-8}$	$< 7.5 \times 10^{-8}$	0.8
rfaE	Confirmed	Even	$< 8 \times 10^{-8}$	$< 7.5 \times 10^{-8}$	1.0
rfaC	Loci intact				

Table 4.7. Results of initial (untargeted) screening and targeted re-screening of phage LL12 against the Keio *E. coli* knockout collection. In the initial screen, all mutants yielding a positive result were screened a second time against both independently-generated gene knockouts present in the Keio collection, denoted as the representatives from the even- and odd-numbered plate sets. Mutants with a positive result from either set were then tested for their efficiency of plating (EOP) by both spot titer and full-plate titration methods. Mutants exhibiting a significant EOP defect (less than ~0.05) were used for further study. The presence of the appropriate gene deletion was confirmed by PCR and sequencing. One knockout from each even/odd pair was selected for P1 transduction of the *kan*-marked deletion into the parental *E. coli* strain BW25113 background, re-tested for EOP defects and complemented *in trans*. Selected gene knockouts from the collection that were not identified in the initial screen were targeted for re-screening (bottom panel). Mutants were cultured from the Keio collection, the presence of the appropriate gene deletion confirmed by PCR, and the EOP determined in the knockout and its complemented counterpart. Blank cells denote that data was not collected, usually because the desired EOP defects were not observed.

Initial screen of Keio collection

Gene deletion	Results of first screen	Results of second screen		EOP (spot titer)		EOP (plate titer)		Keio mutant used for further work	Deletion confirmed by PCR	Transducible by P1*vir*	EOP (plate titer) in P1 transductant	EOP (plate titer) in complemented strain
		Odd	Even	Odd	Even	Odd	Even					
pflC	+	-	-									
cusB	+	-	-									
ompR	+	+	-	0.5	0.5							
envZ	+	-	-									
ompC	+	+	+	0.3	0.3							
yncJ	+	-	-									

Table 4.7 Continued.

Gene deletion	Results of first screen	Results of second screen		EOP (spot titer)		EOP (plate titer)		Keio mutant used for further work	Deletion confirmed by PCR	Transducible by P1*vir*	EOP (plate titer) in P1 transductant	EOP (plate titer) in complemen-ted strain
		Odd	Even	Odd	Even	Odd	Even					
ycbL	+	-	-									
yqjB	+	nt	nt	1.9	1.1							
yaaW	+	-	-									
yggT	+	+	+	0.1	0.1							
rfaP	+	+	+	0.06	1.1	0.04	1.2	Odd	Confirmed	Yes	0.02	1.0
rfaD	+	-	-									
rfaG	+	+	-	7.1 x 10^{-6}	1.0	5.1 x 10^{-6}	1.3	Odd	Confirmed	No		1.1
rfaF	+	-	-									
rfaH	+	+	-	0.4	0.8							
rfaY	+	+	+	0.3	0.2							
rffE	+	+	-	0.2	0.1							
lpcA	+	-	+	0.8	< 4 x 10^{-9}	1.5	< 4.4 x 10^{-9}	Even	Confirmed	No		1.1
trmC	+	-	-									
ygbF	+	-	-									
etp	+	-	-									

Table 4.7 Continued

Targeted re-screening

Gene deletion	Deletion confirmed by PCR	Keio mutant used for further work	EOP (spot titer)	EOP (plate titer)	EOP (plate titer) in complemented strain
rfaE	Confirmed	Even	$< 4 \times 10^{-9}$	$< 4.4 \times 10^{-9}$	1.1
rfaI	Confirmed	Odd	0.6		
rfaB	Confirmed	Even	0.4		
rfaC	Loci intact				
rfaF	Loci intact				

The efficiency of plating (EOP) of phage LL5 on the retained mutant strains was determined by spot titer. The observed plating efficiency of phage LL5 was reduced by at least ~20-fold in eight mutants. This plating defect was confirmed by titration of LL5 in full plate assays, in which only four mutants showed an EOP reduction of ~20-fold or greater. In order to confirm the plating phenotype in a clean genetic background, the kanamycin resistance cassettes from these Keio mutants were transduced by P1 into the parental *E. coli* strain BW25113. Markers could be transduced from all four Keio mutants into the parental strain, and three showed a similar plating defect as the corresponding Keio mutant, indicating the phenotype was linked to the disrupted locus (Table 4.5). One mutant, *rfaQ*, showed a ~25-fold reduction in EOP in the Keio mutant but its P1 transductant exhibited only a very mild EOP defect of 0.3 (Table 4.6), despite having the *rfaQ* deletion confirmed in the transductant by PCR and sequencing. This suggests an abnormality or additional defect in the original *rfaQ* Keio mutant; this mutant was not examined further.

The same approach was applied to confirm the phenotypes of the Keio mutants identified from the screens against phage LL12. Spot titer assays showed that the EOP of phage LL12 was reduced in only three of the nine initially identified Keio mutants. This plating defect could be replicated via full plate plaque assay in all three mutants. Only one of these mutants could be P1 transduced into the parental strain BW25113 and same plating phenotype was observed in the P1 transductant as in the Keio mutant (Table 4.5). The other two Keio mutants were resistant to P1 infection and could not be transduced.

Based on the genes identified in these initial screens, additional mutants from the odd- and even-numbered Keio sets were subjected to targeted re-screening by directly determining the phage EOP by the spot method. For both phages, genes involved in LPS biosynthesis (*rfaP* for LL5, and *rfaP, rfaG* and *lpcA* for LL12, Tables 4.6, 4.7, 4.8) were identified and confirmed, but these genes represented only parts of the known biosynthetic pathway. Additional Keio mutants in *lpcA, rfaE, rfaC, rfaF, rfaI* and *rfaB* were confirmed by PCR and sequencing of the mutant locus; *lpcA, rfaE, rfaI* and *rfaB* were found to contain the appropriate deletions. Strong EOP defects ($< 10^{-8}$) were identified in the *lpcA* and *rfaE* mutants against LL5, and in *rfaE* in LL12 (an EOP defect in *lpcA* against LL12 was already identified in the initial screen) (Table 4.8). These mutants could not be transduced due to P1 resistance.

Five genes were determined to be required for efficient propagation of phage LL5, as evident by the reduced plating efficiency (Table 4.8). The Keio mutant strains deleted for genes *rfaP, lpcA* and *rfaE* showed severe plating defects with EOP of phage LL5 less than 8×10^{-8}. This plating defect was also observed on the P1 transduced *rfaP* mutant. The plating efficiency of phage LL5 could be restored when the respective genes were provided *in trans* (Table 4.8). All plating phenotypes were complemented in the P1 transduced strains when applicable, and in the original Keio strains when the mutations could not be transduced by P1 due to host resistance. Two additional Keio mutants, *secB* and *ppiB,* exhibited milder defects in supporting phage LL5 growth, with EOP

Table 4.8. *E. coli* genes required for phages LL5 and LL12 propagation. The genes required for phage infection cycle can be determined by testing the efficiency of plating. Five genes were found to be required for phage LL5 infection cycle, whereas 4 genes were required for phage LL12 infection cycle. The kanamycin resistance cassette in the Keio strain was P1 transduced into parental BW25113, whenever possible. The P1 transductants are denoted by "#" and the original Keio mutants are denoted by "*". The plating phenotype was complemented in P1 transductants, when applicable. The data represents average and standard deviation of three biological repeats.

Phages	Gene	EOP	Complemented EOP
LL5	*lpcA**	$< 8 \times 10^{-8}$	0.8 ± 0.2
	*rfaE**	$< 8 \times 10^{-8}$	1.0 ± 0.5
	rfaP$^{\#}$	$< 8 \times 10^{-8}$	2.2 ± 2.0
	secB$^{\#}$	0.06 ± 0.02	0.2 ± 0.04
	ppiB$^{\#}$	0.09 ± 0.05	1.5 ± 0.3
LL12	*lpcA**	$< 4.4 \times 10^{-9}$	1.1 ± 0.4
	*rfaE**	$< 4.4 \times 10^{-9}$	1.1 ± 0.5
	rfaP$^{\#}$	0.02 ± 0.01	1.0 ± 0.4
	*rfaG**	$5.1 \times 10^{-6} \pm 1.0 \times 10^{-6}$	1.1 ± 0.2

reductions of ~10- to 100-fold relative to the parental *E. coli* strain BW25113; these mutants could be transduced by P1 into the parental background and could also be complemented *in trans* (Table 4.8).

The three genes *lpcA*, *rfaE* and *rfaP* are parts of the LPS biosynthesis pathway. The severe plating defects associated with multiple genes in this pathway strongly indicate that phage LL5 uses LPS as its receptor, and LPS defects result in major blocks to phage infection. Of these three genes, only *rfaP* appeared as a "hit" in the initial screen of the 3,985 Keio mutants against phage LL5 (Table 4.6). RfaP adds phosphate or 2-aminoethyl diphosphate (PPEtN) to heptose (Hep) I of the inner core LPS, and the LPS of *rfaP* mutants does not have any phosphoryl substituents on Hep I or Hep II, and also lacks Hep III (Fig. 4.3) (Yethon *et al.*, 1998). Hep I is transferred to LPS from the nucleotide precursor molecule ADP-L-*glycero*-D-*manno*-heptose, which is synthesized in a separate pathway involving *lpcA*, *rfaE* and *rfaD* (Fig. 4.3) (Gronow & Brade, 2001). In the absence of either LpcA or RfaE, heptoseless LPS core is formed (Valvano *et al.*, 2000), whereas *rfaD* mutants can still incorporate the stereoisomer D-*glycero*-D-*manno*-heptose into the LPS (Coleman & Leive, 1979). Keio *lpcA* and *rfaE* mutants with heptoseless LPS core cannot be infected by bacteriophage P1, rendering P1 transduction of the genetic marker ineffective (Valvano *et al.*, 2000).

As the *rfaP* mutant is the last step in the LPS pathway with a strong observed plating defect for LL5, this raises a question about the role of Hep II phosphorylation and the

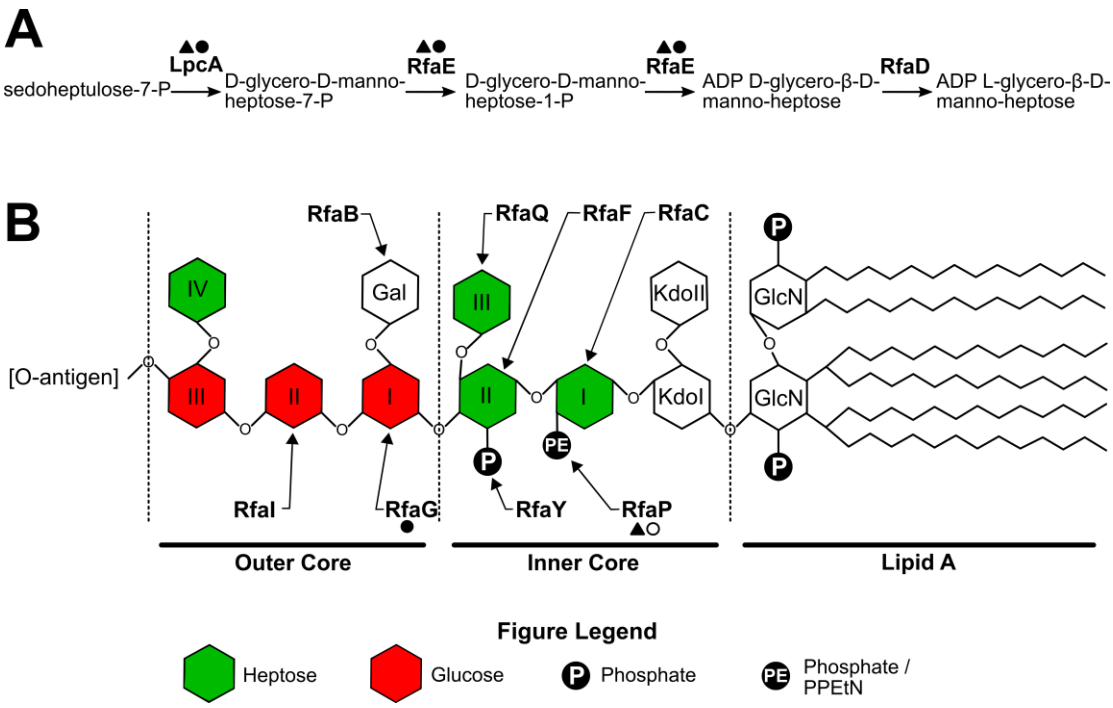


Fig. 4.3. Genes and biosynthetic pathway of the *E. coli* lipopolysaccharide (LPS) required for replication of phages LL5 and LL12. Proteins in the pathway are denoted in bold and label each step in biosynthesis. Black triangles indicate proteins that are required by LL5 for growth, and black circles indicate proteins required by LL12. Panel A: The nucleotide sugar precursor ADP L-glycero-β-D-manno-heptose is used as a substrate for the transfer of heptose (green in panel B) to the E. coli core LPS. ADP L-glycero-β-D-manno-heptose is synthesized from sedoheptulose-7-P via a pathway comprised of LpcA, RfaE and RfaD. Both LpcA and RfaE are required for growth of phages LL5 and LL12. Panel B: LPS is composed of four distinct domains: Lipid A, inner core, outer core and O-antigen. The enzymes responsible for the addition of sugar residues and phosphoryl constituents relevant to this study are denoted. RfaC, RfaF, RfaQ, RfaG, RfaI and RfaB add hexo or hepto sugar residues to LPS, and RfaP and RfaY add phosphoryl substituents to the heptose residues I and II, respectively. RfaP is required for infection by phage LL5 and the open circle by RfaP indicates a milder plating defect (~50-fold reduction in plating efficiency) for phage LL12 in the absence of this protein. RfaG is required for LL12 infection. Panel A is adapted from (Gronow & Brade, 2001) and panel B from (Clifton *et al.*, 2013).

addition of Hep III in the LL5 host recognition mechanism. RfaY adds the phosphoryl group to Hep II and RfaQ links Hep III to Hep II (Yethon *et al.*, 1998). The Keio *rfaY* mutant was ruled out during the initial screen and the plating defect of phage LL5 on the *rfaQ* transductant was mild (Table 4.6). From this information, it can be inferred that the plating defect of phage LL5 on the *rfaP* mutant is solely due to the absence of phosphoryl substituents on Hep I. Moreover, no other molecule is linked to the phosphoryl group of Hep I, thus it can be concluded that inner core of LPS is used by phage LL5 as its receptor and the phosphoryl group of Hep I is required by phage LL5 to infect its *E. coli* host. The LPS receptor requirement of phage LL5 appears similar to those of phage TLS (German & Misra, 2001). Phage TLS utilizes outer membrane protein TolC and the sugar residues Hep I (and its phosphoryl group), and Hep II of the LPS inner core to recognize and infect its *E. coli* host (German & Misra, 2001). The putative tail fibers of LL5 bear similarity to those of TLS, so it was intriguing why *tolC* did not come up as in screen. The requirement of *tolC* for LL5 infection still needs to be tested.

Apart from the three genes that were involved in LPS biosynthesis, the plating efficiency of phage LL5 was also reduced in Keio strains with deletion in two other genes, *secB* and *ppiB*. The EOP of phage LL5 was 0.06 and 0.09 in *secB* and *ppiB* strains, respectively. The plating defect in both strains could be restored to normal levels when the gene was supplied *in trans* (Table 4.8). Both SecB and PpiB are chaperones that

contribute to protein translocation and proline peptide bond isomerization, respectively (Bechtluft *et al.*, 2010; Unal & Steinert, 2014).

Proteins, once synthesized in cytoplasm, are sorted into compartments of the cell by different protein transport systems. SecB, a tetrameric cytoplasmic chaperone, is a component of the general secretory (Sec) system that transports proteins synthesized in the cytoplasm, post-translationally, to the extra-cytoplasmic compartments. Post-translational transport is primarily preferred for periplasmic and outer membrane proteins (Denks *et al.*, 2014; Findik & Randall, 2017). SecB binds to polypeptides and keeps them in an unfolded state until cytoplasmic ATPase SecA directs the bound polypeptide to the SecYEG transmembrane channel (Findik & Randall, 2017). The translocation of polypeptides across the SecYEG channel is powered by the cytoplasmic ATPase SecA (Denks *et al.*, 2014). Eighteen *E. coli* proteins have been reported to be dependent on SecB-mediated translocation (Bechtluft *et al.*, 2010) (Findik & Randall, 2017). We do not know if any of these SecB-dependent proteins play a role in the infection cycle of phage LL5, or if the potential accumulation of cytoplasmic protein aggregates in *secB* mutants hampers phage replication. In the absence of SecB, other cytoplasmic chaperones have been reported to be upregulated to stabilize secretory proteins during their delayed translocation and/or to rescue protein aggregates (Baars *et al.*, 2006). This compensatory mechanism by other chaperones may be the reason why the EOP defect of phage LL5 in the *secB* mutant is relatively mild (EOP = 0.06).

Another chaperone affecting the plating efficiency of phage LL5 is PpiB, which belongs to peptidyl-prolyl *cis*/*trans* isomerase (PPIase) superfamily of proteins, catalyzing protein folding at the peptide bonds preceding proline residues (Unal & Steinert, 2014). Although PPIases play a role in several biological processes, there is no evidence of any biological process depending solely on any PPIases (Unal & Steinert, 2014). The genome of *E*. *coli* K-12 encodes eight PPIases, belonging to three families: FKPBs, cyclophilins and parvulins (Unal & Steinert, 2014). The cyclophilins family consist of PpiA and PpiB, which are periplasmic and cytoplasmic proteins, respectively (Unal & Steinert, 2014). To our knowledge, there is only one reported instance of the requirement of a PPIase for a phage infection cycle: SlyD, belonging to FKBP family of PPIases, has been shown to be required for plaque formation by the ssDNA phage ΦX174 (Roof *et al.*, 1994). SlyD is required to stabilize ΦX174 lysis protein E, so that it can accumulate to optimum levels to lyse the host cell (Bernhardt *et al.*, 2002). Since the infection cycle of phage LL5 has not been characterized, it is difficult to explain which aspect of phage replication is affected by the absence of PpiB.

Four genes were determined to be required for efficient propagation of phage LL12, as evident by the plating defect (Table 4.8). All Keio strains in which phage LL12 showed plating defects were deleted for genes in the LPS biosynthesis pathway. Phage LL12 showed severe plating defects (EOP < 10^{-8}) in *lpcA* and *rfaE* deletions, and an EOP of ~10^{-6} in the *rfaG* deletion. The plating defect of phage LL12 was milder (EOP ~0.02) in the *rfaP* deletion (Table 4.8). The plating defects of phage LL12 could be restored when

the respective genes were supplied *in trans*. All complementation assays were conducted in the P1 transduced strains when applicable, and in the original Keio strains when the mutations could not be transduced by P1 (Table 3).

The functions of genes *lpcA*, *rfaE* and *rfaP* in LPS biosynthesis have been explained in context of phage LL5 above. RfaG links glucose (Glc) I to Hep II of the LPS inner core (Fig. 4.3) (Parker *et al.*, 1992; Yethon *et al.*, 2000), and marks the start of the outer core domain of the *E. coli* LPS. Sugar residues Glc II and galactose (Gal) are linked to Glc I by RfaI and RfaB respectively (Schnaitman & Klena, 1993). The plating efficiency of phage LL12 in the respective *rfaI* and *rfaB* mutants were close to wild type (~0.5) suggesting that Glc II and the Gal sidechain do not play significant roles in phage LL12 infection (Table 4.7). The strongly reduced EOP of phage LL12 on *rfaG* deletions suggests a crucial role of the outer core Glc I in the host recognition mechanism of phage LL12. This Hep II - Glc I linkage is conserved in K-12, and R1 - R4 LPS core types in *E. coli* (Amor *et al.*, 2000). As shown in Table 4.4, phage LL12 is able to infect *E. coli* strains with K-12, R1 and R3 LPS core types, which is consistent with the finding that the Gal sidechain residue linked to Glc I in the K-12 core and the residues downstream of Glc I are not involved in phage receptor binding. To our knowledge, LL12 is the first candidate from the group of V5-like phages for which the host receptor has been characterized. Based upon the sequence similarities of their tail fibers, other closely related V5-like phages such as rV5, ΦAPCEc02, and the O157:H7 typing phages 5 and 14 are likely to use the same or similar receptors as phage LL12.

From genetic analysis, we have established that phage LL12 recognizes the *E. coli* LPS core as its host receptor in a K-12 background. However, LL12 is also able to infect multiple different serotypes of *E. coli* with varying O-antigen (Table 4.4). In the case of phage P1, the extensive O-antigen expressed by hosts such as *E. coli* O157:H7 and *Salmonella* Typhimurium is able to obscure the LPS core and loss of the O-antigen results in bacterial sensitivity to this phage (Ornellas & Stocker, 1974; Ho & Waldor, 2007). This observation suggests that phage LL12 has developed a mechanism to deal with the presence of O-antigen that may obscure its receptor in the LPS core. Several phages are known to have evolved mechanisms to reach the cell surface to recognize these polysaccharide coats to facilitate infection. The tail spike protein (TSP) of *Salmonella* phage P22 recognizes O-antigen as its receptor and also has endorhamnosidase activity and cleaves its glycosidic linkages resulting in the shortening of the O-antigen (Andres *et al.*, 2010a; Andres *et al.*, 2010b). Coliphage G7C also expresses tail spikes with enzymatic activity against O-antigen that is involved in phage adsorption (Prokhorov *et al.*, 2017).

Conclusions

The Keio collection is a library of single-gene deletion of non-essential genes in *E. coli* K-12 strain BW25113. The library consists of 7970 mutants, with each mutant generated independently and distributed in odd- and even-numbered plates (Baba *et al.*, 2006). Phages LL5 and LL12 were initially isolated against pathogenic *E. coli* hosts as

candidates for therapeutic use, and phage LL12 was shown to infect representatives of several prominent STEC serovars. Both phages LL5 and LL12 were screened against the Keio library to investigate the host factors required for successful phage propagation. Initial screens suggested that twenty-one *E. coli* genes were necessary for phage LL5 and LL12 propagation (Table 4.5), but on further analysis a total of 5 and 3 *E. coli* genes were found to affect the propagation of phages LL5 and LL12, respectively (Tables 4.5, 4.8). Based on these figures, the screening process resulted in a false positive rate of 86%; the false negative rate of this screen cannot be calculated but target re-testing of genes of interest identified two additional genes not identified in the initial screens. These observations highlight the generally noisy nature of high-throughput screens and the requirement for additional confirmatory experiments following screening.

Through successive verification screen, spot and full plate plaque assays, it was established that three *E. coli* genes were needed for each of phage LL5 and LL12 propagation (Table 4.8), which gives a false positive rate of ~86% for both phages. Of the genes required for propagation of either phage LL5 or LL12, ~60% of the genes constitute the LPS biosynthesis pathway. It was noteworthy that all the genes pertaining to the LPS pathway did not come in the initial screen "hits", so targeted screen by spot-titer method was done for phage LL5 against Keio mutants *lpcA* and *rfaE*, and for phage LL12 against Keio mutant *rfaE* (Table 4.8); the gene disruption in Keio mutants *lpcA* and *rfaE* was verified beforehand. Both phages LL5 and LL12 exhibited plating defects on the respective Keio mutants.

The genes presented in this study may not be exhaustive list of genes necessary for phage LL5 or LL12 propagation because of limitations of these high throughput screens. If any biological pathway is comprised of multiple genes, it is likely that some genes will appear as “hits” in the initial screen and the role of other genes of that particular pathway can be tested by using verified gene mutants. The Keio collection, nevertheless, is a great library for initial screen, but data should be interpreted only after exhaustive follow-up of the mutants.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Co-evolution of phages and their bacterial hosts have resulted in development of several defense and counter-defense strategies (Chapter I). This dissertation focuses on the study of antirestriction system of coliphage P1, employed to overcome bacterial type I restriction and modification (R-M) systems (Chapters II and III). Two components of P1 antirestriction system, DarA and DarB, were described previously (Iida *et al.*, 1987). From genetic analysis, it was demonstrated that DarA protects P1 DNA from type I *Eco*A, whereas DarB protects P1 DNA from type I *Eco*B and *Eco*K systems (Iida *et al.*, 1987). Chapter II of this dissertation provides evidence suggesting P1 antirestriction is comprised of other components besides DarA and DarB. It has been shown that DarB and Ulx are required for protection against *Eco*B and *Eco*K systems, whereas Hdf, DarA, and DdrA are required for protection against *Eco*A, *Eco*B and *Eco*K systems. Interestingly, disruption of *ddrB* provided increased protection against *Eco*B and *Eco*K systems. Biochemical evidence supported by genetic analysis, suggests that components of P1 antirestriction system are incorporated into P1 virions in a definite order. Hdf and DarA are incorporated first, followed by DdrA. DarB, Ulx and DdrB are incorporated next. Specific order of incorporation of DarB, Ulx and DdrB is not known, but it seems that Ulx can be incorporated only in the presence of DarB (Chapter II). The role of Hdf and DarA in capsid morphogenesis has also been demonstrated in Chapter II. When either *hdf* or *darA* is disrupted, ~80% of the virion progenies have aberrant small heads.

The roles played by Hdf and DarA in head-size determination and incorporation of antirestriction proteins suggest that P1 antirestriction system is linked to capsid morphogenesis.

Chapter III further explores P1 antirestriction system following upon information provided in Chapter II. Since the P1 antirestriction components were encoded in two operons separated by genes of unrelated functions, it was hypothesized that other regions of P1 could encode for other unknown components of the P1 antirestriction system. To determine if there are any other components of P1 antirestriction system, isogenic knockouts of P1 genes of unknown functions, were constructed. Out of 29 genes deleted, 24 were found not to be required for P1 propagation at tested laboratory conditions. Isogenic P1 mutants disrupted for these 24 genes were tested for restriction phenotype by plating assays. None of these 24 genes were found to be required for antirestriction function. However, five genes *pmgA*, *pmgB*, *pmgC*, *pmgG* and *pmgR* were found to be required for P1 propagation. The essentiality of these genes still need to be confirmed by *in trans* complementation. After the phenotype of these genes have been confirmed, lysates of these mutants can be observed under transmission electron microscopy to see if these mutants have any morphogenesis defect. Specific roles of these genes in virion morphogenesis, if any, can then be further studied.

To understand the mechanism of antirestriction, incorporation of DarB into P1 capsids was further studied (Chapter III). Evidence presented in Chapter III suggests that P1

antirestriction components are packaged before DNA. It has been demonstrated in Chapter II that proteins of P1 antirestriction system contribute to a significant mass in P1 capsids. Since P1Δ*darA* cannot package any antirestriction components, it could have greater internal volume compared to that of P1. As DNA is packaged into P1 procapsids by headful mechanism (Bachi & Arber, 1977; Lobocka *et al.*, 2004), it can be hypothesized that P1Δ*darA* can package more DNA because of the availability of more internal capsid volume. This can be tested by pulsed-field gel electrophoresis of phage DNA (Lingohr *et al.*, 2009).

SDS-PAGE analysis of procapsid/tail fractions of P1Δ*pacA* and P1Δ*darA_pacA* lysates indicate presence of novel proteins in lanes of both P1Δ*pacA* and P1Δ*darA_pacA*, which are missing in the lane of CsCl purified P1 (Chapter III). Proteins present in procapsids, but not in mature virions, are indicative of scaffolding core (Dokland, 1999). These novel proteins can be identified by mass spectrometry and if determined to be virion-associated, their role in capsid morphogenesis can be further explored.

Moreover, evidence suggests that incorporation of DarB into procapsid is guided a signal provided by N-terminal residues of DarB (Chapter III). By testing incorporation of recombinant foreign proteins fused at their N-terminus with N-terminal DarB residues into P1 virions, it can be determined if these residues are sufficient for incorporation. Moreover, it can also be determined if these residues are sufficient for ejection of proteins into host following infection. In this study, the first 9 and 30 N-terminal

residues of DarB have been fused to TetR-mCherry fusion protein. By testing the incorporation of these fusion proteins into P1Δ*darB*, the necessity and sufficiency of these selected N-terminal DarB residues for capsid targeting can be determined.

Even though genetic evidence has established the role of DarB in protecting P1 DNA from host restriction, biochemical activity of protection still needs to be determined *in vitro*. As protein purification protocol for DarB has been optimized, further biochemical assay can be conducted with purified DarB to elucidate biochemical mechanisms for its antirestriction activity. Since DarB can be incorporated into P1 capsids only in the presence of other P1 proteins Hdf, DarA, DdrA, it is possible that DarB is interacting directly with some of these proteins. Also, DarB might be interacting with host components such as type I R-M subunits and host chaperones for its activity. Since the affinity tag for DarB purification has been optimized, same tag can be exploited to determine interacting partners of DarB by conducting pull-down assays.

High throughput genetic screens were conducted for novel T1-like coliphage LL5 and rV5-like coliphage LL12 against the *E. coli* Keio collection to understand components of phage-host interactions. Receptors for both phages LL5 and LL12 were characterized and two chaperones, PpiB and SecB were demonstrated to be required for efficient propagation of phage LL5. Bioinformatic analysis suggests that both phages encode multiples sets of tail fibers. The receptor binding proteins of both phage LL5 and LL12

have not been characterized. Further experiments can be designed to characterize the receptor binding proteins of phages LL5 and LL12.

REFERENCES

Abremski, K. & L.W. Black, (1979) The function of bacteriophage T4 internal protein I in a restrictive strain of *Escherichia coli*. *Virology* **97**: 439-447.

Adams, M.H., (1959) *Bacteriophages*. Interscience Publishers, New York, NY.

Aksyuk, A.A. & M.G. Rossmann, (2011) Bacteriophage assembly. *Viruses* **3**: 172-203.

Amor, K., D.E. Heinrichs, E. Frirdich, K. Ziebell, R.P. Johnson & C. Whitfield, (2000) Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun* **68**: 1116-1124.

Andres, D., U. Baxa, C. Hanke, R. Seckler & S. Barbirz, (2010a) Carbohydrate binding of *Salmonella* phage P22 tailspike protein and its role during host cell infection. *Biochem Soc Trans* **38**: 1386-1389.

Andres, D., C. Hanke, U. Baxa, A. Seul, S. Barbirz & R. Seckler, (2010b) Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles *in vitro*. *J Biol Chem* **285**: 36768-36775.

Arber, W. & D. Dussoix, (1962) Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J Mol Biol* **5**: 18-36.

Arber, W. & D. Wauters-Willems, (1970) Host specificity of DNA produced by *Escherichia coli*. XII. The two restriction and modification systems of strain 15T$^-$. *Mol Gen Genet* **108**: 203-217.

Arisaka, F., (2005) Assembly and infection process of bacteriophage T4. *Chaos* **15**: 047502.

Atanasiu, C., T.J. Su, S.S. Sturrock & D.T. Dryden, (2002) Interaction of the ocr gene 0.3 protein of bacteriophage T7 with *Eco*KI restriction/modification enzyme. *Nucleic Acids Res* **30**: 3936-3944.

Baars, L., A.J. Ytterberg, D. Drew, S. Wagner, C. Thilo, K.J. van Wijk & J.W. de Gier, (2006) Defining the role of the *Escherichia coli* chaperone SecB using comparative proteomics. *Journal of Biological Chemistry* **281**: 10024-10034.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, *et al.*, (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Bachi, B. & W. Arber, (1977) Physical mapping of *Bgl*II, *Bam*HI, *Eco*RI, *Hind*III and *Pst*I restriction fragments of bacteriophage P1 DNA. *Mol Gen Genet* **153**: 311-324.

Bair, C.L., D. Rifat & L.W. Black, (2007) Exclusion of glucosyl-hydroxymethylcytosine DNA containing bacteriophages is overcome by the injected protein inhibitor IPI*. *J Mol Biol* **366**: 779-789.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, *et al.*, (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Barrangou, R. & L.A. Marraffini, (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* **54**: 234-244.

Bazinet, C. & J. King, (1988) Initiation of P22 procapsid assembly in vivo. *J Mol Biol* **202**: 77-86.

Bechtluft, P., N. Nouwen, S.J. Tans & A.J. Driessen, (2010) SecB--a chaperone dedicated to protein translocation. *Mol Biosyst* **6**: 620-627.

Bernhardt, T.G., W.D. Roof & R. Young, (2002) The *Escherichia coli* FKBP-type PPIase SlyD is required for the stabilization of the E lysis protein of bacteriophage phi X174. *Mol Microbiol* **45**: 99-108.

Bertani, G., (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* **62**: 293-300.

Bertani, G. & J.J. Weigle, (1953) Host controlled variation in bacterial viruses. *J Bacteriol* **65**: 113-121.

Bickle, T.A. & D.H. Kruger, (1993) Biology of DNA restriction. *Microbiol Rev* **57**: 434-450.

Black, L.W. & C. Ahmad-Zadeh, (1971) Internal proteins of bacteriophage T4D: their characterization and relation to head structure and assembly. *J Mol Biol* **57**: 71-92.

Black, L.W. & V.B. Rao, (2012) Structure, assembly, and DNA packaging of the bacteriophage T4 head. *Adv Virus Res* **82**: 119-153.

Blount, Z.D., (2015) The unexhausted potential of *E. coli*. *Elife* **4**.

Bondy-Denomy, J., A. Pawluk, K.L. Maxwell & A.R. Davidson, (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429-432.

Booth, W.T., C.R. Schlachter, S. Pote, N. Ussin, N.J. Mank, V. Klapper, *et al.*, (2018) Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS Omega* **3**: 760-768.

Botstein, D. & I. Herskowitz, (1974) Properties of hybrids between *Salmonella* phage P22 and coliphage lambda. *Nature* **251**: 584-589.

Boulanger, P., (2009) Purification of bacteriophages and SDS-PAGE analysis of phage structural proteins from ghost particles. *Methods Mol Biol* **502**: 227-238.

Brass, A.L., D.M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R.J. Xavier, *et al.*, (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **319**: 921-926.

Brussow, H. & R.W. Hendrix, (2002) Phage genomics: small is beautiful. *Cell* **108**: 13-16.

Burckhardt, J., J. Weisemann, D.L. Hamilton & R. Yuan, (1981) Complexes formed between the restriction endonuclease *Eco*K and heteroduplex DNA. *J Mol Biol* **153**: 425-440.

Calendar, R., (1988) *The Bacteriophages*, p. 614. Springer US, Plenum Press, New York.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer & T.L. Madden, (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Casjens, S.R., (2011) The DNA-packaging nanomotor of tailed bacteriophages. *Nat Rev Microbiol* **9**: 647-657.

Cherepanov, P.P. & W. Wackernagel, (1995) Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**: 9-14.

Choi, K.H., J. McPartland, I. Kaganman, V.D. Bowman, L.B. Rothman-Denes & M.G. Rossmann, (2008) Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol* **378**: 726-736.

Chopin, M.C., A. Chopin & E. Bidnenko, (2005) Phage abortive infection in *lactococci*: variations on a theme. *Curr Opin Microbiol* **8**: 473-479.

Clifton, L.A., M.W. Skoda, E.L. Daulton, A.V. Hughes, A.P. Le Brun, J.H. Lakey & S.A. Holt, (2013) Asymmetric phospholipid: lipopolysaccharide bilayers; a Gram-negative bacterial outer membrane mimic. *J R Soc Interface* **10**: 20130810.

Coleman, W.G., Jr. & L. Leive, (1979) Two mutations which affect the barrier function of the *Escherichia coli* K-12 outer membrane. *J Bacteriol* **139**: 899-910.

Cornelissen, A., P.J. Ceyssens, V.N. Krylov, J.P. Noben, G. Volckaert & R. Lavigne, (2012) Identification of EPS-degrading activity within the tail spikes of the novel *Pseudomonas putida* phage AF. *Virology* **434**: 251-256.

Cumby, N., A.M. Edwards, A.R. Davidson & K.L. Maxwell, (2012) The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol* **194**: 5012-5019.

Dalmasso, M., R. Strain, H. Neve, C.M. Franz, F.J. Cousin, R.P. Ross & C. Hill, (2016) Three New *Escherichia coli* Phages from the Human Gut Show Promising Potential for Phage Therapy. *PLoS One* **11**: e0156773.

Datsenko, K.A. & B.L. Wanner, (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**: 6640-6645.

Delcher, A.L., D. Harmon, S. Kasif, O. White & S.L. Salzberg, (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**: 4636-4641.

Denks, K., A. Vogt, I. Sachelaru, N.A. Petriman, R. Kudva & H.G. Koch, (2014) The Sec translocon mediated protein transport in prokaryotes and eukaryotes. *Mol Membr Biol* **31**: 58-84.

Deveau, H., R. Barrangou, J.E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, *et al.*, (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**: 1390-1400.

Dokland, T., (1999) Scaffolding proteins and their role in viral assembly. *Cell Mol Life Sci* **56**: 580-603.

Dowah, A.S.A. & M.R.J. Clokie, (2018) Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophys Rev* **10**: 535-542.

Durmaz, E. & T.R. Klaenhammer, (2007) Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. *J Bacteriol* **189**: 1417-1425.

Dussoix, D. & W. Arber, (1962) Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage lambda. *J Mol Biol* **5**: 37-49.

Echols, H., (1972) Developmental pathways for the temperate phage: lysis vs lysogeny. *Annu Rev Genet* **6**: 157-190.

Emond, E., B.J. Holler, I. Boucher, P.A. Vandenbergh, E.R. Vedamuthu, J.K. Kondo & S. Moineau, (1997) Phenotypic and genetic characterization of the bacteriophage

abortive infection mechanism AbiK from *Lactococcus lactis*. *Appl Environ Microbiol* **63**: 1274-1283.

Escherich, T., (1988) The intestinal bacteria of the neonate and breast-fed infant. 1884. *Rev Infect Dis* **10**: 1220-1225.

Fillol-Salom, A., R. Martínez-Rubio, R.F. Abdulrahman, J. Chen, R. Davies & J.R. Penadés, (2018) Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *The ISME Journal*.

Findik, B.T. & L.L. Randall, (2017) Determination of the intracellular concentration of the export chaperone SecB in *Escherichia coli*. *PLoS One* **12**: e0183231.

Fineran, P.C., T.R. Blower, I.J. Foulds, D.P. Humphreys, K.S. Lilley & G.P. Salmond, (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* **106**: 894-899.

Fokine, A. & M.G. Rossmann, (2014) Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **4**: e28281.

Garneau, J.R., F. Depardieu, L.C. Fortier, D. Bikard & M. Monot, (2017) PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci Rep* **7**: 8292.

Georgopoulos, C., (2006) Toothpicks, serendipity and the emergence of the *Escherichia coli* DnaK (Hsp70) and GroEL (Hsp60) chaperone machines. *Genetics* **174**: 1699-1707.

German, G.J. & R. Misra, (2001) The TolC protein of Escherichia coli serves as a cell-surface receptor for the newly characterized TLS bacteriophage. *J Mol Biol* **308**: 579-585.

Gill, J.J., J.D. Berry, W.K. Russell, L. Lessor, D.A. Escobar-Garcia, D. Hernandez, *et al.*, (2012) The Caulobacter crescentus phage phiCbK: genomics of a canonical phage. *BMC Genomics* **13**: 542.

Gill, J.J. & P. Hyman, (2010) Phage choice, isolation, and preparation for phage therapy. *Curr Pharm Biotechnol* **11**: 2-14.

Gill, J.J., E.J. Summer, W.K. Russell, S.M. Cologna, T.M. Carlile, A.C. Fuller, *et al.*, (2011) Genomes and characterization of phages Bcep22 and BcepIL02, founders of a novel phage type in *Burkholderia cenocepacia*. *J Bacteriol* **193**: 5300-5313.

Golomidova, A.K., E.E. Kulikov, N.S. Prokhorov, R.C. Guerrero-Ferreira, V.N. Ksenzenko, K.K. Tarasyan & A.V. Letarov, (2015) Complete genome sequences of T5-

related *Escherichia coli* bacteriophages DT57C and DT571/2 isolated from horse feces. *Arch Virol* **160**: 3133-3137.

Golomidova, A.K., E.E. Kulikov, N.S. Prokhorov, C. Guerrero-Ferreira Rcapital Es, Y.A. Knirel, E.S. Kostryukova, *et al.*, (2016) Branched Lateral Tail Fiber Organization in T5-Like Bacteriophages DT57C and DT571/2 is Revealed by Genetic and Functional Analysis. *Viruses* **8**.

Green, M.R. & J. Sambrook, (2012) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Gronow, S. & H. Brade, (2001) Lipopolysaccharide biosynthesis: which steps do bacteria need to survive? *J Endotoxin Res* **7**: 3-23.

Gross, R.J., T. Cheasty & B. Rowe, (1977) Isolation of bacteriophages specific for the K1 polysaccharide antigen of *Escherichia coli*. *J Clin Microbiol* **6**: 548-550.

Guidolin, A., J.M. Zingg & W. Arber, (1989a) Organization of the bacteriophage P1 tail-fibre operon. *Gene* **76**: 239-243.

Guidolin, A., J.M. Zingg, H. Lehnherr & W. Arber, (1989b) Bacteriophage P1 tail-fibre and *dar* operons are expressed from homologous phage-specific late promoter sequences. *J Mol Biol* **208**: 615-622.

Guzman, L.M., D. Belin, M.J. Carson & J. Beckwith, (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J Bacteriol* **177**: 4121-4130.

Gyles, C.L., (2007) Shiga toxin-producing *Escherichia coli*: an overview. *J Anim Sci* **85**: E45-62.

Hanlon, G.W., S.P. Denyer, C.J. Olliff & L.J. Ibrahim, (2001) Reduction in exopolysaccharide viscosity as an aid to bacteriophage penetration through *Pseudomonas aeruginosa* biofilms. *Appl Environ Microbiol* **67**: 2746-2753.

Hao, L., A. Sakurai, T. Watanabe, E. Sorensen, C.A. Nidom, M.A. Newton, *et al.*, (2008) Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature* **454**: 890-893.

Hatfull, G.F. & R.W. Hendrix, (2011) Bacteriophages and their genomes. *Curr Opin Virol* **1**: 298-303.

Hayashi, K., N. Morooka, Y. Yamamoto, K. Fujita, K. Isono, S. Choi, *et al.*, (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2**: 2006 0007.

Heinrich, J., M. Velleman & H. Schuster, (1995) The tripartite immunity system of phages P1 and P7. *FEMS Microbiol Rev* **17**: 121-126.

Hendrix, R.W., (2002) Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**: 471-480.

Ho, T.D. & M.K. Waldor, (2007) Enterohemorrhagic *Escherichia coli* O157:H7 *gal* mutants are sensitive to bacteriophage P1 and defective in intestinal colonization. *Infect Immun* **75**: 1661-1666.

Hochman, L., N. Segev, N. Sternberg & G. Cohen, (1983) Site-specific recombinational circularization of bacteriophage P1 DNA. *Virology* **131**: 11-17.

Hong, Y.R. & L.W. Black, (1993) Protein folding studies in vivo with a bacteriophage T4 expression-packaging-processing vector that delivers encapsidated fusion proteins into bacteria. *Virology* **194**: 481-490.

Hubacek, J. & S.W. Glover, (1970) Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J Mol Biol* **50**: 111-127.

Iida, S., (1984) Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. *Virology* **134**: 421-434.

Iida, S. & W. Arber, (1977) Plaque forming specialized transducing phage P1: isolation of P1CmSmSu, a precursor of P1Cm. *Mol Gen Genet* **153**: 259-269.

Iida, S., R. Hiestand-Nauer, H. Sandmeier, H. Lehnherr & W. Arber, (1998) Accessory genes in the *darA* operon of bacteriophage P1 affect antirestriction function, generalized transduction, head morphogenesis, and host cell lysis. *Virology* **251**: 49-58.

Iida, S., J. Meyer, K.E. Kennedy & W. Arber, (1982) A site-specific, conservative recombination system carried by bacteriophage P1. Mapping the recombinase gene *cin* and the cross-over sites *cix* for the inversion of the C segment. *EMBO J* **1**: 1445-1453.

Iida, S., M.B. Streiff, T.A. Bickle & W. Arber, (1987) Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of *darA*$^{-}$ phages. *Virology* **157**: 156-166.

Ikeda, H. & J. Tomizawa, (1968) Prophage P1, and extrachromosomal replication unit. *Cold Spring Harb Symp Quant Biol* **33**: 791-798.

Inclan, Y.F., A. Persat, A. Greninger, J. Von Dollen, J. Johnson, N. Krogan, *et al.*, (2016) A scaffold protein connects type IV pili with the Chp chemosensory system to mediate activation of virulence signaling in *Pseudomonas aeruginosa*. *Mol Microbiol* **101**: 590-605.

Jin, Y., S.M. Sdao, J.A. Dover, N.B. Porcek, C.M. Knobler, W.M. Gelbart & K.N. Parent, (2015) Bacteriophage P22 ejects all of its internal proteins before its genome. *Virology* **485**: 128-134.

Jones, P., D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, *et al.*, (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Kao, S.H. & W.H. McClain, (1980) Roles of bacteriophage T4 gene *5* and gene *s* products in cell lysis. *J Virol* **34**: 104-107.

Kaper, J.B., J.P. Nataro & H.L. Mobley, (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**: 123-140.

King, J., R. Griffin-Shea & M.T. Fuller, (1980) Scaffolding proteins and the genetic control of virus shell assembly. *Q Rev Biol* **55**: 369-393.

King, J., E.V. Lenk & D. Botstein, (1973) Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. *J Mol Biol* **80**: 697-731.

Kirsch, K.R., T.M. Taylor, D. Griffin, A. Castillo, D.B. Marx & L. Smith, (2014) Growth of Shiga toxin-producing *Escherichia coli* (STEC) and impacts of chilling and post-inoculation storage on STEC attachment to beef surfaces. *Food Microbiol* **44**: 236-242.

Kitagawa, M., T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga & H. Mori, (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res* **12**: 291-299.

Kliem, M. & B. Dreiseikelmann, (1989) The superimmunity gene sim of bacteriophage P1 causes superinfection exclusion. *Virology* **171**: 350-355.

Kondo, E. & S. Mitsuhashi, (1964) Drug resistance of enteric bacteria. IV. Active transducing bacteriophage P1 *CM* produced by the combination of R factor with bacteriophage P1. *J Bacteriol* **88**: 1266-1276.

Koonin, E.V., K.S. Makarova & F. Zhang, (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**: 67-78.

Kornberg, S.R., S.B. Zimmerman & A. Kornberg, (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. *J Biol Chem* **236**: 1487-1493.

Kropinski, A.M., T. Waddell, J. Meng, K. Franklin, H.W. Ackermann, R. Ahmed, *et al.*, (2013) The host-range, genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. *Virol J* **10**: 76.

Kruger, D.H. & T.A. Bickle, (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**: 345-360.

Labrie, S.J., J.E. Samson & S. Moineau, (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**: 317-327.

Laemmli, U.K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lambert, J.M., R.S. Bongers & M. Kleerebezem, (2007) Cre-*lox*-based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. *Appl Environ Microbiol* **73**: 1126-1135.

Laslett, D. & B. Canback, (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **32**: 11-16.

Lehman, I.R. & E.A. Pratt, (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. *J Biol Chem* **235**: 3254-3259.

Leiman, P.G., S. Kanamaru, V.V. Mesyanzhinov, F. Arisaka & M.G. Rossmann, (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**: 2356-2370.

Lennox, E.S., (1955) Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1**: 190-206.

Leon, L.M., S.D. Mendoza & J. Bondy-Denomy, (2018) How bacteria control the CRISPR-Cas arsenal. *Curr Opin Microbiol* **42**: 87-95.

Lindsay, J.A., A. Ruzin, H.F. Ross, N. Kurepina & R.P. Novick, (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol* **29**: 527-543.

Lingohr, E., S. Frost & R.P. Johnson, (2009) Determination of bacteriophage genome size by pulsed-field gel electrophoresis. *Methods Mol Biol* **502**: 19-25.

Liu, M., R. Deora, S.R. Doulatov, M. Gingery, F.A. Eiserling, A. Preston, *et al.*, (2002) Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. *Science* **295**: 2091-2094.

Lobocka, M.B., D.J. Rose, G. Plunkett, 3rd, M. Rusin, A. Samojedny, H. Lehnherr, *et al.*, (2004) Genome of bacteriophage P1. *J Bacteriol* **186**: 7032-7068.

Loenen, W.A., D.T. Dryden, E.A. Raleigh & G.G. Wilson, (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Res* **42**: 20-44.

Loenen, W.A. & N.E. Murray, (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol* **190**: 11-22.

Lowe, T.M. & P.P. Chan, (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**: W54-57.

Lu, M.J. & U. Henning, (1989) The immunity (*imm*) gene of *Escherichia coli* bacteriophage T4. *J Virol* **63**: 3472-3478.

Maillou, J. & B. Dreiseikelmann, (1990) The *sim* gene of *Escherichia coli* phage P1: nucleotide sequence and purification of the processed protein. *Virology* **175**: 500-507.

Maniloff, J. & H.W. Ackermann, (1998) Taxonomy of bacterial viruses: establishment of tailed virus genera and the order *Caudovirales*. *Arch Virol* **143**: 2051-2063.

Marchler-Bauer, A., Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, *et al.*, (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200-D203.

Mateu, M.G., (2013) Assembly, stability and dynamics of virus capsids. *Arch Biochem Biophys* **531**: 65-79.

Maynard, N.D., E.W. Birch, J.C. Sanghvi, L. Chen, M.V. Gutschow & M.W. Covert, (2010) A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLoS Genet* **6**: e1001017.

Meselson, M. & R. Yuan, (1968) DNA restriction enzyme from *E. coli*. *Nature* **217**: 1110-1114.

Mise, K. & W. Arber, (1976) Plaque-forming transducing bacteriophage P1 derivatives and their behaviour in lysogenic conditions. *Virology* **69**: 191-205.

Mullaney, J.M. & L.W. Black, (1996) Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J Mol Biol* **261**: 372-385.

Murray, N.E., (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**: 412-434.

Nelson, D., (2004) Phage taxonomy: we agree to disagree. *J Bacteriol* **186**: 7029-7031.

Noguchi, H., T. Taniguchi & T. Itoh, (2008) MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15**: 387-396.

Nordstrom, K. & A. Forsgren, (1974) Effect of protein A on adsorption of bacteriophages to *Staphylococcus aureus*. *J Virol* **14**: 198-202.

Ornellas, E.P. & B.A. Stocker, (1974) Relation of lipopolysaccharide character to P1 sensitivity in *Salmonella typhimurium*. *Virology* **60**: 491-502.

Park, T., D.K. Struck, J.F. Deaton & R. Young, (2006) Topological dynamics of holins in programmed bacterial lysis. *Proc Natl Acad Sci U S A* **103**: 19713-19718.

Parker, C.T., A.W. Kloser, C.A. Schnaitman, M.A. Stein, S. Gottesman & B.W. Gibson, (1992) Role of the *rfaG* and *rfaP* genes in determining the lipopolysaccharide core structure and cell surface properties of *Escherichia coli* K-12. *J Bacteriol* **174**: 2525-2538.

Parma, D.H., M. Snyder, S. Sobolevski, M. Nawroz, E. Brody & L. Gold, (1992) The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev* **6**: 497-510.

Parreira, R., S.D. Ehrlich & M.C. Chopin, (1996) Dramatic decay of phage transcripts in lactococcal cells carrying the abortive infection determinant AbiB. *Mol Microbiol* **19**: 221-230.

Pawluk, A., N. Amrani, Y. Zhang, B. Garcia, Y. Hidalgo-Reyes, J. Lee, *et al.*, (2016a) Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* **167**: 1829-1838 e1829.

Pawluk, A., A.R. Davidson & K.L. Maxwell, (2018) Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol* **16**: 12-17.

Pawluk, A., R.H. Staals, C. Taylor, B.N. Watson, S. Saha, P.C. Fineran, *et al.*, (2016b) Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol* **1**: 16085.

Piya, D., L. Vara, W.K. Russell, R. Young & J.J. Gill, (2017) The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Mol Microbiol* **105**: 399-412.

Prokhorov, N.S., C. Riccio, E.L. Zdorovenko, M.M. Shneider, C. Browning, Y.A. Knirel, *et al.*, (2017) Function of bacteriophage G7C esterase tailspike in host cell adsorption. *Mol Microbiol* **105**: 385-398.

Qimron, U., B. Marintcheva, S. Tabor & C.C. Richardson, (2006) Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci U S A* **103**: 19039-19044.

Rajaure, M., J. Berry, R. Kongari, J. Cahill & R. Young, (2015) Membrane fusion during phage lysis. *Proc Natl Acad Sci U S A* **112**: 5497-5502.

Raleigh, E.A. & G. Wilson, (1986) *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc Natl Acad Sci U S A* **83**: 9070-9074.

Ram, G., J. Chen, K. Kumar, H.F. Ross, C. Ubeda, P.K. Damle, *et al.*, (2012) *Staphylococcal pathogenicity* island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc Natl Acad Sci U S A* **109**: 16300-16305.

Rauch, B.J., M.R. Silvis, J.F. Hultquist, C.S. Waters, M.J. McGregor, N.J. Krogan & J. Bondy-Denomy, (2017) Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**: 150-158 e110.

Rifat, D., N.T. Wright, K.M. Varney, D.J. Weber & L.W. Black, (2008) Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. *J Mol Biol* **375**: 720-734.

Roof, W.D., S.M. Horne, K.D. Young & R. Young, (1994) *slyD*, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl *cis-trans*-isomerases. *J Biol Chem* **269**: 2902-2910.

Rosner, J.L., (1972) Formation, induction, and curing of bacteriophage P1 lysogens. *Virology* **48**: 679-689.

Samson, J.E., A.H. Magadan, M. Sabri & S. Moineau, (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**: 675-687.

Sauer, B., (1987) Functional expression of the *cre-lox* site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**: 2087-2096.

Schmidt, C., M. Velleman & W. Arber, (1996) Three functions of bacteriophage P1 involved in cell lysis. *J Bacteriol* **178**: 1099-1104.

Schnaitman, C.A. & J.D. Klena, (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* **57**: 655-682.

Schneider, C.A., W.S. Rasband & K.W. Eliceiri, (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671-675.

Scholl, D., S. Adhya & C. Merril, (2005) *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microbiol* **71**: 4872-4874.

Schwarzer, D., F.F. Buettner, C. Browning, S. Nazarov, W. Rabsch, A. Bethe, *et al.*, (2012) A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. *J Virol* **86**: 10384-10398.

Seed, K.D., (2015) Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathog* **11**: e1004847.

Seed, K.D., S.M. Faruque, J.J. Mekalanos, S.B. Calderwood, F. Qadri & A. Camilli, (2012) Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. *PLoS Pathog* **8**: e1002917.

Seed, K.D., D.W. Lazinski, S.B. Calderwood & A. Camilli, (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**: 489-491.

Segev, N., A. Laub & G. Cohen, (1980) A circular form of bacteriophage P1 DNA made in lytically infected cells of *Escherichia coli*. Characterization and kinetics of formation. *Virology* **101**: 261-271.

Shao, Q., J.T. Trinh, C.S. McIntosh, B. Christenson, G. Balazsi & L. Zeng, (2017) Lysis-lysogeny coexistence: prophage integration during lytic development. *Microbiologyopen* **6**.

Shevchenko, A., H. Tomas, J. Havlis, J.V. Olsen & M. Mann, (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856-2860.

Showe, M.K. & L.W. Black, (1973) Assembly core of bacteriophage T4: an intermediate in head formation. *Nat New Biol* **242**: 70-75.

Simpson, R.J., (2006) Bulk precipitation of proteins by ammonium sulfate. *CSH Protoc* **2006**.

Skorupski, K., J.C. Pierce, B. Sauer & N. Sternberg, (1992) Bacteriophage P1 genes involved in the recognition and cleavage of the phage packaging site (*pac*). *J Mol Biol* **223**: 977-989.

Skorupski, K., B. Sauer & N. Sternberg, (1994a) Faithful cleavage of the P1 packaging site (pac) requires two phage proteins, PacA and PacB, and two *Escherichia coli* proteins, IHF and HU. *J Mol Biol* **243**: 268-282.

Skorupski, K., N. Sternberg & B. Sauer, (1994b) Purification and DNA-binding activity of the PacA subunit of the bacteriophage P1 pacase enzyme. *J Mol Biol* **243**: 258-267.

Soding, J., A. Biegert & A.N. Lupas, (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**: W244-248.

Spoerel, N., P. Herrlich & T.A. Bickle, (1979) A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature* **278**: 30-34.

Steinbacher, S., S. Miller, U. Baxa, A. Weintraub & R. Seckler, (1997) Interaction of *Salmonella* phage P22 with its O-antigen receptor studied by X-ray crystallography. *Biol Chem* **378**: 337-343.

Stern, A. & R. Sorek, (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**: 43-51.

Sternberg, N., (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc Natl Acad Sci U S A* **87**: 103-107.

Sternberg, N. & D. Hamilton, (1981) Bacteriophage P1 site-specific recombination. I. Recombination between *lox*P sites. *J Mol Biol* **150**: 467-486.

Sternberg, N., B. Sauer, R. Hoess & K. Abremski, (1986) Bacteriophage P1 *cre* gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. *J Mol Biol* **187**: 197-212.

Streiff, M.B., S. Iida & T.A. Bickle, (1987) Expression and proteolytic processing of the *darA* antirestriction gene product of bacteriophage P1. *Virology* **157**: 167-171.

Studier, F.W., (1973) Analysis of bacteriophage T7 early RNAs and proteins on slab gels. *J Mol Biol* **79**: 237-248.

Studier, F.W. & N.R. Movva, (1976) SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J Virol* **19**: 136-145.

Summer, E.J., (2009) Preparation of a phage DNA fragment library for whole genome shotgun sequencing. *Methods Mol Biol* **502**: 27-46.

Susskind, M.M., D. Botstein & A. Wright, (1974a) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter *sie*A+ lysogens. *Virology* **62**: 350-366.

Susskind, M.M., A. Wright & D. Botstein, (1971) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. II. Genetic evidence for two exclusion systems. *Virology* **45**: 638-652.

Susskind, M.M., A. Wright & D. Botstein, (1974b) Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. IV. Genetics and physiology of *sie*B exclusion. *Virology* **62**: 367-384.

Tallent, S.M., T.B. Langston, R.G. Moran & G.E. Christie, (2007) Transducing particles of *Staphylococcus aureus* pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. *J Bacteriol* **189**: 7520-7524.

Taylor, D.N., D.H. Hamer & D.R. Shlim, (2017) Medications for the prevention and treatment of travellers' diarrhea. *J Travel Med* **24**: S17-S22.

Thomas, D. & P. Prevelige, Jr., (1991) A pilot protein participates in the initiation of P22 procapsid assembly. *Virology* **182**: 673-681.

Thomason, L.C., N. Costantino & D.L. Court, (2007) *E. coli* genome manipulation by P1 transduction. *Curr Protoc Mol Biol* **Chapter 1**: Unit 1 17.

Tock, M.R. & D.T. Dryden, (2005) The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**: 466-472.

Traub, F. & M. Maeder, (1984) Formation of the prohead core of bacteriophage T4 in vivo. *J Virol* **49**: 892-901.

Tribble, D.R., (2017) Resistant pathogens as causes of traveller's diarrhea globally and impact(s) on treatment failure and recommendations. *J Travel Med* **24**: S6-S12.

Unal, C.M. & M. Steinert, (2014) Microbial peptidyl-prolyl *cis*/*trans* isomerases (PPIases): virulence factors and potential alternative drug targets. *Microbiol Mol Biol Rev* **78**: 544-571.

Valentine, R.C., B.M. Shapiro & E.R. Stadtman, (1968) Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. *Biochemistry* **7**: 2143-2152.

Vallee, M. & J.B. Cornett, (1972) A new gene of bacteriophage T4 determining immunity against superinfecting ghosts and phage in T4-infected *Escherichia coli*. *Virology* **48**: 777-784.

Valvano, M.A., C.L. Marolda, M. Bittner, M. Glaskin-Clay, T.L. Simon & J.D. Klena, (2000) The *rfaE* gene from *Escherichia coli* encodes a bifunctional protein involved in

biosynthesis of the lipopolysaccharide core precursor ADP-L-glycero-D-manno-heptose. *J Bacteriol* **182**: 488-497.

Van Valen, L., (1973) A new evolutionary law. Evol. Theory 1: 1—30. 1974. Two modes of evolution. *Nature* **252**: 298-299.

Veesler, D. & J.E. Johnson, (2012) Virus maturation. *Annu Rev Biophys* **41**: 473-496.

Volkin, E., (1954) THE LINKAGE OF GLUCOSE IN COLIPHAGE NUCLEIC ACIDS1. *Journal of the American Chemical Society* **76**: 5892-5893.

Walker, D.H., Jr. & T.F. Anderson, (1970) Morphological variants of coliphage P1. *J Virol* **5**: 765-782.

Walkinshaw, M.D., P. Taylor, S.S. Sturrock, C. Atanasiu, T. Berge, R.M. Henderson, *et al.*, (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol Cell* **9**: 187-194.

Wingfield, P., (2001) Protein precipitation using ammonium sulfate. *Curr Protoc Protein Sci* **Appendix 3**: Appendix 3F.

Wood, W.B., (1966) Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol* **16**: 118-133.

Xu, M., D.K. Struck, J. Deaton, I.N. Wang & R. Young, (2004) A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. *Proc Natl Acad Sci U S A* **101**: 6415-6420.

Yarmolinsky, M.B., (2004) Bacteriophage P1 in retrospect and in prospect. *J Bacteriol* **186**: 7025-7028.

Yethon, J.A., D.E. Heinrichs, M.A. Monteiro, M.B. Perry & C. Whitfield, (1998) Involvement of *waaY*, *waaQ*, and *waaP* in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane. *J Biol Chem* **273**: 26310-26316.

Yethon, J.A., E. Vinogradov, M.B. Perry & C. Whitfield, (2000) Mutation of the lipopolysaccharide core glycosyltransferase encoded by *waaG* destabilizes the outer membrane of *Escherichia coli* by interfering with core phosphorylation. *J Bacteriol* **182**: 5620-5623.

Young, R., (2013) Phage lysis: do we have the hole story yet? *Curr Opin Microbiol* **16**: 790-797.

Young, R. & J.J. Gill, (2015) MICROBIOLOGY. Phage therapy redux--What is to be done? *Science* **350**: 1163-1164.

Yuan, R., T.A. Bickle, W. Ebbers & C. Brack, (1975) Multiple steps in DNA recognition by restriction endonuclease from *E. coli* K. *Nature* **256**: 556-560.

Zabeau, M., S. Friedman, M. Van Montagu & J. Schell, (1980) The *ral* gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in *E. coli*. *Mol Gen Genet* **179**: 63-73.

Zabrovitz, S., N. Segev & G. Cohen, (1977) Growth of bacteriophage P1 in recombination-deficient hosts of *Escherichia coli*. *Virology* **80**: 233-248.

Zerbino, D.R., (2010) Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* **Chapter 11**: Unit 11 15.