EVOLVED TO BREAK FREE: EVOLUTION AND DIVERSITY IN SINGLE GENE LYSIS

A Dissertation

by

KARTHIK REDDY CHAMAKURA

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
Committee Members,	Jennifer K. Herman
	Lanying Zeng
	Deborah A. Siegele
Head of Department,	Gregory D. Reinhart

May 2017

Major Subject: Biochemistry

ABSTRACT

Small lytic phages compelled by the same evolutionary pressures as large dsDNA phages have evolved mechanisms to lyse their hosts efficiently and at optimal times to ensure maximum fecundity. The ssDNA *Microviridae* and ssRNA *Leviviridae*, with extremely small genomes (<6kb), effect lysis with a single gene. The most studied single-gene lysis systems (SGL) are gene *E* of the canonical microvirus ϕX174 and genes A_2 and *L* of the paradigm ssRNA phages Qβ and MS2, respectively. All three lysis proteins lack muralytic activity, and two, A_2 and E, have been shown to function as "protein antibiotics" by acting as noncompetitive inhibitors of conserved peptidoglycan (PG) biosynthesis enzymes, MurA and MraY, respectively. The third, L of MS2, does not inhibit PG biosynthesis but instead is hypothesized to trigger host autolytic response through an unknown mechanism.

The work described in this dissertation identifies a chaperone involved in MS2 L-mediated lysis, the minimal lytic domain of L, and a new paradigm in SGL that resolves a long-standing debate in peptidoglycan biosynthesis. Using a genetic approach, *dnaJ* was shown to be required for L-mediated lysis. DnaJ was shown to form a complex with L and this interaction is abrogated in the absence of the highly basic N-terminal half of L. Mutational analysis of L has revealed that most critical residues are clustered around a highly-conserved LS motif. To determine if SGL evolved at different genomic locations target other steps in PG biosynthesis we studied of the lysis mechanism of Lys^M from an *E. coli* phage M. Both genetic and biochemical analysis have demonstrated that Lys^M

targets MurJ, the proposed Lipid II flippase candidate, and blocks Lipid II flipping to effect host lysis. Thus, the results presented in this dissertation provide the most convincing evidence in favor of MurJ being the Lipid II flippase. Additionally, the results highlight the plasticity of ssRNA phages to evolve inhibitors against proteins involved in PG biosynthesis and maintenance.

DEDICATION

This dissertation is dedicated to my late grandparents (Santhosh Reddy Chamakura and Ganga Reddy Baireddy)

ACKNOWLEDGEMENTS

This work would not have been possible without the contributions and support of several individuals. First, I would like to thank my mentor Dr. Ry Young for giving me an opportunity to work in his lab. I could not have asked for a better mentor and working with him was an amazing learning experience. His enthusiasm for research, teaching, and new ideas have enabled me to have a fabulous graduate experience. I would like to thank my committee members, Dr. Debbie Siegele, Dr. Jennifer Herman, Dr. Lanying Zeng, for their excellent outside guidance throughout the course of my graduate career. I would like to thank my outside collaborators Dr. Thomas G. Bernhardt, Dr. Natividad Ruiz, Dr. Lok-To Sham, Dr. Rebecca M. Davis, and Dr. Hongbaek Cho for their help in the results presented in Chapter 3. Special thanks to undergraduate student researchers Garrett Edwards, Kayla LeSage, Lorna Min, Jennifer Tran, Kalyn Weiss, Elizabeth Tran, and Zachary Williams for their contributions. Mentoring undergraduate students was one of the best experiences during my graduate career. I would like to thank my current and former lab mates, Dr. Samir Moussa, Dr. Gabby Kutty, Dr. Catrina Reed, Dr. Kenneth To, Dr. Manoj Rajaure, Rohit Kongari, Dr. Yi Chen, Adriana Hernandez, Jesse Cahill, and Jacob Lancaster, for their support and encouragement. I am very grateful for their help and very thankful to have worked with them. Special thanks to Daisy Wilbert, staff of Center for Phage Technology, and the staff of the Bio/Bio department. I want to acknowledge everybody who has contributed to my achievements either as a friend,

teacher, or by being a good colleague. Finally, I would like to thank my family for all their support without which this endeavor would not have been successful.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Ry Young, Dr. Jennifer Herman, and Dr. Lanying Zeng of the Department of Biochemistry and Biophysics and Dr. Deborah Siegele of the Department of Biology.

The labs of Dr. Thomas G. Bernhardt of the Department of Microbiology and Immunobiology at Harvard Medical School and Dr. Natividad Ruiz of the Department of Microbiology at Ohio State University provided part of the data shown in chapter 3. Dr. Lok-To Sham (Bernhardt lab) of the Department of Microbiology and Immunobiology at Harvard Medical School conducted the colicin M assays and *amj* rescue experiment depicted in Chapter 3. Dr. Rebecca M Davis (Ruiz lab) of the Department of Microbiology at Ohio State University constructed *murJ* haploid strains, performed their function, and expression tests and performed SCAM analysis.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

This work was made possible in part by NIH under Grant Number GM27099. The Center for Phage Technology at Texas A&M University jointly sponsored by Texas A&M AgriLife provided additional funding.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or Texas A&M Agrilife.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Viruses of bacteria or bacteriophages are natural biocontrol agents that play a critical role in maintaining and driving bacterial diversity in every environmental niche (1). In fact, it is estimated that phage outnumber bacteria by ~10-fold, with an estimated 10^{31} phages on the planet (2). This astounding number is a testament of their reproductive and evolutionary success. As obligate intracellular parasites of bacteria, phages have evolved a wide variety of highly efficient mechanisms to both enter and exit their host (3, 4).

As the most abundant and well-studied group of phages, the tailed phages of *Caudovirales*, begin their infection cycle by adsorbing to the host cell. Adsorption of a phage on the host entails a "ligand-receptor" binding between the phage tail fibers (or attachment proteins) and the surface features on the host (5, 6). Upon recognition of the host receptor, the phage tail machinery springs into action and locally breaches the host cell envelope to deliver its genetic material and accessory proteins (3). This local breach in the envelope is thought to be immediately repaired or plugged by host and/or phage proteins, albeit molecular evidence for either process is lacking (7). The molecules and mechanisms of phage DNA entry remain an understudied aspect of phage biology. Nevertheless, based on a handful of well-studied phages, it is evident that phages have evolved a variety of molecules to successfully breach the host cell envelope at the time of infection.

The time after DNA ejection and the appearance of the first virus-like particles in the host is referred to as the latent period. During this time the host biosynthetic machinery is hijacked by the phage to synthesize viral DNA and proteins. Four independent late-stage processes viz., head assembly, tail assembly, DNA packaging, and lysis, are initiated during the latent period. The phage particles begin to assemble and continue to accumulate intracellularly until a genetically programmed lysis event. Lysis of the host concludes the latent period of phage infection.

Lysis of the host results in release of progeny virions into the environment and the precise timing of this event has profound impact on the fecundity of the phage (8, 9). The genetic ablation of the lysis event has no effect on the rate of intracellular accumulation of phage particles. In fact for phage mutants with inactivated lysis genes, a 10-fold increase in progeny virions per cell has been reported (10, 11). Yet normally lysis occurs at an optimal lysis time that provides the phage an overall fitness advantage. To effect lysis, a phage must compromise the host envelope, which in Gram-negative bacteria is comprised of the asymmetric outer membrane (OM), a thin layer of highly cross-linked peptidoglycan (PG), and the cytoplasmic or inner membrane (IM). To date there are two broad strategies to effect lysis, one involving an intricate, precisely timed, step-by-step breakdown of each barrier of the host envelope (12). This is achieved by proteins that have evolved to specifically disrupt each layer of the host envelope. This system is reliant on a precise molecular clock (holins) that disrupts the cytoplasmic membrane at an allele-specific time. Once the cytoplasmic membrane is compromised, the peptidoglycan is degraded by endolysins and the OM is disrupted by spanins (13).

This system is employed by large dsDNA phages that have ample genomic space to evolve a variety of lysis proteins and their regulators (4) (Figure 1.1). The other strategy of lysis is employed by small lytic phages with genomes less than 6kb. Limited by their small genome size and compelled by the evolutionary pressure to lyse at an optimal time, they too have evolved genes to effect lysis. In contrast to dsDNA phages, these small viruses encode a single gene to overcome all the barriers of the host envelope (14). The work described below is concerned with "single gene lysis" of host cells by small lytic phages, the evolution of their lysis genes, and the diversity of the host targets and lysis mechanisms.

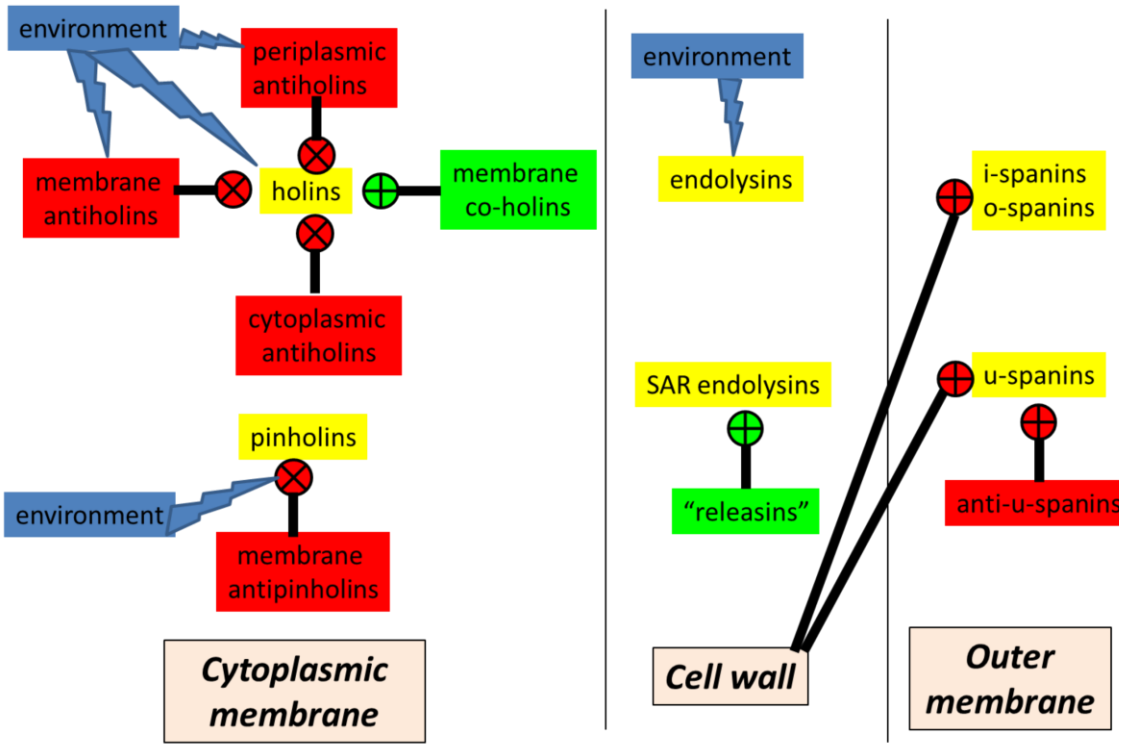

Figure 1.1 The various lysis proteins and lysis regulators of *Caudovirales*.
The cytoplasmic membrane, the cell wall (PG), and the OM barriers of the Gram-negative cell envelope are shown as vertical columns. The six different classes of lysis proteins (yellow) that compromise different layers of the cell envelope are shown in their respective columns. The positive and negative regulators of the lysis proteins are highlighted in green and red respectively. The lysis proteins that are capable of sensing environmental cues are indicated by blue lightning bolt.

Phages with small genome size

Bacteriophages with genome sizes <6 kb represent some of the simplest viruses in the biosphere. The genetic material of these viruses is composed of either single stranded or double stranded nucleic acids. Based on the type of nucleic acids, these viruses can be broadly classified into RNA and DNA phages. RNA genomes can be either ssRNA or dsRNA, while DNA viruses with small genomes are mostly ssDNA (15). From a historical standpoint, these viruses were initially isolated with the purpose of using them as model organisms for studying some basic biological processes such as DNA and RNA replication, transcription, translation, and macromolecular assembly (16, 17).

ssDNA phages

Based on the symmetry of viral structure, ssDNA phages are classified in two broad groups, as linear filamentous ssDNA phages and "cubic" ssDNA phages. Filamentous phages (M13, and f1) composed of circular ssDNA encased in long flexible tube were isolated as a group of male-specific phages. Instead of lysing their hosts, filamentous phages extrude through the host membrane at a steady rate. Since filamentous phages are not lytic, they will not be covered in this dissertation. On the other hand, small icosahedral bacteriophages with closed circular (+) sense ssDNA of the family *Microviridae* represent the most abundant and ubiquitous small genome phages (18, 19).

ssDNA viruses: the Microviridae

Microviridae is a family of small icosahedral phages with covalently closed circular (+) sense ssDNA genomes. They are ubiquitous, having even been found for intracellular bacteria such as *Chlamydia* (20, 21). The prototypic member of *Microviridae* is the famous *E. coli* phage ϕX174; its unique features/properties have enabled some of the most important achievements in molecular biology. The DNA genome of ϕX174 was the first DNA whole genome sequencing project to be completed (22, 23). Because of its simplicity and availability of complete genome sequence, it served as model system for studying processes such as DNA replication, transcription, gene expression, and viral assembly (17). Its DNA was also the first to undergo site-directed mutagenesis and the first fully synthetic genome (24, 25).

The 5386-nucleotide genome of ϕX174 encodes 11 genes, of which 9 are essential (26). There are four genes that are either completely or partly embedded in other genes, some as alternate starts in a single frame and others as alternate reading frames (Figure 1.2). The gene *A** is embedded within the same reading frame of gene *A*. Genes *B* and *E* are completely embedded in alternate reading frames of genes *A* and *D* respectively, while gene *K* overlaps both gene *A* and gene *C* (17).

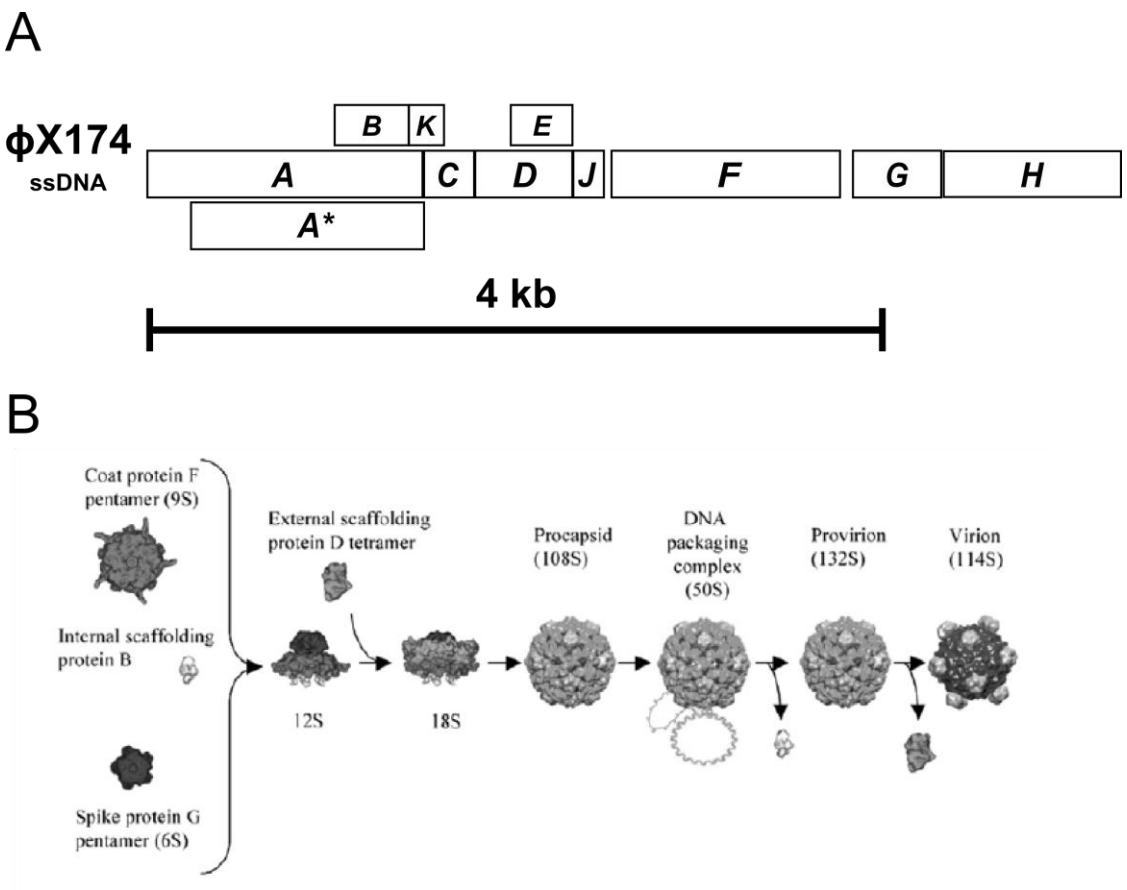

Figure 1.2 Genome organization and capsid assembly of ϕX174
(A) Genome map of ϕX174. (B) ϕX174 virion assembly pathway. This figure was adapted from (27).

The infection of ϕX174 begins with the recognition of terminal glucose and galactose residues in the LPS core of *E. coli* C (28). The common laboratory strain *E. coli* K-12 is naturally resistant to ϕX174 but it can be sensitized by altering the structure of LPS core (29). The glucose-binding site is present in the coat protein, made up of six conserved residues make up the binding site (30, 31). The binding of virions to the LPS core is reversible and dependent on the presence of Ca^{+2} ions, which stabilizes the side-chains of the binding site residues (32). The binding of glucose is a reversible reaction, which presumably allows the virus to roll along the surface of the host until proper recognition of the secondary receptor (33). The binding to secondary receptor is an irreversible reaction and results in the ejection of DNA from the capsid (34). The DNA ejection requires the pilot protein H, which is present in 10-12 copies per virion, at one of the 5-fold vertices (17). It was speculated that an oligomer of H-protein forms a DNA channel through the cell wall (35). Indeed, recent crystallographic and cryo-EM studies have shown that H-protein form a 22 Å conduit (tail-like structure) that traverses the entire length of the periplasmic space and contacts the IM (36). The central pore of the tail-like structure is wide enough to accommodate two strands of ssDNA. Upon penetration into the host cell, genome replication occurs and the process can be divided into three stages. In the first stage, the (+) sense ssDNA is converted into a circular dsDNA species known as the "replicative form I DNA" (RF I DNA) (37). During second stage, (+) sense DNA is amplified through rolling circle replication of RF I DNA (38). In the final stage, the circular ssDNA genome is synthesized and packaged into the procapsids. For viral genes to be transcribed and translated the negative strand synthesis

must first occur to serve as template for transcription. Unlike large ds DNA phages, temporal regulation of expression of genes is completely under the control of cis-acting elements like promoters and terminators, and Shine-Dalgarno (SD) sequences and also on mRNA structure and stability. The promoters for genes are found upstream of genes A, B, and D, and terminator sequences after genes J, F, G, and H. The terminators in ϕX174 are not efficient and allow leak-through transcription and translation (17). The lysis gene *E* is on all the transcripts, unlike in dsDNA phages where lysis genes are late genes (39, 40).

The viral morphogenesis begins with the formation of two types of pentamers, one composed of the spike protein G and the other made up of coat protein F (Figure 1.2) (27). The binding of internal scaffolding protein B to F pentamers (F_5) facilitates the interaction of G protein pentamers (G_5) and tetramers of the external scaffolding protein D to F_5 and B complex. Incorporation of H protein into the spike complex completes the formation of the procapsid. The packaging of DNA along with the highly basic J protein initiates the maturation process. During this, the B protein is extruded and the external scaffold protein D is released to form a mature virion (27).

ϕX174 has a burst size of ~200 virions per cell with cell lysis at ~20 after infection. The lysis protein E, a 91 aa protein with an essential N-terminal transmembrane domain (TMD) is responsible for effecting host lysis (41, 42). E-mediated lysis is described in detail in later sections of Chapter I.

dsRNA phages

The phage ϕ6, founding member of the family *Cystoviridae*, was isolated in 1973 from bean plants as a type IV pili-specific, lipid-containing, and segmented dsRNA phage (43, 44). Its genome contains three dsRNA segments designated as L, M, and S with sizes 6.5 kbp, 4 kbp, and 3 kbp respectively (Figure 1.3). The largest of the dsRNA segments, L, encodes the major structural protein (P1) of the inner core, RNA-dependent RNA polymerase (P2), packaging NTPase (P4), and accessory protein P7 (45). Together these four components form the dodecahedron inner core, which is encased in a shell of P8, encoded by gene 8 on segment S. The P8 shell is in turn surrounded by lipid-containing membrane that harbors proteins for host recognition and accessory proteins (Figure 1.3) (46). The assembly of particles begins with the formation of the inner core. The nucleation for inner core assembly begins with the interaction between tetramers of P1 and hexamers of P4 (47). At this stage, empty procapsids appear to be invaginated or deflated with 5-fold vertices sloping towards the center of the capsid (48).

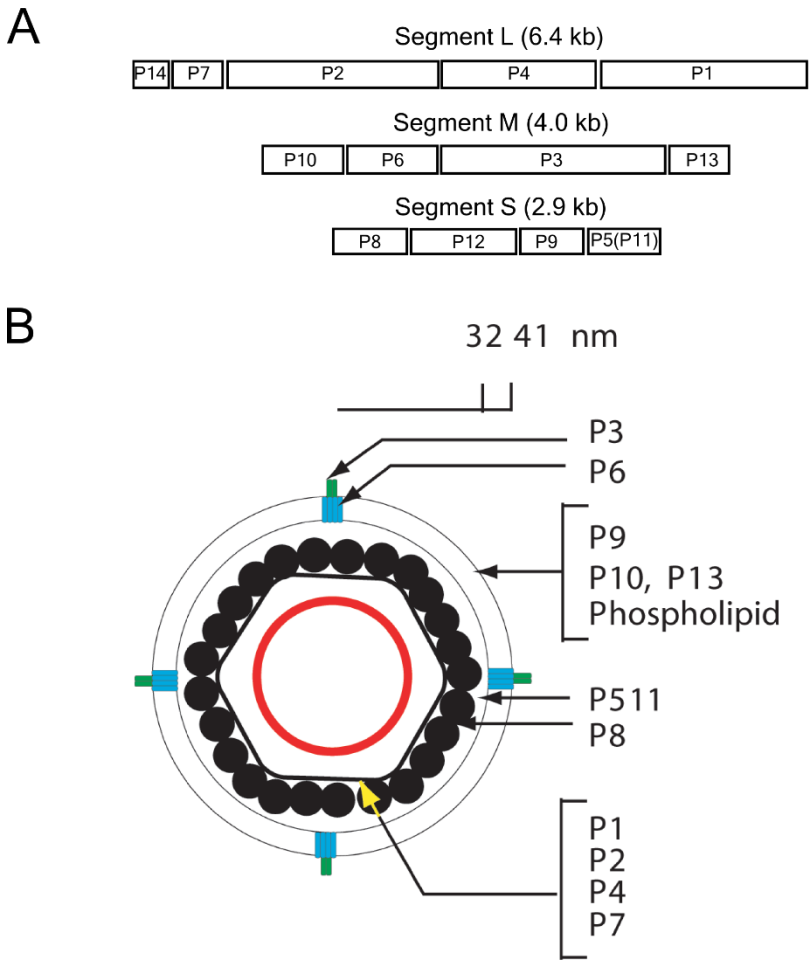

Figure 1.3 Genome organization and structure of cystovirus
(A) Genome organization of segmented dsRNA phage phi6. (B) A cartoon rendition of cystovirus structure. dsRNA (red circle) is enclosed in a core of P1, P2, P4 and P7. The core structure is encapsulated within a shell of P8, which is in turn covered by the lipid-containing membrane. Attachment protein P6 and P3 are in the membrane along with P9 and P10. This figure was adapted from (46)

The (+) strands of S, M, and L segments are packaged into the procapsid sequentially in 5' to 3' direction (49). It has been proposed that the surface of empty procapsid specifically binds to (+) strand of the S segment and initiates packaging through the hexameric NTPase, P4, present at twelve 5-fold vertices (50). Hydrolysis of NTPs by P4 drives the packaging of RNA into the capsids. Packaging of (+) strand of the S segment induces conformational change that exposes binding site for M segment. After M segment is packaged another conformational change facilitates the packaging of L segment. Packaging of L segment activates RNA-dependent RNA polymerase P2 to synthesize (-) strands to double the RNA content within the capsid. At this stage, the capsid becomes inflated and the hexameric P4 protrudes outside as knobs at every vertex, this structure constitutes the viral core. The negative strand of dsRNA serves as a template to transcribe (+) sense RNA that is extruded out of the capsid to be translated in the host cytoplasm. The same (+) sense RNA molecules are packaged into progeny virions (46). The viral core then serves as a template for the assembly of nucleocapsid protein P8. The nucleocapsid encapsulated virions can interact with host cytoplasmic membrane and cause infection in spheroplasts (51).

The assembly of envelope around the nucleocapsid core is the final step in the virion maturation pathway. Unlike other enveloped viruses, envelopes in *Cystoviridae* are not picked up during budding but the components are assembled in the host cytoplasm. Membrane proteins P9 and P12 are the minimum components required for envelopment. The host attachment protein P3 is anchored to the envelope through protein P6 and is the last component assembled on to the virion (46).

The lysis of the host brings about the end of the infection cycle. The exact mechanism of host lysis is not known. It has been suggested that the lytic enzyme P5 in the inner core that facilitates PG degradation during infection also plays a role in lysis. In addition to P5, a small membrane associated protein P10 was also shown to be required for host lysis. It is thought that P10 acts as a holin to release excess intracellular P5 into the periplasm, where it degrades PG and causes host lysis (52, 53). The mechanism by which P10 facilitates host lysis could represent a new mode of lysis.

dsRNA phages other than phi6 have been isolated mostly from the leaves of plants (54). The phage phi 7, phi9, and phi10 share high degree of sequence identity with phi6 but phi8, phi12, and phi13 are quite different from phi6 both in nucleotide and amino acid sequence. Structurally these viruses are very similar in appearance but at the molecular level there are two major differences between phi6-like and other phages. First, phi6 has P6 and P3 to attach to type IV pili while Phi 8 has multiple P3 proteins and it attaches to the host LPS. All phages except phi8 have a P8 shell that encapsulates the inner core (46).

Recently Krishnamurthy et.al., have used bioinformatics to identify, among others, a novel dsRNA phage from whole cell transcriptome data of a *Streptomyces* sp. This was reported as the first ever RNA phage found for Gram-positive bacteria (55). This finding raises the possibility of existence of other RNA phages (both ds and ssRNA phages) against other Gram-positive species.

ssRNA phages

The ssRNA phages, the simplest viruses ever discovered, were first isolated as male-specific phages of *E. coli* in 1961 by Loeb and Zinder (56). These are icosahedral viruses of 23-28 nm in size that share remarkably similar properties among all the known members (57). The capsids have T=3 symmetry with approximately 180 copies of coat protein arranged in 20 hexagonal capsomers forming the faces and 12 pentameric capsomers occupying the vertices of the capsid. The capsid encloses approximately 4000 bases of positive-strand ssRNA attached to a single copy of the maturation protein (Mat) (58). The ssRNA constitutes about 32% of the total weight of the virus particle. The molecular weights of these viruses range from 3.6-3.8 million with sedimentation constants ranging from 77-81S (57). Their relative abundance and readily available viral polycistronic ssRNA made them an ideal source for messenger RNA to study protein translation. This has been one of the most significant contributions of ssRNA phages to the field of molecular biology (59).

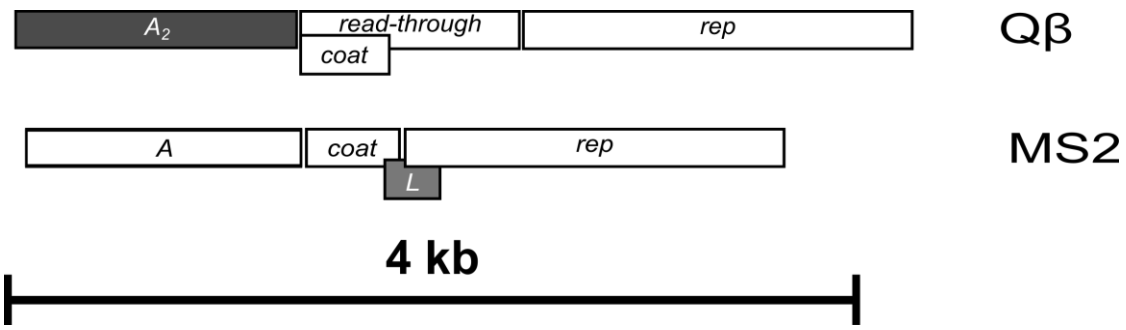

Figure 1.4 Genome organization of prototypical allolevivirus (Qβ) and levivirus (MS2).
The typical genome organization of the two canonically defined groups of *Leviviridae, the* Allolevivirus Qβ and the Levivirus MS2. Both Qβ and MS2 share the same basic core three-gene structure of *mat, coat*, and *rep*. In Qβ the maturation gene (A_2) is also the lysis gene. The product of *coat* gene in Qβ is extended at a low frequency as a read-through product, *A1*. In contrast, the lysis gene in MS2 has evolved as a separate gene, *L*. The lysis genes are shaded grey.

Alloleviviridae and Leviviridae

The ssRNA phages of the family Leviviridae were grouped into four different serological groups viz., Groups I, and II (f2, MS2, R17, fr) and groups III, and IV (Qβ) (60, 61). Based on genome organization, these four groups were later classified in the sixth report of the International Committee on Taxonomy of Viruses into *Leviviridae, Alloleviviridae*, and *Unclassified Leviviridae* (Figure 1.4) (62).

Infection, replication, morphogenesis, and release of progeny

The ssRNA phage infection process begins with the attachment of the virion to the sides of retractable pili produced by bacteria (63). Although ssRNA phages have been found against every retractable pili-like appendage, only coliphages that infect through F pili will be discussed below.

Brinton, Gemski and Carnahan (64) first demonstrated that the presence of F pili is necessary for sexual conjugation in various Gram-negative bacteria. The genes for the F pilus are carried on the conjugative F plasmid and the cells that carry the F plasmid are called F^+ (male) cells. The F plasmid, which is ~ 100 kb and carries genes for the DNA transfer apparatus, is maintained either as a single-copy plasmid (F+ strains) or as an integrated element in the host chromosome (Hfr strains). The F pili are flexible, dynamic, filamentous structures that protrude from the surface of male cell, contact the F^- cell, and allows for transfer of DNA to the female cell (65, 66). The F pilus constitutes on average 10,000 monomers of pilin, a processed product of the *traA* gene. The initial product of *traA* gene is a 121 aa propilin, which gets proteolytically processed after the first 51 residues; N-proximal residue of the C-terminal 70 amino acids is acetylated to form mature pilin (67). Mediated by the *tra* complex, the pilin monomers oligomerize to form 1-20 μm helical structures with an outer diameter of 8 nm and a 2.5 nm hollow core visible in negatively stained electron micrographs (68, 69). Upon contact with a F^- cell, the pilus retracts to bring both donor and recipient cell in apposition to facilitate the transfer of DNA (67). However, cell-to-cell contact is not required for DNA

translocation (70). The physical mechanism of DNA transfer and the role of pilus as a facilitator or as a DNA conduit are still unknown.

The ssRNA coliphages adsorb to the sides of F pili through an interaction between the attachment protein (A in MS2; A_2 in Qβ) of phage and the pilin (71). This interaction is reversible at 4°C; the phage can also reversibly attach to cell-free preparations of F pili. However, when attached to cell-associated pili at 37°C the process is irreversible (72). An ionic strength of 0.1 M or greater favors maximum phage adsorption and is not dependent on the presence of specific cations (73). Mutational analysis has revealed that residues 17-28 of the mature pilin form the exposed sides of pilus and are involved in phage attachment (74). The F pilus-phage interaction leads to cleavage of the Mat protein into two polypeptides, of mass 15 kDa and 24 kDa (74). Cleavage of A provides the necessary signal from F pilus to initiate the ejection of RNA from the capsid. The mechanism of how a ~1000 nm long RNA genome is unraveled from the capsid is unknown. RNA ejection into the cytosol, accompanied by the Mat protein, is associated with a rapid decrease in cellular NTP levels (75). It is estimated that about 10 phage genomes/cell can enter in the first 5 min of infection (76). The RNA entry is dependent of the presence of either Mg^{+2}, Ca^{+2}, Ba^{+2}, or Sr^{+2} in the medium (77). However, it is not known what stage of RNA penetration constitutes the requirement for divalent cations.

The role of the F pilus in the transport of the gRNA has been the subject of several models, none of which supported by molecular evidence. Brinton (1965) first suggested that RNA was transported through the hollow tube of the F pilus. In support of

this model, the hollow core of the pilus is lined by basic residues, which might interact with the negative charge of phosphate backbone of the viral nucleic acid (65). Another model is the retraction of pili upon attachment of RNA phage (Figure 1.5). The pilus retraction proceeds through sequential depolymerization of the pilin subunits at the cell cytoplasmic membrane (78). This retraction may bring the phage RNA-Mat complex to the base of the pilus to be internalized. Support for this model comes from the observation that upon infection with RNA phage PP7, the polar pili of *Pseudomonas aeruginosa* contract by about 50% in length (79).

Upon entering the cell, the phage RNA functions directly as a mRNA, resulting in the synthesis of the viral proteins by the host machinery. Complex regulatory elements constituted in the secondary and tertiary structure of the RNA control replication and translation of the same RNA molecule.

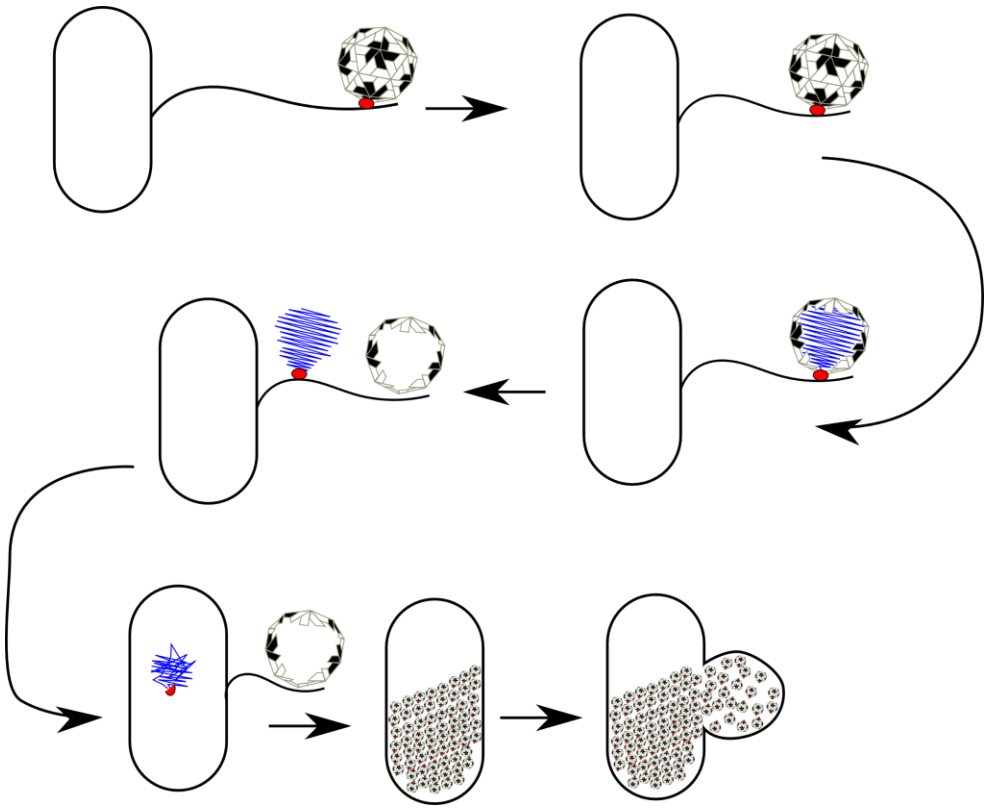

Figure 1.5 ssRNA phage infection cycle
A generalized schematic of ssRNA phage infection. Host recognition begins with the attachment of the Mat protein on the virion to a retractable pilus. The retraction of the pilus stimulates the release of the ssRNA genome bound to Mat. The Mat protein is cleaved into two fragments and the ssRNA genome bound to two half fragments of Mat enter the host through an unknown mechanism. The (strand is synthesized from the (+) strand and serves as a template for amplification of more (+) sense strands, which leads to increased viral protein synthesis and concurrent assembly of phage particles. Lysis of the host terminates the infection cycle.

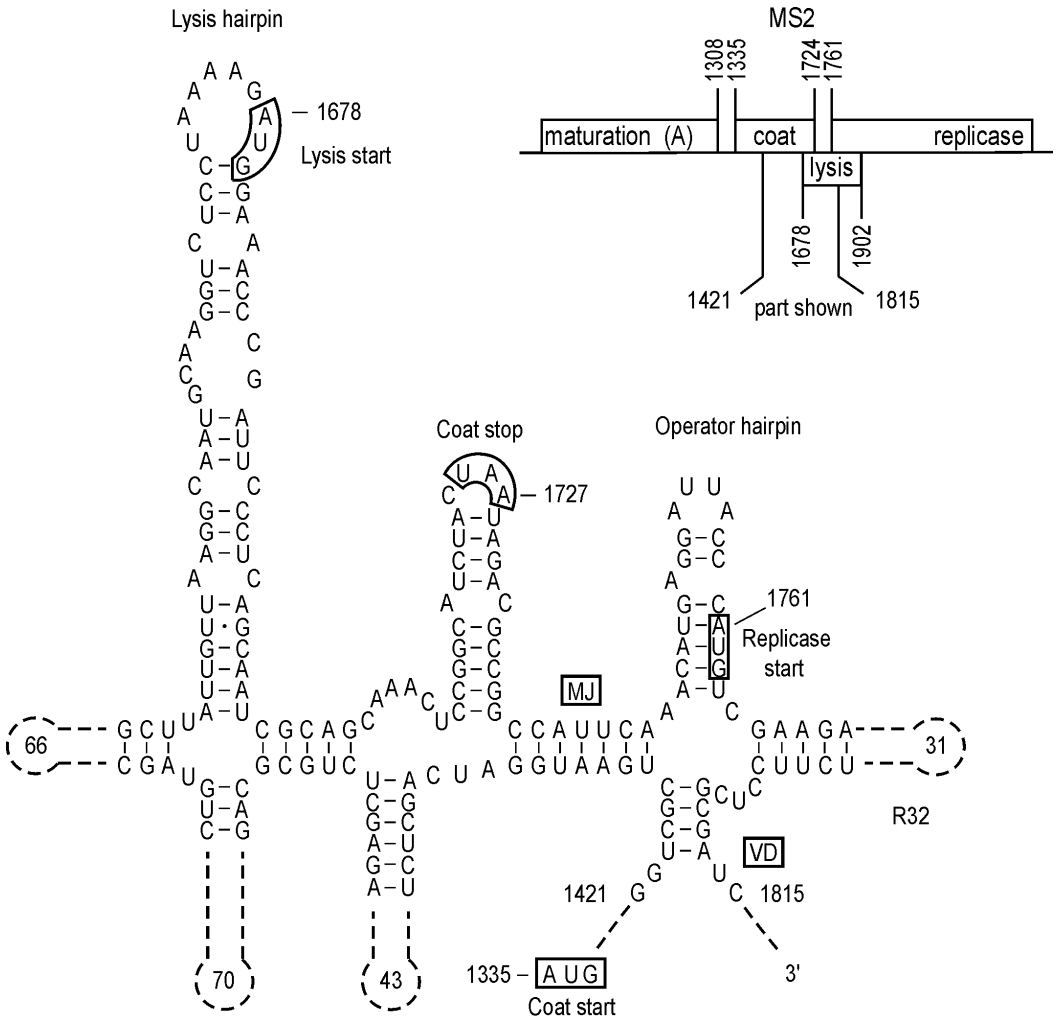

Figure 1.6 Secondary structure of MS2 RNA encompassing 3' region of *coat* and 5' regions of *L* and *rep*

Genome map of MS2 is shown in the top right corner with various nucleotide positions in the genome are indicated by numbers. The secondary structure of RNA sequence between nt 1421 and 1815 is shown with lysis, coat, and operator hair-pins indicated. The start and stop codons of three different gene are boxed on the sequence. This figure was reprinted from (16).

The *coat* gene is the first of the four phage proteins to be translated as the RNA secondary structure occluding the 5' untranslated region, the initiator "hair-pin", is relatively weak compared to the other genes (Figure 1.6). Translation of *coat* disrupts the base-pairing in an element, the "MJ stem", upstream of the *rep* gene, allowing ribosomes to initiate translation of the *rep* cistron. Once Rep is synthesized, the host protein ribosomal protein S1, EF-Tu, and EF-Ts form a complex with Rep to form the RNA-dependent RNA polymerase holoenzyme (RdR). This enzyme complex synthesizes the (–) strand by using (+) strand as a template, after which the (–) strand is in turn used as a template to synthesize additional (+) strand gRNAs. The RdR initiates (–) strand synthesis at the penultimate nucleotide at the 3' end of (+) strand and ends with copying of three G residues at the 5' end of (+) strand. The (–) strand synthesis is completed by the addition of an untemplated A residue at the 3' end of the (–) strand. This process essentially creates the same CCCA 3' end in the (–) strand as the in the (+) strand. For RNA replication, EF-Tu first binds to the 3' end of (–) strand and recruits the Rep core enzyme subunit. Deletional analysis has shown that C-terminal residues of the core enzyme also play a role in recognizing the template. The long-distance RNA interactions near the 3' end aid in capture of the 3' $CCCA_{OH}$ end by the core enzyme to begin RNA synthesis. The levels of coat protein control the intracellular levels of Rep. At sufficiently high concentration of coat protein, dimers of coat bind to the operator hairpin (Figure 1.6) and prevent initiation of new rounds of rep translation. Maturation protein is present in single copy per virion so the levels of Mat protein also tightly regulated. The SD sequence of *mat* is normally sequestered in one of the stems of a 5'

clover-leaf-like structure. However, in a nascent (+) strand, because of an intermediary "kinetic trap" state, the 5' structure is slow to reach the equilibrium, thus allowing ribosomes to access the SD site and initiate production of Mat. Thus, of the three core genes only *mat*, and *rep* are under the control RNA regulatory element, whereas *coat* lacks any known regulatory elements. The competition for the S binding site upstream of *coat* gene between ribosomes and replicase (Figure 1.6) acts as the only negative regulation for coat expression (16).

Assembly of the virions begins with the binding of the Mat protein to the (+) strand RNA genome (gRNA). This binding primarily happens at the 3' end (V2 stem loop in MS2) (Figure 1.6) of the genome and a secondary site located with the *mat* gene at position 400. Binding of the 3' end by Mat prevents (–) strand synthesis and initiates genome packaging. The second step in assembly is the construction of the capsid around the RNA. The capsid nucleation event was thought to be mediated by the operator hair-pin at the start site of the *rep* gene (Figure 1.6) (16). Recently Dai et.al., using high-resolution cryo-EM studies on MS2, have shown that three consecutive stem-loops bind to three neighboring coat protein dimers, with the replicase operator being in the middle. Additionally, they found that 13 other stem-loops showed a relatively strong interaction with the coat dimers and are clustered around the operator hair-pin (80). Although, coat protein will form icosahedral capsids in the absence of RNA, the presence of the nucleation stem loops significantly enhances the formation of capsids even at low concentrations (81).

The function of the lysis gene terminates the infection cycle. In MS2, the lysis gene, *L*, is overlaps both *coat* and *rep* genes (82). To control the timing of the lysis event, translation of L is restricted mainly by the "lysis hair-pin" (Figure 1.6). Unlike the translation of *rep*, the expression of *L* is not coupled to the translation of the *coat* gene and subsequent unwinding of the hair-pin. Instead, translation of *L* happens only if the *coat* gene terminates near the *L* start site. It has been suggested that ribosomes that terminate at the *coat* stop codon drift on the RNA and initiate new round of translation at near-by start sites. For MS2, it is estimated that the probability of successful reinitiation at the lysis start site is only 5% (16).

ssRNA phages that infect bacteria other than *E. coli*

Given the simplicity of ssRNA phages, it is likely that these viruses initially evolved to infect ancestral forms of modern eubacteria. Thus, ssRNA phages could have possibly co-evolved to infect every genus of eubacteria. Most the ssRNA phages isolated thus far have been those that infect the genus *Escherichia*. However, there are a few that have been isolated against the genera of *Acinetobacter*, *Pseudomonas*, and *Caulobacter* (83-85) (Figure 1.7). Among these phages, the ones that infect *Caulobacter* have been classified into three different groups based on serological and species specificities, with each group infecting only one of the following three Caulobacter species: *bacteroides*, *fuseformis*, and *crescentus*. Irrespective of the type of pili used and the host species that ssRNA phages infect, their physical properties are remarkably similar. The virions are almost indistinguishable in negatively stained electron micrographs, the diameter of the particles ranging from 21-26 nm, and the sedimentation coefficient of virions ranging

from 70-88S. By extension, the genomic RNAs from different ssRNA phages also have similar sedimentation coefficients of ~ 30S (86). However, subtle morphological differences between couple of ssRNA coliphages have been recently elucidated through high-resolution cryo-EM asymmetric reconstructions (87) (80). One of the interesting features of coat proteins of ssRNA coliphages is the absence of histidine residues (88). However, in the case of Caulobacter phage phiCb5, the coat protein lacks methionine instead of histidine. An analysis of available phage genomes has revealed that phages that infect exclusively *Pseudomonas* or *Acinetobacter* have all 20 amino acids. The rationale of this unusual phenomenon in the RNA phage infection cycle is not known.

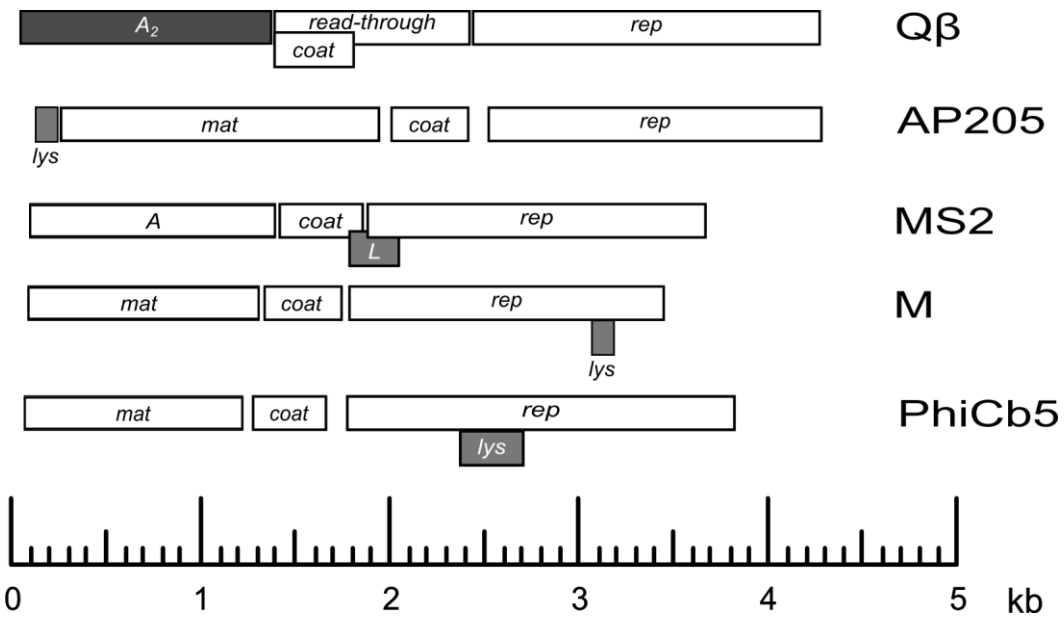

Figure 1.7 Evolution of lysis genes at different genomic locations.
Genome maps of five different phage-types drawn to scale. Lysis genes are highlighted in grey.

Cell wall structure and biosynthesis

A bacterial cell would exist just as a fragile membranous structure of lipids, proteins, and nucleic acids if it were not for the high tensile-strength peptidoglycan (PG) layer, or the sacculus. This layer allows the cell to tolerate high internal osmotic pressures (3-30 atm) and still maintain shape (89). The PG layer is composed of glycan strands made up of repeating disaccharide units of Mur*N*Ac-pentapeptide and Gluc*N*Ac (Figure 1.8). In Gram-negative bacteria, this layer is composed of 2-3 glycan strands, whereas in Gram-positive bacteria the PG layer has20-80 glycan strands (90-92). Studies on the *E. coli* sacculi show that, on average, a single layer of glycan chains run almost perpendicular to the long axis of the cell (93). To provide additional strength, the glycan strands are cross-linked through the formation of a peptide bridge between pentapeptide side chains of Mur*N*ACs of adjacent strands (Figure 1.8). These cross-links make the PG layer a very large complex polymer that can be isolated and visualized under an optical microscope (92). In Gram-negative bacteria, the PG layer is covalently linked through peptide linkages to the Braun's lipoprotein (Lpp), which is anchored in the inner leaflet of OM (91).

Biosynthesis of PG sacculus can be divided into three different stages: the cytoplasmic stage, the membrane stage, and the periplasmic stage.

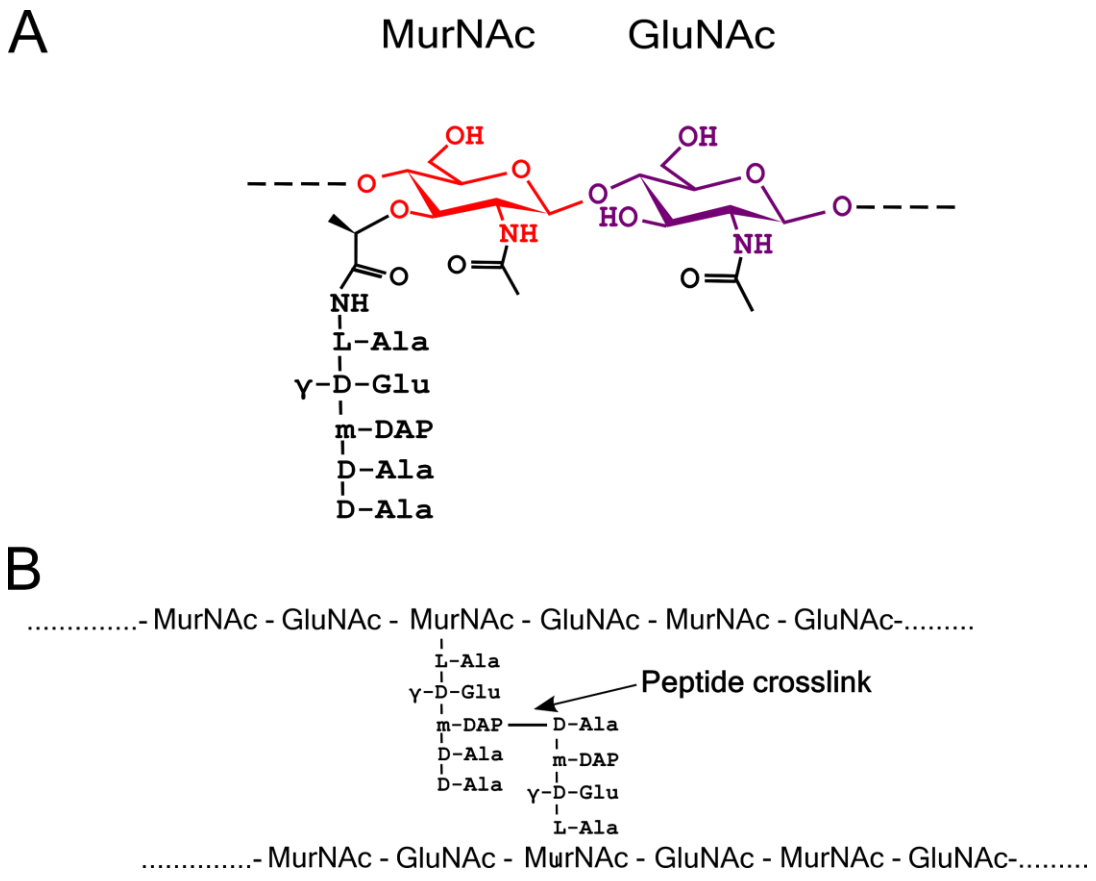

Figure 1.8 Structure of peptidoglycan
(A) The structure of the disaccharide pentapeptide building block of *E. coli* glycan strands is shown. The dashes on either end represent the repeating units of building blocks in a glycan strand. A three-letter code is used for amino acids in the peptide side chain. (B) The major peptide crosslink (3-4 linkage) found between glycan strands in *E. coli* is shown. The crosslink occurs between the free amino group of m-DAP of the glycan strand (top strand) and the carboxylic acid group of the penultimate D-Ala of another glycan strand (bottom strand). The terminal D-Ala of the second strand (bottom strand) is lost in the crosslinking reaction.

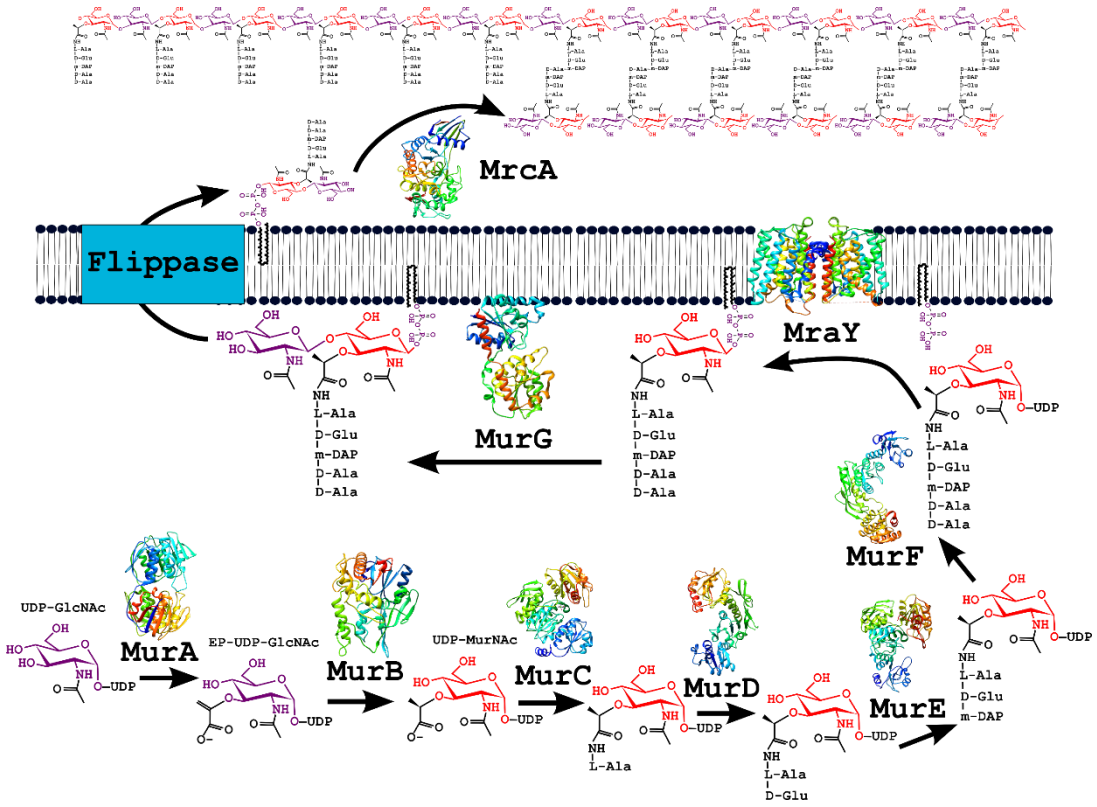

Figure 1.9 The peptidoglycan biosynthesis pathway
The conserved PG biosynthesis pathway from the modification of UDP-Glc*N*Ac to incorporation in periplasmic PG is shown. The enzymes that catalyze each step of the reaction are shown next to the arrow. The Lipid II flippase is indicated by a generic blue box in the membrane representing both the flippase candidates FtsW and MurJ. The undecaprenol lipid carrier is shown as a hair-pin like structure in the bilayer. The PDB numbers of the enzyme structures shown in the figure are as follows: MurA (PDB: 3KR6); MurB (PDB:1MBT); MurC (PDB: 2F00); MurD (PDB: 1E0D); MurE (PDB: 1E8C); MurF (PDB: 1GG4); MurG (PDB: 1NLM); MraY (PDB: 4J72); MrcA (PDB: 2C6W).

The PG biosynthesis begins in the cytoplasm and it is the transfer of enolpyruvyl moiety from phospho-enolpyruvate (PEP) to uridine diphosphate (UDP)-*N*-acetylglucosamine (Glc*N*Ac). This step is catalyzed by the essential and highly conserved enzyme MurA (94) (Figure 1.9). The structures of MurA from several different species has been solved in both free and ligand-bound states (95-97). The structures revealed two globular domains that form a clam shell-like structure in the apoenzyme state with the active site located in the cleft (open state). Binding of the both substrates PEP and UDP-Glc*N*Ac in the cleft brings the two globular domains closer and encompasses the substrates to form the catalytically active closed state. The next step in PG biosynthesis is the NADPH-dependent reduction of enolpyruvyl UDP-Glc*N*Ac to UDP-*N*-acetylmuramic acid (Mur*N*Ac) by the enzyme MurB (98) (Figure 1.9). The next series of steps involve sequential ligation of the pentapeptide amino acids to the lactyl moiety of UDP-Mur*N*Ac. The four Mur ligases (MurC-MurF) catalyze these reactions in an ATP-dependent manner. First, MurC catalyzes the addition of L-Ala to UDP-Mur*N*Ac, then MurD adds D-Glu to the product of MurC reaction (UDP-Mur*N*Ac-L-Ala). The third residue in the pentapeptide chain is the unusual amino acid mesodiaminopimelic acid (meso-A_2pm or mDAP), which is added to the product of MurD (UDP-Mur*N*Ac-L-Ala-D-Glu) by the enzyme MurE. The MurC-E ligases are structurally very similar and have an N-terminal nucleotide-activated substrate binding Rossmann-type α/β fold, a central ATP binding domain, and a C-terminal amino acid binding domain, which provides specificity. The binding all three substrates (ATP, UDP-Mur*N*Ac, and L-Ala/D-Glu/mDAP) occurs in a sequential manner. This sequential

binding of the substrates in Mur ligases is essential for forming the high-energy acyl phosphate intermediate in the ligation reaction. MurF catalyzes the last cytoplasmic step in PG biosynthesis, which is the ligation of D-Ala-D-Ala dipeptide to UDP-Mur*N*Ac-L-Ala-D-Glu-mDAP). Unlike MurC-MurE ligases, MurF lacks the N-terminal Rossmann-type α/β domain that binds to nucleotide-activated substrate, instead the nucleotide binding function is carried out by the central ATP-binding domain that extends outward. The D-Ala-D-Ala substrate synthesis is carried out separately in the cell by an ATP-dependent D-Ala-D-Ala ligase. The D-Ala is in turn made by racemization of L-Ala by an alanine racemase. The final cytoplasmic product of PG biosynthesis is the UDP-Mur*N*Ac-L-Ala-D-Glu-mDAP-D-Ala-D-Ala (UDP-Mur*N*Ac-pep5).

The first membrane-linked step in PG synthesis begins with the transfer of UDP-Mur*N*Ac-pep5 to the lipid carrier undecaprenyl phosphate (C55-P or UndP) (Figure 1.9). This reaction is catalyzed on the cytoplasmic face of the IM by the integral membrane protein MraY to generate a product commonly known as Lipid I. MraY has 10 transmembrane domains, five relatively long cytoplasmic loops, and six periplasmic loops (99). However, the crystal structure of MraY from *Aquifex aeolicus* revealed that two of the six periplasmic loops are structured, comprising a β-hairpin and an alpha helix (100). The catalytic active site, Mg^{+2}, and UDP-Mur*N*Ac-pep5 binding sites are in a cleft formed by TMDs 5, 3,4,8 and 9b which has access to the cytoplasm. An inverted U-shaped groove near TM9b is proposed to the binding site for the Und-P lipid carrier. Single-step and two-step mechanisms have been proposed for the catalytic activity MraY, but these mechanisms are yet to be confirmed (99). The next lipid-linked step in

the pathway is the transfer of Glc*N*Ac from UDP-Glc*N*Ac to the hydroxyl group at position C4 of the membrane linked Lipid I to yield Lipid II (Figure 1.9). MurG, a soluble protein that functions at the membrane surface, catalyzes this step and follows an ordered Bi-Bi mechanism where the enzyme first binds UDP-Glc*N*Ac, then Lipid I is recruited to the enzyme to complete the transfer of the sugar. MurG consists of two domains linked by a hinge region; the N-terminal domain contains a membrane-associating hydrophobic region and the Lipid I binding region, while the C-terminal domain binds UDP-Glc*N*Ac (98, 99). The Lipid II is flipped across the cytoplasmic membrane and the identity of the enzyme(s) that carry out this function has been a contentious issue and is covered under the sub-heading "The flippase wars".

The extracellular steps of PG biosynthesis utilize the high energy stored in the phosphodiester-muramic acid bond of the flipped Lipid II and in the D-Ala-D-Ala peptide bond to drive the polymerization and cross-linking reactions respectively. These steps are carried out by mono- or bi- functional penicillin binding proteins (PBPs). Based on the structure and function of their N-terminal domains, PBPs are divided into Class A (bifunctional) and Class B (monofunctional) enzymes. The Class A enzymes have a N-terminal proximal glycosyl-transferase (GT) domain and a C-terminal transpeptidase (TP) domain. The Class B enzymes have only C-terminal TP domain and the N-terminal domain has non-penicillin binding functions. The GT domain catalyze the initiation and elongation of Lipid II polymerization. During initiation step, two Lipid II monomers are linked b-1,4 linkage between NAM-NAG to form Lipid IV and subsequent addition of new Lipid II molecules with the loss of undecaprenyl-

pyrophosphate elongates the polymer (98). Recently, RodA a member of the SEDS superfamily was shown to catalyze GT reactions (101, 102). Although, the extent to which RodA drives the GT reactions is not known. The TP domain catalyzes transpeptidation to cross-link peptide side chains of adjacent glycan strands. Additionally, TP domains catalyze carboxypeptidation reaction to cleave the terminal D-Ala from the pep5 side chain.

The "flippase wars"

After Lipid II is formed in the inner leaflet of IM, it must be flipped to the outside of the cell (the periplasm in Gram-negative bacteria) to be incorporated into the PG (Figure 1.9). It is estimated that there are only a few hundred Lipid II molecules in the cell at any given moment, yet thousands of flipping events per min are needed to build the cell wall. In addition, the number of C55-P lipid carriers are also limited (~7000 molecules/cell) and are shared by several different pathways (103-107). These observations led to a suggestion that Lipid II requires a transporter that rapidly recycles the lipid carrier. The identity of this transporter has been sought for decades.

For several years, FtsW was as the only serious Lipid II flippase contender, partly because it fit the genetic criteria viz.: an essential integral membrane protein, associated with divisome, interacting with class A (bifunctional) and class B (transpeptidase only) penicillin-binding proteins (PBPs). Moreover, the gene, *ftsW* was localized in a division and cell wall biogenesis gene cluster. Also, the lack of good biochemical assays for measuring Lipid II flipping activity, either *in vivo* or *in vitro*, obscured the confirmatory experiments. The Breukink group developed a FRET-based

assay and demonstrated that FtsW translocated a Lipid II analog across reconstituted liposomes that harbored purified FtsW (108). Based on these experiments it was concluded that FtsW is the flippase for septal PG synthesis and its paralog RodA was the flippase for cell elongation. Using the same FRET-based assay, two charged residues R145 and K153 on TMD4 were shown to be required for translocation of Lipid II (109). A pore-like structure within FtsW was suggested to accommodate Lipid II. Apart from Lipid II analogues, FtsW was also shown to facilitate transport of different membrane phospholipids. However, the charged residues on TMD4 were not required for this activity. Based on the in vitro activity of FtsW, it was proposed that other members of shape, elongation, division, and sporulation (SEDS) superfamily might also function as Lipid II flippases.

In opposition to the FtsW/RodA hypothesis which was based on in vitro biochemical data, Natividad Ruiz used a reductionist bioinformatic approach to identify the essential gene, *murJ* (*mviN*), as the only Lipid II flippase in *E. coli* (110). Simultaneously, an essential gene screen also identified *mviN*, another name for *murJ*, as a protein involved in murein synthesis (111). MurJ belongs to a well-characterized family of transporters, the multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. Depletion of MurJ either by a shifting a thermosensitive mutant to the restrictive temperature or by turning off an inducible promoter leads to accumulation of Lipid II intermediates, nucleotide precursors, loss of cell shape and ultimately lysis. MurJ has 14 TMDs with N- and C-termini in the cytoplasm and a central solvent-exposed cavity which is lined by the transmembrane domains (TMD) TMD1, TMD2,

TMD7, and TMD8 (112). Charged residues R18, R24, R52, and R270, which lie within a solvent-exposed cavity, were found to be essential for MurJ function (113). These charged residues are hypothesized to be involved in interaction with Lipid II and possibly to facilitate energy coupling during the flipping activity. However, the notion that MurJ was the long-sought Lipid II flippase lacked biochemical evidence, until 2014, when Sham et al., developed an *in vivo* flippase assay (114). This assay was based on the ability of colicin M (ColM) to specifically cleave periplasmic Lipid II and thereby generating a soluble ColM product (Figure 1.10) (115). In *E. coli*, upon treatment with ColM, the periplasmic Lipid II pools are converted into the soluble ColM product, proceeding until all the cellular Lipid II molecules are depleted. In the central experiment, when MurJ was chemically inactivated using thiol-modification, Lipid II and other nucleotide precursors accumulated inside the cell with a concurrent decrease in the ColM product. When this assay was repeated with an *E. coli rodA* strain depleted of FtsW, the ColM product, and thus Lipid II translocation, was unperturbed. This lead the authors to conclude that FtsW is not the flippase and, by extension, the SEDS superfamily members are not the Lipid II flippases.

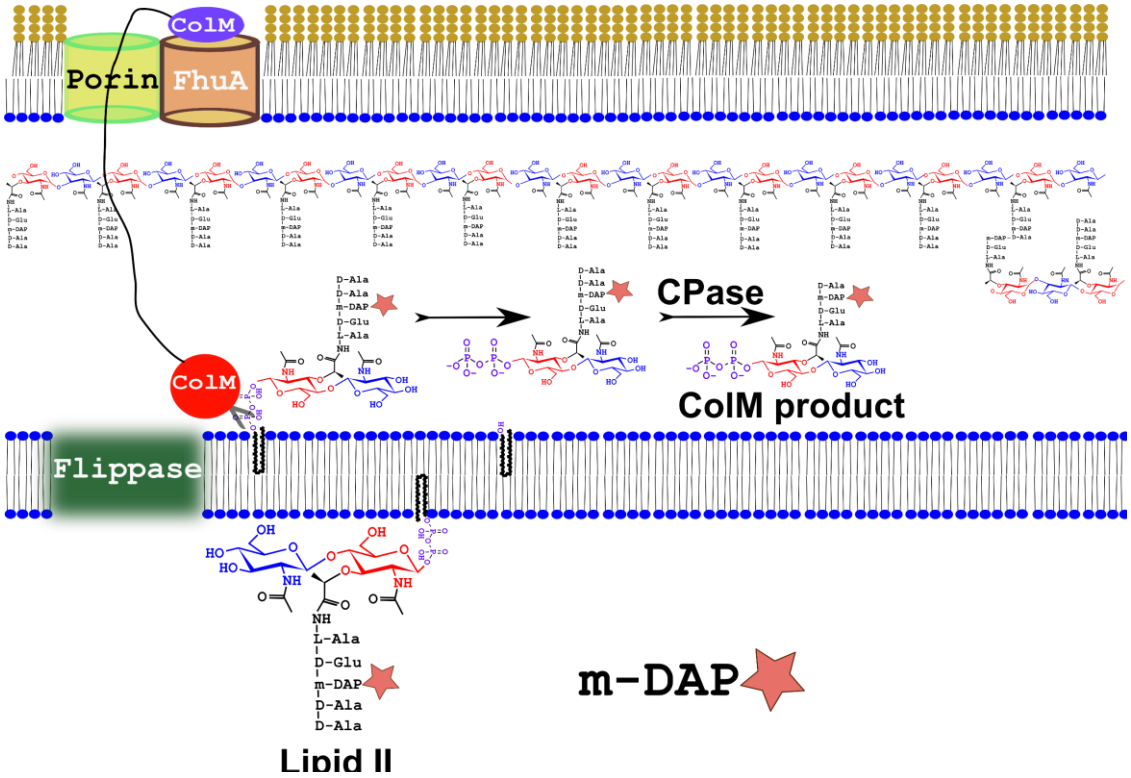

Figure 1.10 ColM assay for Lipid II flippase activity
Diagram showing the cleavage of periplasm exposed Lipid II by ColM. The C-terminal domain of ColM binds to FhuA and the N-terminal domain gains access to the periplasm, presumably through an OM porin. In the periplasm, the N-terminal domain of ColM cleaves periplasmic Lipid II into undecaprenol (Und-OH) and soluble pyrophosphate-Mur*N*Ac-Pep5-Glc*N*Ac, which is processed by periplasmic carboxypeptidases to yield pyrophosphate-Mur*N*Ac-Pep4-Glc*N*Ac (ColM product). Radiolabeled [^{3}H]-mDAP is indicated by a star.

MurJ is highly conserved among the Gram-negative bacteria but not in Gram-positive bacteria. In Firmicutes such as *Bacillus*, the closet functional homolog of MurJ is YtgP. In *B. subtilis*, YtgP and other members of the MOP family of proteins (SpoVB, YkvU, and YabM) are not essential under laboratory conditions (116). Sporulation defects associated with SpoVB null mutants are complemented by *E. coli murJ*; conversely, both *spoVB* and *ytgP* can complement a *murJ* deletion or depletion (117). However, deletion of all 10 MOP members in *B. subtilis* had no significant growth or morphological defects, suggesting the existence of an alternate flippase. Subsequently, *ydaH* (renamed as "*amj*", for alternate to murJ) was discovered in a synthetic-lethal screen with *ytgP* (118). Accordingly, the ColM assay was used to show that Amj translocated Lipid II; moreover, *amj* was shown to complement a *murJ* deletion in *E. coli*. Unlike MurJ, Amj is predicted have only 6 TMDs and is hypothesized to dimerize to form a central cavity to accommodate Lipid II during its translocation.

The identity of the Lipid II flippase and the mechanism of Lipid II translocation remains as the last unresolved mystery in the core PG biosynthesis pathway.

Lysis strategies of small lytic phages

Small lytic phages are known to cause lysis of Gram-negative bacterial host without comprising a muralytic activity. The lysis proteins of these phages were thus termed as "amurins" as they subvert PG without actively degrading the PG meshwork. The three canonical amurins and the current knowledge concerning these amurins are reviewed below.

The E lysis protein of ϕX174

The first single-gene lysis system to be studied genetically was *E* of ϕX174, A genetic approach was employed to show that the host target of E is MraY (119). This genetic data contradicted the previously proposed "membrane tunnel" model by Lubitz and coworkers (120) and established a much simpler mechanism of action for E. As noted above, MraY is a universally conserved integral membrane protein that catalyzes the synthesis of the first lipid linked intermediate (Lipid I) in the pathway for PG precursor biosynthesis (104). MraY has 10 transmembrane domains (TMDs) with N-out/C-out topology (100, 121). E functions as a non-competitive inhibitor of MraY (122), thereby inhibiting PG biosynthesis and causing a septal collapse in cells that attempt to divide (Figure 1.11). Several mutants of MraY have been isolated with most of them being localized to only two (TMD 5 and TMD 9) of the ten TM helices (119, 122). The first 35 aa of E comprise a hydrophobic domain that is essential for function, whereas the last 56 aa are not conserved, highly basic and can be replaced by LacZ or other protein domains (123). In addition, the minimal structure of E that can cause lysis was determined to be the first 29 residues and within this a minimal 18 residue region was found to be critical for function (124).

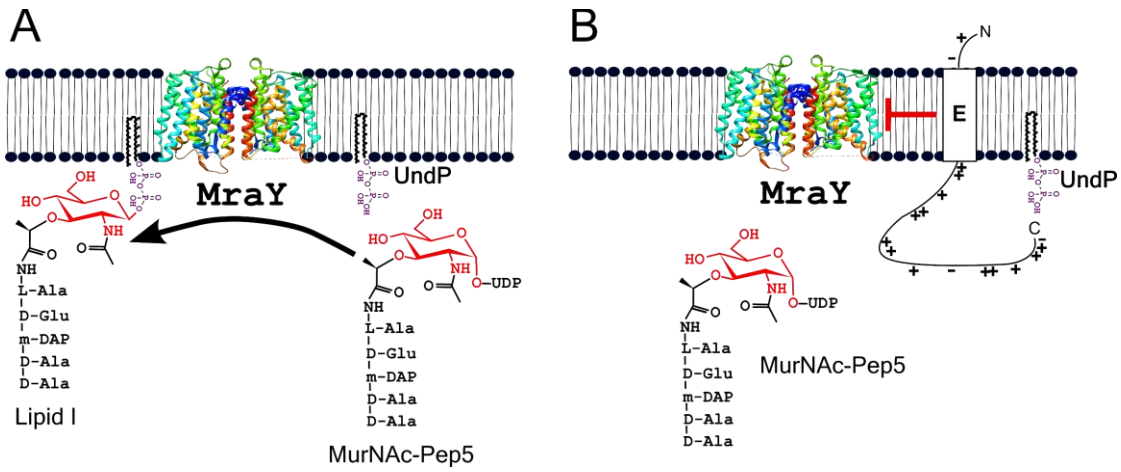

Figure 1.11 E of ϕX174 inhibits the first lipid-linked step in PG biosynthesis
(A) MraY transfers MurNAc-pentapeptide to the lipid carrier undecaprenol pyro-phosphate (UndP) to form lipid I. Structure of MraY (PDB: 4J72). (B) The lysis protein E inhibits the transfer of MurNAc-pentapeptide to UndP and this leads to an accumulation of cytoplasmic PG intermediates and eventual lysis.

Interestingly, E requires a host protein SlyD, to function normally (125). SlyD is a FKBP-type peptidyl-prolyl cis-trans isomerase that functions as a protein chaperone in *E. coli* (126). In the absence of SlyD, E is highly unstable, presumably because it fails to fold properly (127). Therefore, in the absence of SlyD the host is resistant to E-mediated lysis. Indeed, the first host mutants that were isolated as E-resistant mutants were deletions in *slyD*. *E* mutants that overcame this dependence were isolated by plating ϕX174 on lawns on *slyD* cells and selecting phage mutants that formed plaques. These mutants were called *Epos* for "plates on slyD. Epos mutants R3H and L19F bypassed the requirement of *slyD* due to increased E protein synthesis, although the Epos protein remained unstable (128, 129).

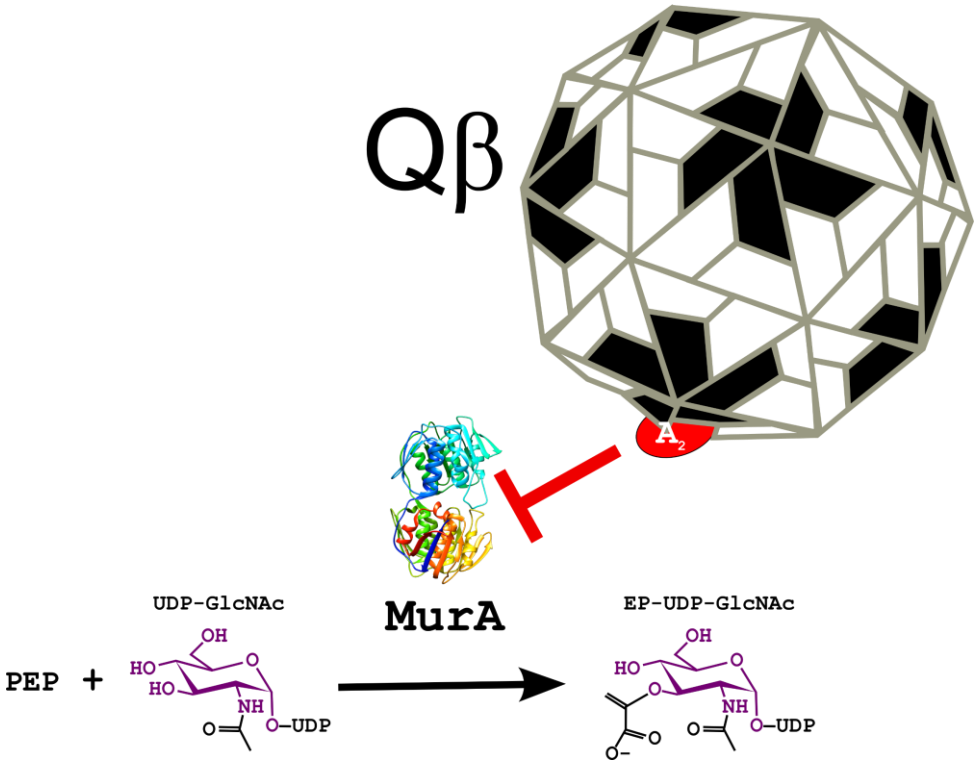

Figure 1.12 A_2 of Qβ inhibits the first committed step in PG biosynthesis
The transfer of enol pyruvate moiety from PEP (phospho enol pyruvate) to UDP-Glc*N*Ac is catalyzed by MurA to form enolpyruvyluridine 5'-diphospho-N-acetylglucosamine (EP-UDP-Glc*N*Ac). The A_2 molecule on a Qβ virion binds to MurA in the closed state, when both PEP and UDP-Glc*N*Ac are bound in the active site.

A_2 of Qβ

A similar genetic approach was used to identify the host target of the lysis protein A_2 of ssRNA phage Qβ (130). To isolate host *rat* mutants (resistant to A-two, A_2 was induced from a multicopy plasmid clone and survivors that failed to lyse were selected. Survivors in the initial selection were subjected to a second round of screening by cross-streaking with Qβ phage, thus isolating *rat* mutants. Mapping and subsequent sequencing of the resistant locus led to identification of a missense mutation L138Q in MurA, the enzyme that catalyzes the committed step in PG biosynthesis (Figure 1.12). Inhibition of MurA precludes the synthesis of PG precursors and incorporation of new subunits into the growing polypeptide chain, leading to a septal collapse (Figure 1.9). Therefore, A_2-mediated inhibition of MurA mimics the lysis phenotype of E. During lysis effected by expression of A_2 or *E*, [^{3}H]-DAP incorporation into the SDS insoluble material ceases prior to the onset of lysis (130, 131). It was shown that Qβ particles inhibited $MurA_{wt}$, but not the *rat* mutant L138Q, in a crude extract; the dissociation constant was estimated at ~10 nM (130). Subsequently, in vitro studies with a soluble MBP-A_2 fusion and MurA suggested that the interaction is direct and dependent on the conformation of MurA, preferentially binding to the closed conformational state of MurA liganded by both substrates (132).

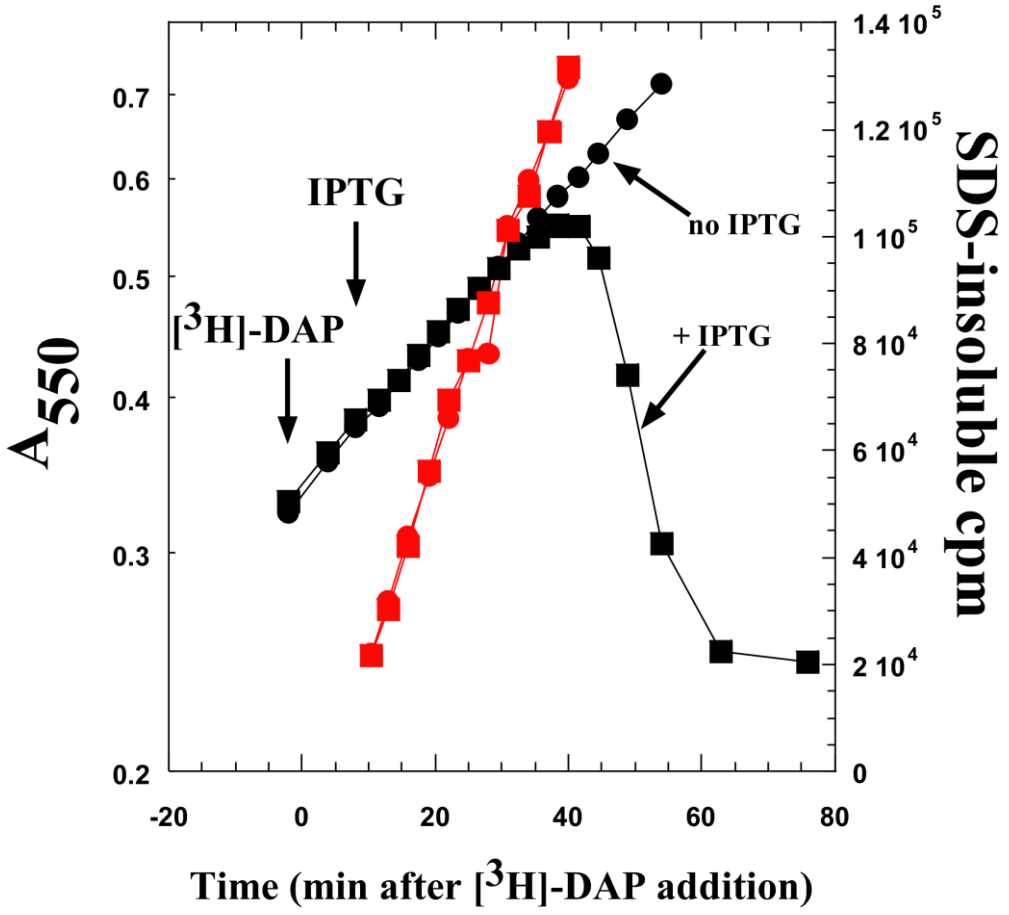

Figure 1.13 L does not affect [^{3}H]-DAP incorporation.
Cultures with *L* expressed from an inducible plasmid (squares) or empty plasmid (circles). Red symbols show the incorporation of labeled [^{3}H]-mDAP into SDS-insoluble material (cpm). Black symbols show growth and lysis monitored by A_{550} measurements. The time points at which [^{3}H]-mDAP and inducer (IPTG) were added are indicated by an arrow. This figure was adapted from (133).

L of MS2

In contrast to E and A_2, lysis protein L from ssRNA phage MS2 does not inhibit [^{3}H]-DAP incorporation into SDS-insoluble material (133) (Figure 1.13). This indicates that L does not target any step in PG precursor pathway or the first step of incorporation of Lipid II into the murein. In addition, the morphology of lysis mediated by L is drastically different. In contrast to E and A_2 (which cause septal collapse indistinguishable from that observed with cell-wall antibiotics like penicillin (14), L-mediated lysis is characterized by blebbing along the lateral cell wall (134). Together these considerations indicate that L represents a new class of amurin, one that does not inhibit PG biosynthesis (Bernhardt 2001, McIntosh 2008). *L* encodes a 75 aa protein that has been reported to be present in the membrane fractions (82). Initial genetic and biochemical characterization of L was attempted in the 1980s by the van Duin and Holtje groups. It was reported that the highly basic N-terminal half of the protein is dispensable for lytic activity but none of the C-terminal truncations were reported to be functional (135). An experiment with a synthetic peptide comprising of C-terminal 25 amino acids of L was reported to dissipate proton motive force (pmf) of *E. coli* inverted membrane vesicles and cause fluorescent dye leakage in reconstituted liposomes (136). Additionally, based on immunoelectron microscopy, L was reported to be enriched in the zones of adhesion between IM and OM, also known as Bayer's patches (137). Biochemical analysis of the murein in cells in which L-mediated lysis had occurred was reported to have a decreased average chain length of glycan strands and altered cross-linking between them (138). Based on the above observations, van Duin's group

proposed that L causes lesions in the IM and thereby activates host autolytic enzymes such as lytic transglycoslyases and D-D endopeptidases to cause rapid lysis (138). Although this remains as an attractive model for L function, the experiments with synthetic peptide lacked any negative controls and was done in an *in vitro* system. Additionally, the reported localization of L in the Bayer's patches in all likelihood is unlikely as these zones of adhesions are now thought of as an artifact of sample preparation for EM imaging (139). Furthermore, L lacks any known signal sequence for periplasmic export, suggesting that it is not capable of sorting to periplasmic structures.

Lysis genes in other phages

There are 34 sequenced ssRNA genomes are available in the NCBI database under the family Leviviridae, which is in turn subdivided into Allolevivirus, Leviviruses, and unclassified Leviviruses. Within these genomes, a large majority of them are highly similar in both genetic architectures and nucleotide sequence. Under Allolevivirus, there is total of 11 phages but due their extremely high sequence identity, five phages (ID2, NL95, SP, TW19, and TW28) are grouped under phage FI and four phages (M11, MX1, ST, and VK) are grouped under Qβ. In effect, there are only two Allolevivirus in the NCBI database. However, based on the genetic architecture and sequence similarity, both could be considered as variants of Qβ. Similarly, under Levivirus there are 20 phages with 9 grouped under BZ13, 7 grouped under MS2, and 4 unclassified leviviruses. In addition to Allolevivirus and Levivirus, there are 4 unclassified phages in Leviviridae that cannot be placed into either of the sub-families (62).

Based on genetic architecture of the lysis gene, all 34 known ssRNA phages can be divided into five phage-types: Qβ, MS2, AP205, and phiCb5. As noted above, in Qβ the Mat gene A_2 at the 5' end of the genome encodes a protein that serves the dual function of pili attachment factor and as the lysis protein. In contrast, the lysis gene *L* in MS2 is embedded in alternate reading frames of both *coat* and *rep* genes. In the cases of phiCb5 and coliphage M, the lysis gene is entirely embedded in the *rep* gene, although in different places (140, 141). The *Acinetobacter baumanii* phage AP205 has its lysis gene separate from the other three genes, located at the5' end of the genome (83). The diverse location of the lysis genes in various RNA phage genomes suggests that lysis genes have evolved more than once and probably as a late addition to the genome after speciation to different hosts (or to different pili). This diversity in genetic architectures of lysis genes also raises the possibility that the underlying mechanism of lysis might also be different. Among the RNA phages only the lysis mechanism of A_2 from Qβ has been solved and it targets PG biosynthesis enzyme MurA. Given the diversity of single gene lysis (SGL) systems and the existence of multiple protein targets in PG biosynthesis and maintenance, it is not difficult to imagine the existence of SGL inhibitors for every step in PG biosynthesis.

A recent survey of publicly available RNA-inclusive metagenomes and RNA virome studies of invertebrate species have led to the identification of ~200 new ssRNA phage genomes (55, 142). Although most of the new genomes are partial, about 80 genomes are either complete or nearly full-length, with all the three core genes present. Some of the new viruses had genomes sizes much larger than all the known *Leviviridae*.

In addition to these core genes, some genomes were found to have additional ORFs of unknown function (55). For example, the genome of AVE000 (4.95 kb) has a 1.2 kb ORF of unknown function upstream of *mat* gene and partially overlapping 259 nucleotides of *mat* gene (Figure 1.14). In AVE001 (5.02 kb), the *mat* gene is unusually large (2.39 kb) when compared to *mat* genes of sequenced leviviruses (range from 1.17-1.60 kb). In a first of its kind discovery, AVE002 was found to have an additional nonoverlapping ORF between *mat* and *rep* genes. One of the genes presumably encodes the coat protein but the other could be a novel lysis gene or a gene encoding a capsid component to enlarge the size of the virion. Unsurprisingly, only one genome (AVE017) of ~200 genomes had an annotated lysis gene. The lysis protein of AVE017 has 38% sequence identity and 43% similarity with the L lysis protein of coliphage MS2.

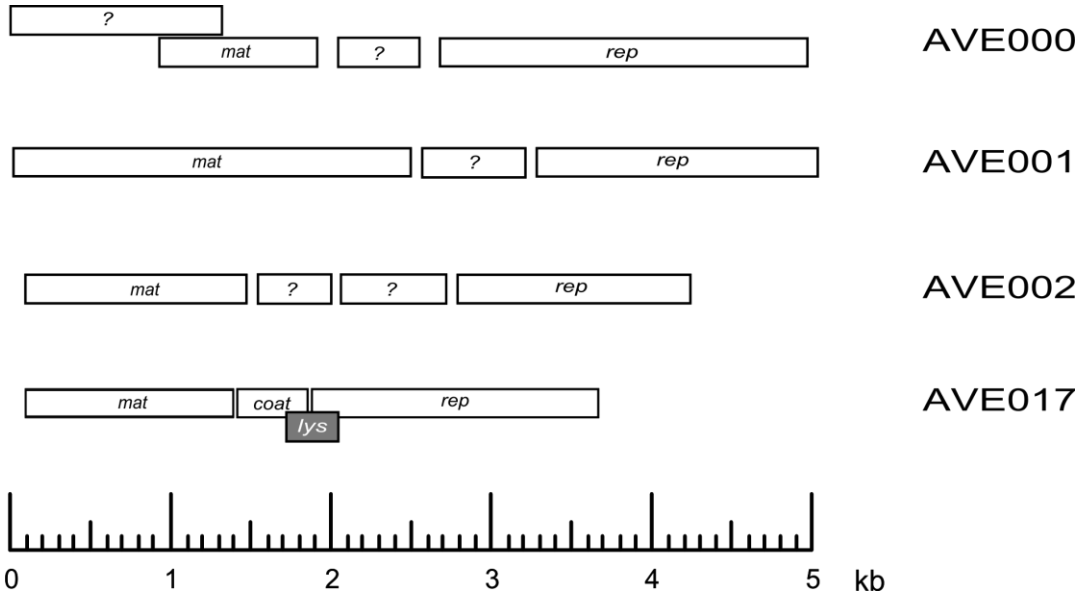

Figure 1.14 Genome organization of novel ssRNA phages with new genes and genetic architectures.
The core genes *mat*, *coat*, and *rep* that share similarity with known homologs are shown. While the novel genes with no assigned function are indicated with a question mark. The genome organization of AVE017 is similar to MS2 and the lysis protein from AVE017 shares 43% similarity with MS2 L.

CHAPTER II

MS2 LYSIS OF ESCHERICHIA COLI DEPENDS ON HOST CHAPERONE DNAJ

Introduction

The single stranded RNA phage MS2 is one of the simplest viruses, encoding just four proteins. Of the four proteins, three of them, RNA-dependent RNA replicase (Rep), major capsid protein (Coat), and maturation protein (A) are involved in viral replication and assembly. The fourth protein is the lysis protein (L), which causes lysis of the host and thus controls the length of infection cycle (14). *L* was the first gene in molecular biology shown to be embedded in two different genes, coat and replicase, in this case in the +1 reading frame of each gene (143) (Figure 2.1). Expression of *L* from a plasmid is necessary and sufficient to elicit lysis (135). *L* is one of the three canonical "single gene lysis" systems (SGL) used by small phages to effect lysis, the other two being *E* from the ssDNA phage ϕX174, representing the ubiquitous *Microviridae*, and A_2 from ssRNA phage Qβ, representing the *Alloleviridae* (14). Genetic and molecular analysis revealed that both E and A_2 inhibit specific steps in the PG biosynthesis pathway: A_2 inhibits MurA, which catalyzes the first committed step and E inhibits MraY, which catalyzes the formation of the first lipid-linked intermediate (119, 130, 131). In both cases, the isolation of dominant mutations conferring resistance to the lethal function of the lysis protein and the mapping of these mutations to the gene encoding the biosynthetic enzyme was the key to deciphering the lytic mechanism. In the case of *E*, the first and most common mutations conferring resistance to lysis mapped to a different gene, *slyD*, encoding a cytoplasmic, FKBP-type cis-trans peptidyl-prolyl

isomerase (128). These mutations were recessive, however, and subsequent investigation revealed that SlyD was required for the stability of E, which has 5 Pro residues in its 91 aa length. Suppressor mutations in *E* were all up-translation alleles, compensating for the instability of the E protein (129).

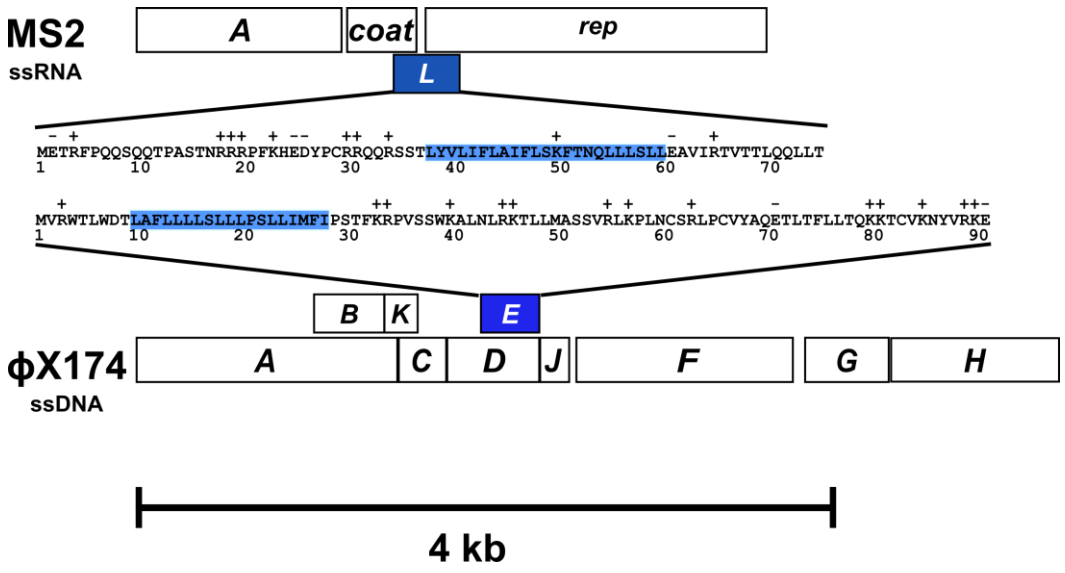

Figure 2.1 Genome organization of ϕX174 and MS2 phages and similarities between their lysis proteins.
Shown are the maps of MS2 and linearized ϕX174 genomes, drawn to scale. The lysis genes of the two phages are shaded blue. The lysis gene, *L*, of MS2 encodes a 75 aa lysis protein and the lysis gene *E*, of ϕX174 encodes a 91 aa protein. The residues spanning the TM domain are highlighted in blue. Basic and acidic residues are indicated above the primary structure of the lysis proteins.

In contrast to E and A_2, which have been called "protein antibiotics" because of the functional resemblance to cell wall antibacterial agents (14), no clear conceptual framework exists for the lytic function of the 75 aa L protein (Fig. 1). L has a hydrophilic N-terminal domain dominated by multiple basic residues and a hydrophobic C-terminal domain, presenting an interesting comparison with ϕX174 E (135). Genetic analysis had shown that the lytic function of E requires only the N-terminal hydrophobic domain and that the highly-charged, basic C-terminal domain could be replaced by unrelated sequences including β-galactosidase and GFP (42, 119). Drawing on this comparison, the van Duin group showed that expression of N-terminal deletions of L retaining as little as 42 C-terminal residues were fully lytic and, indeed, truncations retaining only the last 27 residues had partial function (135). Thus, E and L, although lacking any sequence similarity, seem to have mirror-image organization of functional domains. However, unlike E, induction of L did not lead to a block in PG synthesis, as assessed by incorporation of ^{3}H-diaminopimelate (144). Moreover, a synthetic peptide corresponding to the C-terminal 25 aa was reported to dissipate proton motive force (pmf) of *E. coli* inverted membrane vesicles and cause fluorescent dye leakage in reconstituted liposomes (136). Interestingly, despite the non-essential character of the N-terminal domain and the absence of any other secretory or membrane localization signals, it was reported that L was primarily localized in the periplasmic zones of adhesion between IM and OM, also known as Bayer's patches (137). Moreover, biochemical analysis of the murein from cells that had undergone L-mediated lysis was reported to have a slightly decreased average chain length of glycan strands and slightly

altered cross-linking between them. In addition, L-mediated lysis was blocked in cells grown at low pH, where penicillin-induced autolysis was also inhibited. Taken together, these results were interpreted as support for a general model where L somehow activates host autolytic enzymes, such as lytic transglycoslyases and D-D endopeptidases (138). Unfortunately, in the nearly three decades since these studies, no molecular link between the putative autolytic response, or indeed any host protein, and L has been established.

With the aim of identifying host factors involved in L-mediated lysis a genetic approach was undertaken. Here, we report the identification of one such factor, the host chaperone DnaJ, and the results are discussed in terms of a model for L function and regulation.

Materials and methods

Culture growth, antibodies, and chemicals

Unless indicated, LB broth and agar were used as growth medium. When indicated, media was supplemented with ampicillin (Amp), kanamycin (Kan), chloramphenicol (Cam), and tetracycline (Tet) at concentrations of 100, 40, 10, and 10 μg ml^{-1} respectively. Growth in liquid cultures and lysis was monitored by measuring at A_{550} as previously described (145). When indicated, isopropyl β-D-thiogalactopyranoside (IPTG) (Research Products International), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), and arabinose were added at a final concentration of 1 mM, 10 μg/ml and 0.4%, respectively. Primary antibodies against MS2 L peptide "TPASTNRRRPFKHEDC" were raised in rabbit (Sigma Genosys), mouse anti-His antibodies were purchased from Sigma, and rabbit anti-DnaJ antibodies were

Table 2.1 Strains, phages, and plasmids

Strain, phage, or plasmid	Relevant genotype or description	Reference
Strains		
XL1-Blue	*recA endA1 gyrA96 thi hsdR17 supE44 relA1 lac* [F'::Tn*10 proA+B+lacIq Δ(lacZ)M15*]	Stratagene
MG1655	*ilvG- rfb*-50 *rph*-1	(146)
TB28	MG1655 *lacIZYA <>frt*	(147)
JW0003	Δ*thrC*::*kan*	(148)
JW0014	Δ*dnaJ*::*kan*	(148)
RY15784	F'104 thr-leu Δ*leuA*::*cat*	Laboratory strain
RY15177	HfrH lacI^{q} *tonA*::Tn*10*	(134)
RY34179	TB28 Δ*dnaJ*::*kan*	This study
RY34314	TB28 Δ*thrC*::*kan*	This study
RY34356	RY34314 *thrC*$^{+}$ *dnaJ*$_{P330Q}$	This study
RY34154	TB28 pKC11	This study
RY34155	RY34154 [F'::Tn*10* proA+B+lacIq Δ(lacZ)M15]	This study

Table 2.1 Continued

Strain, phage, or plasmid	Relevant genotype or description	Reference
Phages		
MS2	ssRNA phage MS2	Laboratory stock
pBAD24	*bla araC* P_{ara}	(149)
pRE-L	*L* gene from MS2 cloned under the lambda late promoter pR'	(134)
pUC57-L^{syn} *lacZα*	Synthetic tandem *L* (L^{syn}) and *lacZα* genes cloned between EcoRI and HindIII	GenScript®
pKC11	*bla araC* P_{ara}*::*L^{syn} *lacZα*	This study
pKC12	*bla araC* P_{ara}*::his*$_6$ *-*L^{syn}	This study
pKC13	*bla araC* P_{ara}*::Lodj1*	This study
pKC14	*bla araC* P_{ara}*::Lodj2*	This study
pKC15	*bla araC* P_{ara}*::Lodj3*	This study
pKC16	*bla araC* P_{ara}*::his*$_6$ *-Lodj1*	This study

purchased from Enzo Life Sciences. Secondary Goat anti-mouse-HRP and goat anti-rabbit-HRP antibodies were purchased from Thermo Scientific. Unless otherwise indicated, all chemicals were purchased from Sigma-Aldrich.

Bacterial strains and bacteriophages

The bacterial strains and bacteriophages used in this study are described in Table 2.1 and primers are described in Table 2.2. The $dnaJ_{P330Q}$ allele was moved from the Ill1 mutant into the threonine-auxotroph RY34314 (TB28 Δ*thrC*::Kan^R) by P1 transduction (150), selecting for growth on M9 minimal agar supplemented with 0.2% glucose (151). The *dnaJ* locus was amplified from Thr^+ transductants using primers KC130 and KC131and sequenced to confirm the mutation. The *dnaJ* merodiploids were constructed by mating RY15784 [F'104 *thr-leu* Δ*leuA*::Cam^R] strain with either TB28 or RY34356 (TB28 $thrC^+$ $dnaJ_{P330Q}$) and selecting exconjugants on LB supplemented with both Kan and Cam. Similarly, RY34155 was constructed by mating RY34154 with XL1-Blue and selecting exconjugants on LB supplemented with both Amp and Tet. RY34179 strain was constructed by P1 transduction Δ*dnaJ*::Kan^R from JW0014.

Table 2.2 Primers

Primers	Sequence
KC18	atatatataagcttggatgctttgtgagcaattcg
KC19	atatatatgaattcaccatggaaacccgattcctcag
KC30	cggcagaaaagtccacattg
KC31	cacactaccatcggcgcta
KC130	ccggtgctgatgcttctgc
KC131	gcgcttatgggagtgatcc
KC149	gcatatatgaattcaccatgggcagcagccatcaccatcatca ccacagccaggatccgtcgaagacaaaaagaagttcaactc

Plasmids

The plasmids used in this study are in Table 2.1. A DNA fragment containing synthetic tandem *L* (L^{syn}) and *lacZα* genes was cloned under the pBAD promoter in the plasmid pBAD24, resulting in the plasmid construct pKC11. In this construct, both L^{syn} and *lacZα* genes were codon-optimized for *E. coli* expression (http://www.idtdna.com/CodonOpt) and the synthetic DNA with flanking EcoRI and HindIII sites at the 5' end and 3' was cloned into pUC57 at GenScript®. Using the same restriction sites, the synthetic fragment was moved into pBAD24 by standard techniques. The plasmid pKC12 was constructed by sub-cloning *his6-L* from pRE-His6-L (134). The

DNA of *his6-L* was PCR-amplified with primers KC18 and KC19, gel purified, digested with EcoR1 and HindIII, and cloned into pBAD24. The plasmids pKC13, pKC14, and pKC15 were obtained through selection for *Lodj* alleles (see below). pKC16 was constructed by site-directed mutagenesis (SDM) of pKC12 with primer KC149.

MS2 infection

Cultures of male strains were grown to A_{550} ~0.4 and then diluted to 0.1 in prewarmed LB medium. Then MS2 lysate was added to the freshly diluted cultures at indicated MOIs (multiplicity of infection).

Ill mutant selection

Cultures (25 ml) of RY34155 were grown to A_{550} of 0.2 and induced with arabinose. After ~2 h the lysate was harvested by centrifugation at 10,000 x g for 10 min. The pellet was washed once with PBS (phosphate buffered saline pH 7.2), and the survivors were allowed to recover overnight in 5 ml LB with appropriate antibiotics. The induction procedure was repeated and the survivors from the second round of induction were serially diluted in PBS and plated on LB agar plates supplemented with appropriate antibiotics, IPTG, X-Gal, and arabinose. A total of 6 blue colonies were isolated and screened for MS2 phage resistance by using cross-streaks as previously described (119). Phage-insensitive isolates were studied further by monitoring lysis phenotypes in induced cultures as above. To quantify the expression of L in the mutants, 1 ml sample was mixed with 111 μl of cold 100% trichloroacetic acid (TCA) as previously described (152). The precipitates were collected by centrifugation at 13,000 rpm for 10 min in a microcentrifuge. The pellets were washed with 1 ml cold acetone and air-dried. The

dried pellets were resuspended in 2X sample loading buffer with β-mercaptoethanol and boiled for 10 min. Equivalent amounts of protein were analyzed by SDS-PAGE and Western blotting as previously described (153). Both primary and secondary antibodies were used at 1:3000 dilution.

DNA sequencing and analysis

Sequencing for checking cloning and strain construction was done by Eton Biosciences (San Diego). For whole genome sequencing, genomic DNA extracted with the Qiagen gDNA kit was used to prepare 250 bp paired-end libraries using the Illumina TruSeq Nano DNA LT Library kit (Set A) according to the manufacturer's instructions. Whole genome sequencing was done at the DNA Sequencing Facility of the Institute for Cellular and Molecular Biology at UT Austin. The raw sequencing data was processed on Mutation Analysis Beta 1 (2014-06-27) available at www.cpt.tamu.edu/galaxy/workflow. Briefly, Bowtie2 was used to align the trimmed reads to the MG1655 reference sequence (accession number: NC_000913.3). To facilitate visual display of single nucleotide polymorphisms (SNP) or insertions/deletions (indel), the BAM output from Bowtie2 was processed through Mpileup and BCFtools to generate Variant Call Format (VCF), a standardized text file format. The SNP/indel variants with QUAL scores >100 in both parental (RY34155) and Ill mutants were scored as positive. The positive variants present in Ill mutants but absent in parental strain were further characterized.

Error-prone PCR mutagenesis and selection for *L* overcomes *dnaJ* (LodJ) mutants.

Error-prone PCR mutagenesis of the *L* gene was done using GeneMorph II Random Mutagenesis kit (Agilent Technologies) per the instructions provided in the kit. Briefly, ~900 ng of the *L* gene template (amplified from pRE-L) was used in the reaction with primers KC30 and KC31 to provide ~1 bp change/dsDNA molecule. The randomly mutagenized PCR product was gel-purified, digested with EcoRI and HindIII, ligated into pBAD24 digested with the same enzymes and electroporated into XL1-Blue. The transformants were pooled and the plasmid DNA was extracted using a Qiagen mini-prep kit to generate a library of *L* mutants. Plasmids carrying *Lodj* alleles that restored lysis in the $dnaJ_{P330Q}$ background were obtained by plasmid release. Briefly, ~100 ng of the mutant library was electroporated into RY34356.The transformants were pooled, diluted 1: 20,000, grown in 25 ml of LB supplemented with appropriate antibiotics to A_{550} ~0.2 and induced with arabinose. After 50 min, the culture medium was collected by centrifugation at 10,000 x g. Twenty mls of the supernatant medium was filtered through a 0.22 μm syringe filter (VWR) and then passed through a QIAprep 2.0 Spin Column (Qiagen). The bound DNA was eluted in 20 μl of sterile water. Ten μl of the released plasmids were retransformed into the same strain and the procedure was repeated for another two rounds. At the end of two rounds of amplification, the plasmid DNA was extracted and sequenced.

Pull-down of L and Lodj proteins

LB cultures (500 ml) of TB28 with appropriate plasmids were grown to A_{550} ~0.4, induced with arabinose at t=0 and harvested at allele-specific times by rapid

cooling in ice and centrifugation at 10,000 x g for 10 min: 20 min (*his*$_6$*Lodj*1 allele), 40 min (*L* in TB28), and 60 min (*L* in *dnaJ*^{P330Q}). The cell pellets were resuspended in ~3 ml PBS supplemented with the P8849 Protease Inhibitor Cocktail (Sigma; 1 μl/35 A_{550} units original culture) and passed through an Aminco French Pressure cell at 16,000 psi three times to lyse the cells. The lysate was centrifuged at 10,000 x g for 10 min to remove intact cells. The cleared supernatant was centrifuged at 100,000 x g in TLA 100.3 rotor (Beckman TL100 centrifuge) to collect membrane fractions. The membrane fraction was then resuspended in 1 ml of STE (50 mM Tris pH 8.0, 300 mM NaCl, 1% Empigen BB (Fluka), pH 8.0) and incubated overnight at 4°C with gentle mixing. The detergent extract was centrifuged at 100,000 x g to separate detergent solubilized proteins from detergent-insoluble components. The supernatant (~900 μl) was collected, mixed with 50 μl of Dynabeads® His-tag beads and incubated 5 min at room temperature on a roller drum. The binding and elution protocol was followed per manufacturer's instructions except that beads were washed 5 times in STE. A 20 μl volume from each elution was mixed with 20 μl of 2X sample loading buffer, heated at 100ºC for 10 min, and then analyzed by Western blot.

Results

Selection of Ill mutants

As our first attempt to identify host factors involved in L-mediated lysis, we cloned the *L* gene into a medium copy vector under the control of the lambda late promoter, pR', which in turn is driven by the inducible supply of the lambda late activator, Q, from a low copy lac-ara vector. Induction of cells carrying these two

plasmids (pQ and pRE-*L*) results in lysis in approximately the same time as infection by phage MS2, with comparable levels of L synthesis (Figure 2.2, 2.3). HfrH cells carrying these plasmids were mutagenized with EMS and subjected to two rounds of induced lysis before plating for survivor colonies. Of 3,300 colonies screened, 5 were found to be Lac+, Q+, $MS2^R$ and $M13^S$ (134). However, Hfr mapping and P1 transduction revealed that all five were recessive missense alleles of *pcnB*, encoding the polyA polymerase of *E. coli* (Figure 2.2). As expected, the *pcnB* defect reduced the copy number of the pRE-*L* plasmid to approximately single copy (not shown; (154)), which accounted for the survival of these mutants in our inducible plasmid system. Interestingly, these *pcnB* missense mutants were clustered in the polyA polymerase active site (Figure 2.3) and conferred dominant resistance to MS2 (Figure 2.2), suggesting a heretofore unsuspected role of PcnB in the MS2 infection cycle.

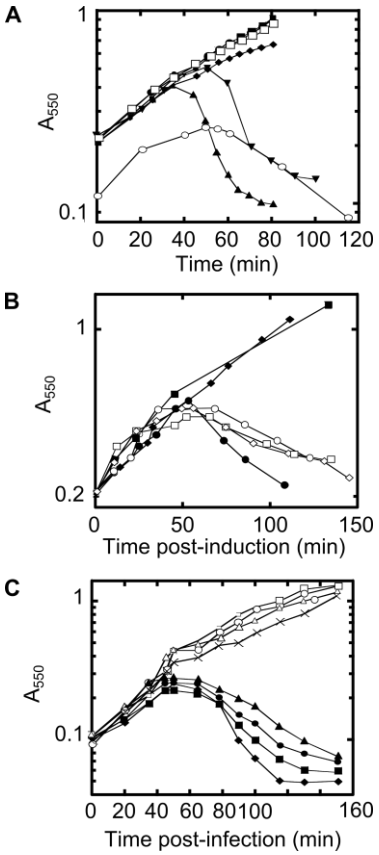

Figure 2.2 Expression of L from the dual plasmid system, pQ pRE-L, is sufficient for lysis

(A) Lysis profile from induction of the dual plasmid system with MS2 L under pR' promoter using 1mM IPTG or 1mM IPTG and 0.2% arabinose (ara) in RY15177 compared to lysis by MS2 infections of RY15177 at MOI of 5. Cultures were grown at 37°C. ●, pQ + IPTG; ■, pRE-*L* + IPTG; ♦, pQ pRE, + IPTG; ▼, pQ pRE-*L*, + IPTG; ▲, pQ pRE-*L*, +IPTG, +ara; □, pQ pRE-*L*, uninduced; ○, MS2. (B) *pcnB* alleles are recessive. Induction of plasmid-borne *L* using IPTG (final concentration of 1mM) at time 0. Host and plasmid: ●, *pcnB*$^+$, vector; ○, *pcnB*$^+$, pCA24N-*pcnB*; ■, BMC1, vector; □, BMC1, pCA24N-*pcnB*; ♦, BMC2, vector; ◊, BMC2, pCA24N-*pcnB*; (C) *pcnB* alleles confer dominant resistance to MS2 infection.

Figure 2.2 Continued. Infection of *pcnB* alleles with MS2 at MOI 5 at time 0. ●, *pcnB*$^{+}$, vector; ■, *pcnB*$^{+}$, pCA24N-*pcnB*; ♦, *pcnB*$^{-}$, vector; ▲, *pcnB*$^{-}$, pCA24N-*pcnB*; □, BMC1, vector; ○, BMC1, pCA24N-*pcnB*; Δ, BMC2, vector; x, BMC2, pCA24N-*pcnB*. The BMC1 and BMC2 *pcnB* mutations are defined in the legend to **Figure 2.3.** This figure was adapted from (134).

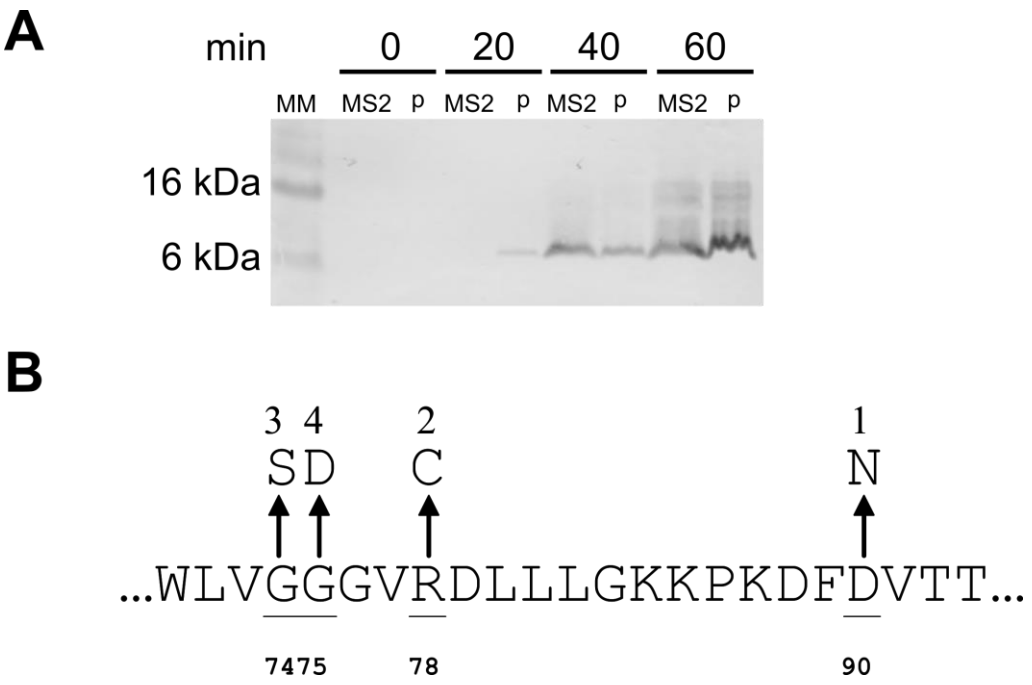

Figure 2.3 Plasmid-borne L is expressed at a level comparable to that attained in MS2 infected cells.
MM represents the molecular mass marker, with sizes in kDa indicated on the left. Samples were collected at the indicated times after induction of the plasmids with IPTG

Figure 2.3 Continued. (p) or infection with MS2 at an MOI of 5 (MS2). (B) The mutations of *pcnB* that confer the dominant MS2-insensitivity are shown as missense changes in the region defined by residues 72 through 93. The strains BMC1 and BMC2 in **Figure 2.2** have the allelic changes marked 1 and 2 in panel B. This figure was adapted from (134).

To avoid host mutants like *pcnB* that reduce either the plasmid copy number or expression of the L gene, we constructed a modified *L* expression vector, pKC11, with *L* and the *lacZ*α gene in tandem under an *ara* promoter. This allowed screening for blue survivor colonies after *L* induction. Expression of *L* from pKC11 was found to cause lysis at ~45 min after induction. From approximately 200 survivor colonies screened, 2 blue colonies with irregular morphology were isolated on X-gal/arabinose plates. Both isolates exhibited an absolute plating defect for MS2 (not shown). We next tested the kinetics of lysis in liquid culture in comparison with the parental host and found that the mutants displayed a 40-50 min delay in the onset of lysis (Figure 2.4). To eliminate the possibility that the mutants confer a non-specific resistance to lysis, we tested lysis by E of ϕX174 and A_2 of Qβ in the mutant hosts and found no significant delay in either case (not shown). Thus, the mutants confer partial resistance specific to L-mediated lysis, enough to delay its onset, and were designated Ill mutants (*insensitivity to L lysis*).

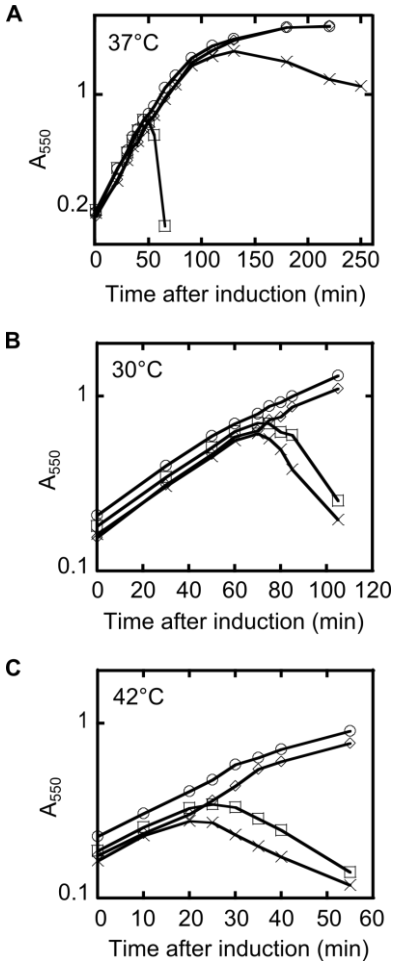

Figure 2.4 L-mediated lysis in *dnaJ*$_{P330Q}$ background is delayed and the allele is recessive to wild-type.
(A) The lysis profile of L in wild-type and *dnaJ*$_{P330Q}$ background at 37°C. Cultures were grown to A_{550} ~0.2 and induced with arabinose 0.4% (w/v) final. Symbols: ○, TB28 and empty vector; ◊, *dnaJ*$_{P330Q}$ and empty vector; □, TB28 and *L*; X, *dnaJ*$_{P330Q}$ and *L*. (B and C) Same as (A) except the strains are merodiploid for *dnaJ* and the lysis profiles were taken at 30°C (B) and 42°C (C).

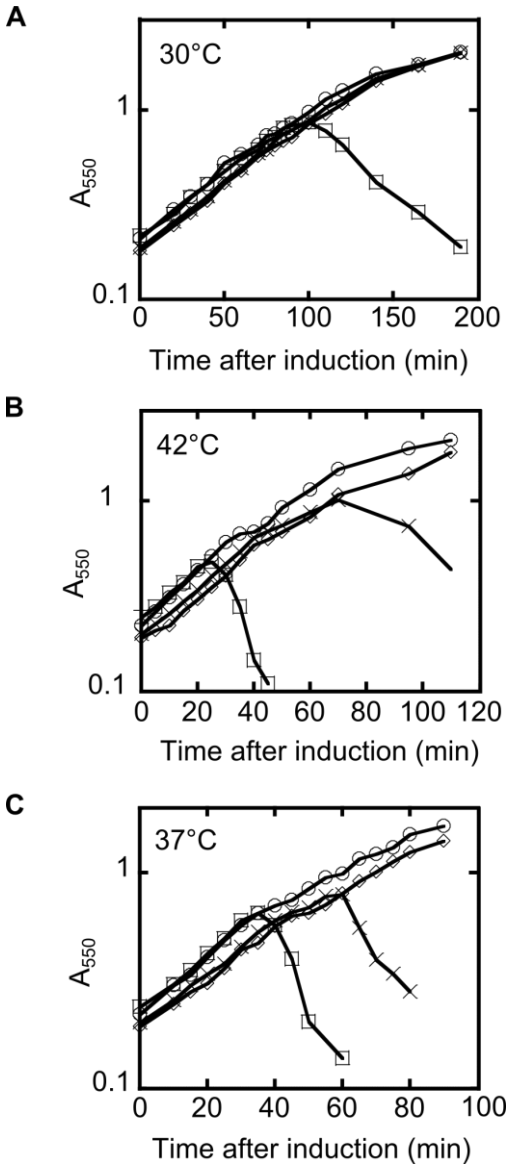

Figure 2.5 L-mediated lysis in *dnaJ*$_{P330Q}$ and Δ*dnaJ* backgrounds.
(A) The lysis profiles of L in wild-type and *dnaJ*$_{P330Q}$ backgrounds at 30°C. Cultures were grown to A_{550} ~0.2 and induced with arabinose 0.4% (w/v) final. (B) Same as (A) except the lysis profiles were taken at 42°C. (C) The lysis profile of L in wild-type and Δ*dnaJ* backgrounds at 37°C. ○, TB28 and empty vector; ◊, *dnaJ*$_{P330Q}$ or Δ*dnaJ* and empty vector; □, TB28 and *L*; X, *dnaJ*$_{P330Q}$ or Δ*dnaJ* and *L*.

The Ill phenotype is due to a *dnaJ* mutation

In order to assign the Ill phenotype to a specific host gene, genomic DNA from two Ill mutants was purified and subjected to whole genome sequencing. An analysis of the genomic data revealed that both the Ill mutants, which may have been siblings, had a P330Q missense mutation in *dnaJ*. To determine whether the *dnaJ* mutation was the sole factor responsible for the lysis phenotype, the mutation was transduced to a new background. The newly constructed $dnaJ_{P330Q}$ mutant retained the observed blue-colony survivor phenotype of the parental mutant (not shown). Moreover, the $dnaJ_{P330Q}$ mutation exhibited strictly recessive behavior, with L-mediated lysis fully restored in a merodiploid (Figure 2.4).

Since DnaJ is a heat shock protein (155), we examined the effect of temperature on the kinetics of L-mediated lysis in wild type and mutant *dnaJ* backgrounds. As shown in Figure 2.5, the $dnaJ_{P330Q}$ allele exhibited an absolute lysis defect at 30°C, as well as significant lysis delays at higher temperatures (Figure 2.5). In addition, the Δ*dnaJ* allele exhibited a more modest delay in lysis when compared with the lysis profile in the Ill mutant (Figure 2.5). Immunoblot analysis showed that L accumulation paralleled the lysis defect at 37°C and 42°C, but was normal at 30°C in the wild-type and $dnaJ_{P330Q}$ backgrounds (Figure 2.6).

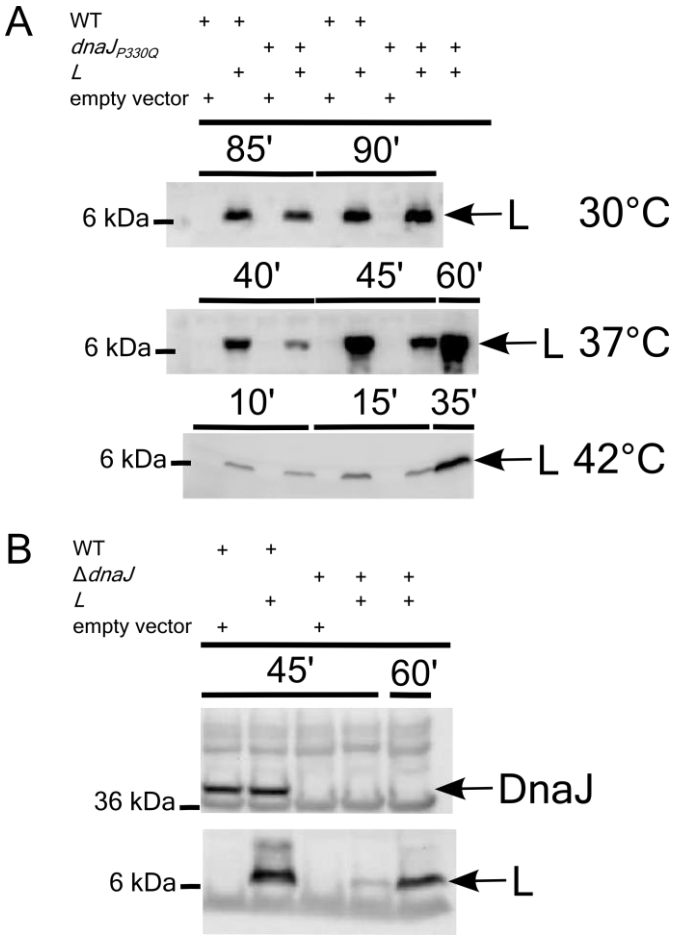

Figure 2.6 L accumulation in *dnaJ* mutants.
(A) L accumulation is delayed in *dnaJ*$_{P330Q}$ background when compared to wild-type at 37°C and 42°C but not at 30°C. (B) L accumulation in wild-type and Δ*dnaJ* at 37°C (bottom). The absence of DnaJ in Δ*dnaJ* is shown by blotting with anti-DnaJ antibody (top). The bands corresponding to L and DnaJ are indicated with an arrow on the right. The molecular mass standard in kDa is represented on the left. The time points of sample collection are shown directly above the blot.

Lodj mutants by-pass DnaJ

To examine L dependence on DnaJ, we sought suppressors in *L* that could overcome the absolute lysis defect at 30°C in the $dnaJ_{P330Q}$ host. To achieve this, we used the plasmid release technique (see Methods) to enrich for *L* alleles with restored lytic competence in the mutant background. Plasmid DNAs prepared from individual lytic transformants, designated as *Lodj* (*L overcomes dnaJ*) alleles, were sequenced. Each *Lodj* mutant plasmid was found to have a single nucleotide deletion that created an *L* gene encoding a protein in which the entire N-terminal half of L was replaced by a few N-terminal residues of the Rep protein (Figure 2.7). To verify that *Lodj* alleles were indeed lytically functional, we cloned the new truncated *L* genes into an inducible plasmid and showed that the lysis timing supported by the *Lodj* alleles was comparable in both parental and $dnaJ_{P330Q}$ mutants hosts at 30°C as well as 37°C (Figure 2.7). To rule out the possibility that it is the presence of the first few amino acid residues of Rep that confers the ability to cause lysis in the $dnaJ_{P330Q}$ background, we used site-directed mutagenesis to replace the first 36 codons of L with His-tag (Figure 2.7), and showed that this construct also functions as a *Lodj* allele (not shown). Taken together, these results indicate that it is the presence of the dispensable, N-terminal, highly basic domain of L that confers the requirement for DnaJ.

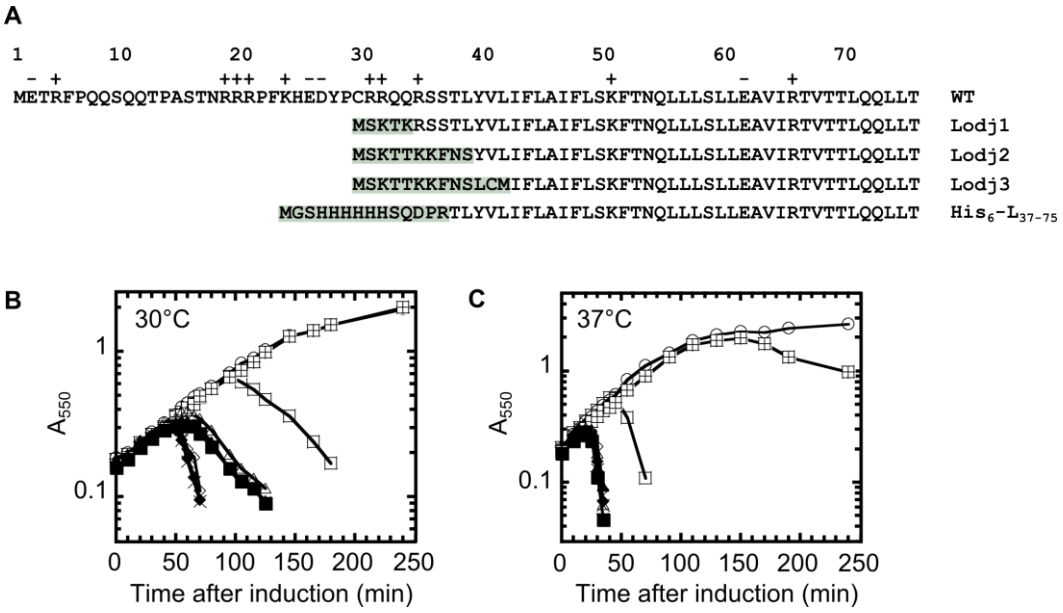

Figure 2.7 The *Lodj* alleles suppress lysis block in *dnaJ*$_{P330Q}$ mutants.
(A) Primary structure of L aligned with three Lodj proteins and His$_6$-L$_{37-75}$. Amino acids that are different from wild type L are highlighted. (B and C) The lysis profiles of L and Lodj1-3 in wild-type and *dnaJ*$_{P330Q}$ backgrounds at 30°C (B) and 37°C (C). ○, TB28 and empty vector; □, TB28 and *L*; ⊞, *dnaJ*$_{P330Q}$ and *L*; ◊, TB28 and *Lodj1*; X, TB28 and *Lodj2*; ♦, TB28 and *LodJ3*; ▲, *dnaJ*$_{P330Q}$ and *Lodj1*; Δ, *dnaJ*$_{P330Q}$ and *Lodj2*; ■, *dnaJ*$_{P330Q}$ and *Lodj3*.

DnaJ interacts with MS2 L

To further investigate the role of DnaJ in L-mediated lysis, we asked if DnaJ and L formed a complex. We constructed both *H_6-L* and *H_6-Lodj*, encoding L and Lodj proteins tagged with an N-terminal hexahistidine sequence, and showed that both were fully functional (not shown). Cultures carrying these plasmid-borne alleles were induced and membrane extracts prepared from samples taken immediately before lysis. Pull-down assays using Dynabeads® (see Methods) showed that DnaJ was associated with L in the parental cells but not in the *$dnaJ_{P330Q}$* Ill mutant (Figure 2.8); this association was abrogated for the LodJ mutant in both parental and Ill mutant backgrounds. Taken together, these results indicate that L and DnaJ form a membrane-associated complex in vivo, and that the complex depends on the non-essential N-terminal segment of L interacting with the C-terminal domain of DnaJ.

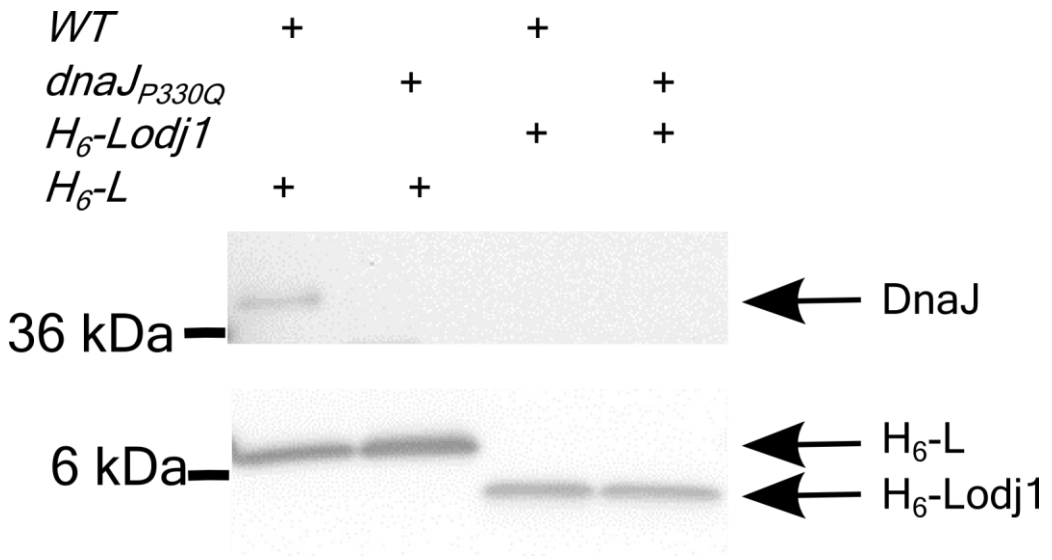

Figure 2.8 DnaJ interacts with L but not with Lodj.
Membrane extracts containing H_6-L or H_6-Lodj1 were solubilized in detergent were bound to anti-his Dynabeads® and the proteins bound to the beads were analyzed by Western blotting with anti-DnaJ (top) and anti-His antibodies (bottom). The bands corresponding to DnaJ, H_6-L, and H_6-Lodj1 are indicated by an arrow on the right. The molecular mass standards are represented on the left.

Discussion

The *Leviviridae*, among the most ancient and possibly the simplest known viruses, effect rapid and efficient lysis of the host without inhibiting cell wall synthesis through the action of the L protein (144). For Gram-negative bacteria, this is the only known lysis phenomenon mediated by antibiotics or by phage that does not involve either inhibition of PG synthesis or elaboration of a muralytic enzyme by the phage. Understanding of how L triggers an autolytic response and identifying the host factors involved in the response might reveal an entirely new perspective for novel antibiotics.

Prior perspectives on L function

Most of the extant work on L-mediated lysis involved a combination of biochemical and electron microscopy analysis of the murein structure in both intact and L-lysed bacteria, leading to a model in which L protein localized to the periplasmic zones of adhesion and caused the inappropriate activation of cellular autolytic functions such as lytic transglycosylases and D-D endopeptidases (137, 138). Moreover, it was shown that a defect in membrane-derived oligosaccharide (MDO) biosynthesis provided resistance to L-lysis, possibly by impeding appropriate localization of the lysis protein (156). However, this model lacked genetic evidence and afforded no clear molecular framework that could lead to mechanistic understanding. In addition, the very existence of zones of adhesion in growing cells has become controversial, since they have not been observed in cryo-electron microscopic images but only in cells that have undergone dehydration and fixation for transmission EM (139). Importantly, the MDO-dependency was not specific to L-mediated lysis, as the lysis protein E of the ssDNA phage ϕX174

displayed a similar dependency (156). At the time, E was also thought to induce autolysis (157). However, we have since shown unambiguously that E is a specific inhibitor of MraY and causes lysis by inhibiting the biosynthesis of Lipid II (119, 131). Thus, it seems likely that the MDO sensitivity of these lysis pathways, one of which involves a blockage of PG synthesis and the other does not, is indirect and non-specific.

The role of DnaJ in L-mediated lysis

Here we have taken a genetic approach to identify host factors required for L lysis, by the conceptually simple approach of selecting for host mutants resistant to L function. The results show that the host chaperone DnaJ is one such factor. We have shown that the P330Q missense change in DnaJ, although preserving the essential heat shock function of DnaJ, confers a defect in L-mediated lysis, absolute at 30ºC and partial at 37ºC. Proline 330 is the most conserved residue in the C-terminal domain of DnaJ, present in 689 of the 690 full-length DnaJ sequences from Proteobacter (http://eggnog.embl.de/version_3.0/index.html). The one exception, from a 2003 genomic sequence of *S. flexneri*, is probably a sequencing error, since more recent *S. flexneri* genomes do not show this variance. Little is known about the biological function of the C-terminus. Proline 330 was chosen as the end of the last DnaJ subdomain before a putative extreme C-terminal domain that was shown to be required for dimerization of DnaJ (158).

The lysis-defective phenotype of the Ill mutant, $dnaJ_{P330Q}$, is L-specific, with no effect on lysis by E or A_2, the other two prototypic single gene lysis proteins. Since the delayed lysis phenotype is specific for L-mediated lysis and not a general defect, the

simplest notion is that the missense change in DnaJ abrogates an important interaction between DnaJ and L. This idea is supported by the finding that DnaJ, but not DnaJ$_{P330Q}$, co-purifies with his-tagged L.

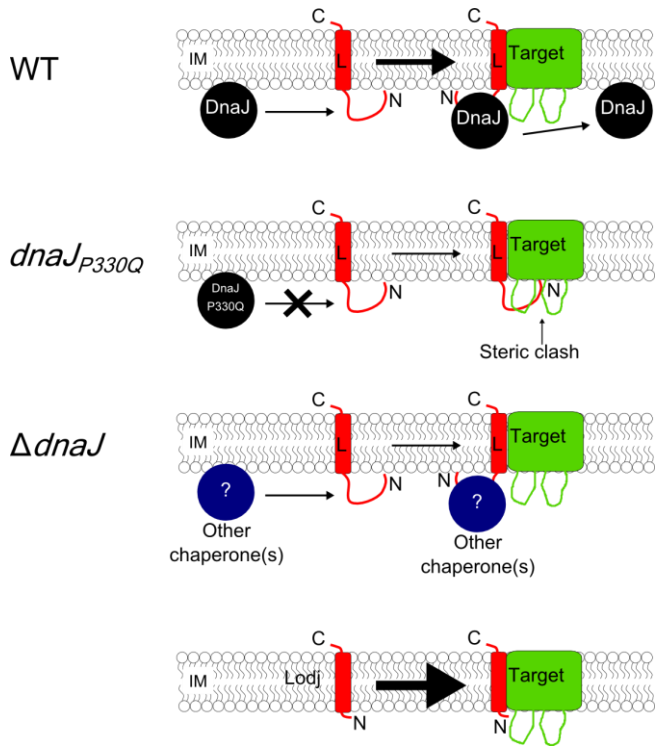

Figure 2.9 Model for the role of DnaJ in L-mediated lysis.
DnaJ interacts with highly basic N-terminal domain of L at the membrane, possibly to remove steric constraints inhibiting the interaction of L (red) with its target (green). The DnaJ$_{P330Q}$ variant loses its ability to interact with L, leading to a less stable interaction between L and its target, resulting in a delayed and gradual onset of lysis. However, in the complete absence of DnaJ, compensatory cellular chaperone activities stabilize the

(**Figure 2.9 Continued**) L-target interaction, resulting in a modest delay in lysis. The products of the Lodj alleles lack the N-terminal domain and thus bypass the requirement for DnaJ. The thickness of the arrows indicates favorability for complex formation: the thicker the arrow, the more favorable the interaction between L and its target.

What role does DnaJ play in L-lysis? It is clearly not the target of L, since deletions of the N-terminal domain of L relieve the dependency on DnaJ and restore normal L-mediated lysis, indeed the Lodj derivatives, expressed from isogenic plasmid environments, evoke lysis much faster than the parental full-length proteins. The simplest idea is that DnaJ is required for proper folding of full-length L but not of the truncated Lodj proteins. Figure 2.9 shows a cartoon rendition of the model. In the WT host, DnaJ interacts with the improperly-folded N-terminal hydrophilic domain, thus avoiding a steric clash with a putative cytoplasmic domain of the (still-unknown) target protein. Once the complex is formed, DnaJ can dissociate and catalyze further L-target events. In support of this notion, there is precedent for DnaJ acting as a chaperone independent of the DnaK-ATPase heat shock activity, in stabilizing a Lys/Arg rich C-terminal domain of TorI, the recombination directionality factor of the KplE1 cryptic prophage of *E. coli* (159). In the Δ*dnaJ* host, compensatory elevations in other, less-specific chaperones, would lead to similar folding events, although with slower kinetics in the absence of the specific interaction. By contrast, the Lodj truncations lack the N-terminal domain and thus do not suffer from the potential steric clash with the target.

From this perspective, the parallel with the lysis protein E and the cytoplasmic chaperone SlyD is striking. SlyD is absolutely required for the proteolytic stability of E, which like L, has a large dispensable domain rich in charged and hydrophilic residues. Like L, removal of this dispensable domain also abrogates the chaperone dependence of E, although the position of the dispensable domain, at the C-terminus in E and the N-terminus in L, is opposite in the two lysis proteins. We suggest these dispensable domains, evidently requiring chaperone activity for proper folding, have evolved as regulatory "damping" features of the two lysis proteins. Among general phage functions like genome replication and virion morphogenesis, lysis is distinct in that maximum efficiency and speed is undesirable. Moreover, in these simple phages with highly constrained genome sizes, both the E and L lysis genes were forced to evolve in alternate reading frames of essential genes. It is logical that the smallest possible lytic domain would emerge from such evolutionary constraint, and, indeed, the essential segments of E and L are the N-terminal 32 and C-terminal 30 residues, respectively. In these severely restricted out-of-frame contexts, it would be much less challenging to evolve a highly-charged region lacking secondary structure and thus compromised in terms of unassisted folding. The acquisition of the crippled domain conferred a requirement for interaction with a cytoplasmic chaperone and thus provided a context for retarding lysis to allow time for assembly of progeny virions. Indeed, this scenario would also confer potential for physiological regulation of lysis timing in the E and L systems via the level or activity of the respective chaperone.

CHAPTER III

MUTATIONAL ANALYSIS OF THE MS2 LYSIS PROTEIN L

Introduction

Small lytic phages with single-stranded nucleic acid chromosomes achieve lysis and release of the progeny virions by the expression of a single gene (14). These phages can be broadly classified into ssDNA and ssRNA phages. The ssDNA phage ϕX174 is the prototype of the ubiquitous family *Microviridae*, with a 5.4 kb genome and ten genes. In the ϕX174 genome, the lysis gene *E* is entirely embedded the +1 reading frame of the essential assembly gene *D* (22). The ssRNA phages defined the family *Leviviridae* and were traditionally further sub-divided into two genera, Allolevivirus and Levivirus, represented by the prototype male-specific coliphages MS2 and Qβ (Figure 3.1). These phages are the simplest viruses, with 3.5- 4.3 kb genomes and three core genes encoding an RNA-dependent RNA polymerase or replicase (Rep), a major capsid protein (Coat), and Mat, the maturation or attachment protein (one molecule per virion, named as A in MS2, A_2 in Qβ). The two genera differ in the structure and role of a fourth essential gene in each case, with the Alloleviviruses having the *A1* gene, a translational read-through extension of coat, and the Leviviruses having the *L* lysis gene, with a reading frame overlapping and out-of-frame with the end of *coat* and the beginning of *rep*. Moreover, the Mat protein of Qβ (A_2) was found to be necessary and sufficient for lysis (160).

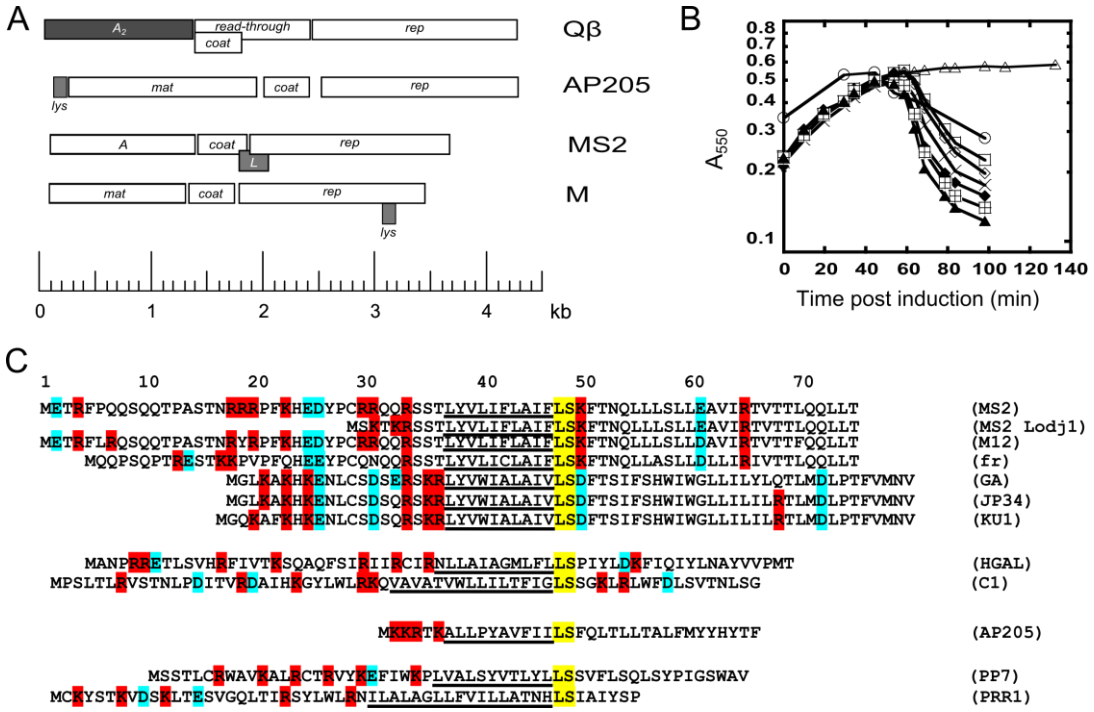

Figure 3.1 Cysteine scanning and alignment of L homologs identifies the importance of the conserved LS motif.
(A) Genome maps of four ssRNA phages representing the four known genetic architectures of lysis genes. The lysis genes are highlighted in grey. (B) Lysis profiles of various cysteine scanning mutants. ▲, *L*; ⊞, L^{C29S}; □, L^{S9C}; ◊, L^{S15C}; ♦, L^{S35C}; ▲, L^{S49C}; X, L^{S58C}; ○, L^{L73C}. (C) L-like lysis proteins from different leviviruses are aligned with respect to the conserved LS motif (yellow), preceded by a stretch of hydrophobic residues (underlined) and highly basic N-termini. Basic and acidic residues are highlighted in red and blue respectively.

The genetic simplicity of the ssDNA and ssRNA phages made it straightforward to define the cistron required for lysis. The term amurin has been proposed for these lysis proteins, which lack muralytic activity but somehow subvert PG integrity (14). The mechanism by which amurins effect lysis remained controversial for decades. One model supported by multiple reports was that ϕX174 E formed a "membrane tunnel" that traversed the entire envelope and allowed release of the cytosolic contents, including the assembled virions (120, 157). Ultimately a genetic approach based on selection of host mutants resistant to the induction of the amurin genes cloned under a plasmid-borne inducible promoter was successful (119). Remarkably, both the ϕX174 E and Qβ A_2 amurins were found to be specific inhibitors of enzymes of the conserved pathway for murein precursor biosynthesis: A_2 inhibits MurA, the first committed step of the pathway, whereas E inhibits MraY, which catalyzes the formation of the first lipid-linked precursor (122, 130, 161). More recently, a Levivirus specific for the conjugational pilus of an IncM R-factor was shown to have a lysis cistron, lys^M, that evolved in a different place in the genome (141). By use of a similar approach, it was shown that Lys^M is a specific inhibitor of MurJ, the Lipid II flippase of *E. coli* (Chapter IV). Thus, for all three of these small phages, the expression of these proteins, either from plasmid-cloned cistrons or in the context of phage infection, caused blockage of the flow of murein precursors, generally leading to lysis as a result of failed septation events. The obvious similarity to murein-specific chemotherapeutic agents has led to these lysis proteins being designated as "protein antibiotics" (14).

Despite this success with three single gene lysis systems, the function of the original amurin, MS2 L, remains unclear. L is a 75 amino acid protein that has been reported to be present in the membrane fractions (82). Work from the van Duin group using plasmid-borne L genes showed that the highly basic N-terminal half of the protein is dispensable for lytic activity, whereas C-terminal truncations were found to be non-functional (135). An experiment with a synthetic peptide comprising of C-terminal 25 amino acids of L was reported to dissipate proton motive force (pmf) of *E. coli* inverted membrane vesicles and cause fluorescent dye leakage in reconstituted liposomes (136). Surprisingly, immuno-electron microscopy studies using an antibody raised against a hybrid L protein, suggested that L was enriched in zones of adhesion between IM and OM, also known as Bayer's bridges (137). Moreover, biochemical analysis of the murein in cells lysed by L was reported to have a decreased average chain length of glycan strands and altered cross-linking (138). These disparate observations were not easily reconciled into a clear conceptual framework. However, it was proposed that L causes pmf-depleting lesions in the IM, thereby somehow activating host autolytic enzymes such as lytic transglycoslyases and D-D endopeptidases (138). Although this "autolysis" remains as an attractive general model for L function, neither an operational schema or molecular details for such a pathway have been forthcoming.

Recently a genetics-based approach, analogous to the approach used for E and A_2, was employed to address the host target of the L amurin. Although a cellular target was not identified, it was demonstrated that the lytic function of L requires an interaction with the host chaperone DnaJ, and that this requirement could be bypassed by deletion of

the basic N-terminal domain (Chapter II). Here we report a comprehensive genetic analysis and identify residues and regions of L that are critical for lysis. Results are discussed in terms of a model for L function.

Table 3.1. Strains and plasmids

Strain and plasmids	Relevant genotype or description	Reference
Strains		
XL1-Blue	*recA endA1 gyrA96 thi hsdR17 supE44 relA1 lac* [F'::Tn10 *proA+B+lacIq Δ(lacZ)M15*]	Stratagene
MG1655	*ilvG- rfb*-50 *rph*-1	(146)
TB28	MG1655 *lacIZYA <>frt*	(147)
Plasmids		
pBAD24	*bla araC* P_{ara}	(149)
pBAD33	*cat araC* P_{ara}	(149)
pQ		(134)
pRE-*L*	*L* gene from MS2 clone under the lambda late promoter pR'	(134)
pRE-L_{C29S}	pRE-*L* with C29S mutation	(134)

Table 3.1 Continued

Strain and plasmids	Relevant genotype or description	Reference
pRE-$L_{C29S, S9C}$	pRE-L_{C29S} with S9C mutation	(134)
pRE-$L_{C29S, S15C}$	pRE-L_{C29S} with S15C mutation	(134)
pRE-$L_{C29S, S35C}$	pRE-L_{C29S} with S35C mutation	(134)
pRE-$L_{C29S, S49C}$	pRE-L_{C29S} with S49C mutation	(134)
pRE-$L_{C29S, S58C}$	pRE-L_{C29S} with S58C mutation	(134)
pKC12	*cat araC* P_{ara}::*his6-*L^{syn}	(Chapter II)
pKC17	*cat araC* P_{ara}::*his6-*L^{syn}	This study
pBAD24 *L*	*bla araC* P_{ara}::*L*	This study

Materials and methods

Bacterial strains, culture growth, plasmids, and reagents

The bacterial strains used in this study are described in Table 3.1 and primers are listed in Table 3.2. LB broth and LB agar were used as growth medium and were supplemented with appropriate antibiotics and inducers. When indicated, ampicillin (Amp), chloramphenicol (Cam), and arabinose were added to the growth media at concentrations 100, 10 μg ml^{-1}, and 0.2-0.4% (w/v) respectively. Bacterial growth and lysis were monitored as previously described (145). The *his6-*L^{syn} gene from pKC12 (*bla araC* P_{ara}::*his6-*L^{syn}) was amplified with primers KC19 and KC30 using Phusion DNA

polymerase (New England Biolabs). The PCR product was gel-purified and digested with enzymes KpnI and HindIII and ligated into pBAD33 digested with the same enzymes to generate pKC17 (*cat araC* P_{ara}::*his6-*L^{syn}). Rabbit polyclonal serum raised against MS2 L peptide "**TPASTNRRRPFKHEDC**" and goat anti-rabbit-HRP (ThermoFisher Scientific) were used as primary and secondary antibodies respectively. Unless otherwise indicated, all chemicals were purchased from Sigma-Aldrich.

Table 3.2 Primers

Primer	Sequence
KC19	ATATATATGAATTCACCATGGAAACCCGATTCCTCAG
KC30	CGGCAGAAAAGTCCACATTG
KC31	CACACTACCATCGGCGCTA

Selection of non-functional mutants

The *L* gene was randomly mutagenized and cloned into pBAD24 as described in Chapter II. The ligated plasmids were transformed into XL1-Blue and plated on LB agar plates supplemented with Amp and arabinose. The colonies that survived on the inducer plates were picked, grown in liquid culture, the plasmid DNA was extracted using a Qiagen mini-prep kit, and the plasmid DNA was sequenced at Eton Biosciences (San Diego) with primers KC30 and KC31. The plasmids with a single-missense change in

the *L* gene were transformed into TB28 and the lysis profiles of the mutants were compared with that of the wild type allele. The mutants that did not cause lysis in liquid culture were scored as non-functional alleles of *L*.

Monitoring accumulation of *L* allele products

To quantify the expression of mutant alleles of *L*, 1 ml sample was collected at 40 min post-induction and mixed with 111 μl of cold 100% TCA as previously described (153). The TCA precipitates were collected by centrifugation at 13,000 rpm for 10 min in a microcentrifuge. The pellets were washed three times with 1 ml cold acetone and air-dried. The dried pellets were resuspended in 2X sample loading buffer with β-mercaptoethanol and boiled for 10 min. Normalized amounts of protein (~0.4 A_{550} units) were analyzed by SDS-PAGE and Western blotting as previously described (153). The antibodies against MS2 L and goat anti-rabbit-HRP were used at 1:3000 dilution.

Membrane fractionation

Cultures of TB28 (500 ml) harboring plasmids with *L* alleles were grown to an A_{550} ~0.2, induced with arabinose, and the cells were centrifuged at 10,000 x g (Sorvall LYNX 6000 Superspeed Centrifuge) for 10 min. The pellets were resuspended in ~3 ml PBS (phosphate buffered saline pH 7.2) supplemented with Protease Inhibitor Cocktail (Sigma; 1 μl/35 A_{550} units original culture) and the cells were lysed by passing three times through an Aminco French Pressure cell at 16,000 psi. The intact cells were cleared from the lysate by centrifuging at 10,000 x g and 500 μl of the supernatant was saved as total fraction. The rest of the supernatant (~2.5 ml) centrifuged at 100,000 x g in TLA100.3 rotor (Beckman TL100 centrifuge) for 1 h and the supernatant and

membrane pellet were collected. The three fractions were normalized to ~0.4 A_{550} units, mixed with 2X sample loading buffer and boiled at 100°C for 10 min. The samples were resolved on SDS-PAGE gels, transferred on to PVDF membranes and Western blotted as described in Chapter II.

Results

Discovery of a serine residue essential for function allows alignment of diverse L-like amurins

In order to facilitate both cysteine-accessibility studies and also experiments with heterobifunctional thiol-specific cross-linkers to identify binding partners, we conducted site-directed mutagenesis on gene L to create a variety of single-Cys alleles, including five Ser-Cys substitutions in the background LC29S where the single natural Cys codon was replaced by Ser. Although most of the substitutions had no effect on L lytic function, a single change, S49C, conferred an absolute lysis defect, both in the LC29S (not shown) and the parental L contexts (Figure 3.1), without affecting L accumulation (not shown). We noticed that this essential Ser residue could be used as part of a Leu-Ser dipeptide motif for aligning the lysis protein sequences not only from the closely related F-specific Leviviruses but also from Leviviruses that use different conjugational or Type IV pili for infection (*E. coli* phages HGAL and C1; *Pseudomonas* phages PP7 and pRR1) and from the *A. baumannii* phage AP205, where the lysis gene evolved in a distinct genomic location (Figure 3.2). These lysis genes have been shown to function in *E. coli* (83, 162, 163). In this alignment, although sequence similarity could not be generally detected, organizing the sequences using the LS motif allowed a putative

domain structure for L-like amurins to be assigned: (1) a positively charged N-terminus (4-36 aa, 3-6 net positive charges); (2) a hydrophobic sequence rich in aromatic and large beta-branched aliphatic residues (10 -17 aa); (3) the LS dipeptide; and (4) a phage-specific C-terminal domain of variable length (4 to 31 residues) but either neutral or containing one net charge. Based on the previous truncation analysis of L and our results indicating the N-terminal domain confers a dependency on the host DnaJ chaperone, Domain 1 would be dispensable (Chapter II).

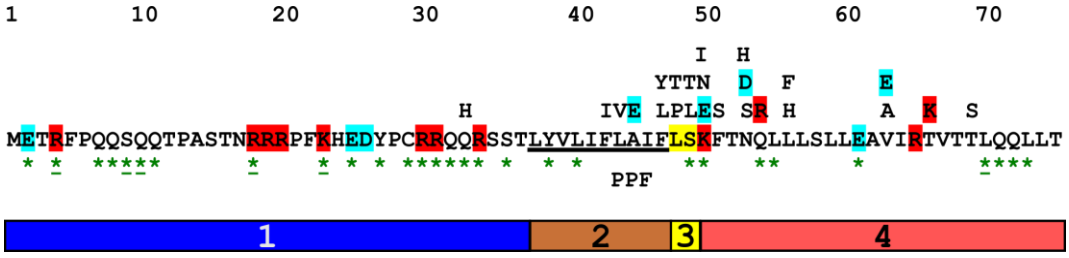

Figure 3.2 Mutational analysis of MS2 *L*.
The L primary structure is represented same as in **Figure 3.1**. The four domains of L are represented by numbered boxes. Domain 1 in blue is the highly-charged domain of L and it is dispensable for function. Domain 2 (brown) is the region composed mainly of hydrophobic residues and it is followed by a conserved LS motif (Domain 3). Domain 4 is variable in length in different L-like proteins. The L-like proteins from phages that

(**Figure 3.2 Continued**) infect through F-pilus all have their Domain 4 (pink) of about the same length. While, L- like proteins from phages that infect through pili other than F have much shorter C-termini. Missense alleles with lysis defects but without a defect in protein accumulation are indicated above the L sequence. Missense changes that do not affect lytic function are indicated below the L sequence. Green asterisks indicate all possible codon positions where a nonsense mutation could be accessed by a single nucleotide change; underlined asterisks indicate positions where no nonsense mutants were obtained in the mutagenesis. Basic and acidic residues are highlighted in red and blue respectively. The numbers above the primary structure indicate the position of the residues within the L protein.

Table 3.3 Single-missense alleles of *L*

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
90	3	G→T	1	M→I	-	-
185	3	G→A	1	M→I	-	-
205	2	T→C	1	M→T	-	-
261	4	G→T	2	E→Stop	-	N.D.
190	8	C→T	3	T→I	-	-
223	7	A→T	3	T→S	-	-
326	17	C→T	6	P→L	-	-
127	22	C→T	8	Q→Stop	-	N.D.
139	24	A→T	8	Q→L	-	-
288	23	A→T	8	Q→L	-	-
391	31	C→T	11	Q→Stop	-	N.D.
65	38	C→T	13	P→L	+	+
345	38	C→T	13	P→L	+	+
331	43	T→G	15	S→A	+	+
50	52	A→G	18	R→G	+	+
177	53	G→T	18	R→I	+	+
172	74	A→G	25	E→G	+	-

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
291	52	A→T	18	R→Stop	-	N.D.
48	55	C→A	19	R→S	+	-
67	56	G→A	19	R→H	+	-
142	58	C→T	20	R→W	+	-
170	58	C→T	20	R→W	+	-
313	59	G→T	20	R→L	+	-
183	67	A→G	23	K→E	+	-
319	67	A→G	23	K→E	+	-
321	67	A→T	23	K→Stop	-	N.D.
343	67	A→G	23	K→E	+	-
384	67	A→T	23	K→Stop	+	N.D.
46	74	A→T	25	E→V	+	-
172	74	A→G	25	E→G	+	-
184	75	G→T	25	E→D	+	-
300	74	A→G	25	E→G	+	-
301	74	A→G	25	E→G	+	-
302	74	A→G	25	E→G	+	-

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
307	75	A→G	25	E→G	+	-
308	75	A→G	25	E→G	+	-
309	75	A→G	25	E→G	+	-
317	74	A→G	25	E→G	+	-
322	74	A→G	25	E→G	+	-
323	74	A→G	25	E→G	+	-
324	74	A→G	25	E→G	+	-
327	74	A→G	25	E→G	+	-
192	77	A→G	26	D→G	+	-
157	81	C→G	27	Y→Stop	-	N.D.
200	87	T→A	29	C→Stop	-	N.D.
268	86	G→A	29	C→R	-	-
294	87	T→A	29	C→Stop	-	N.D.
77	89	G→A	30	R→Q	+	+
265	89	G→T	30	R→L	+	+
344	88	C→T	30	R→Stop	-	N.D.
365	88	C→T	30	R→Stop	-	N.D.

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
166	91	A→T	31	R→Stop	-	N.D.
253	91	A→T	31	R→Stop	-	N.D.
267	92	G→T	31	R→I	+	+
10	99	A→T	33	Q→H	-	-
57	99	A→T	33	Q→H	-	-
196	97	A→T	33	Q→Stop	-	N.D.
361	97	C→T	33	Q→Stop	-	N.D.
243	100	A→T	34	R→Stop	-	N.D.
348	100	A→T	34	R→Stop	-	N.D.
213	107	C→G	36	S→Stop	-	N.D.
358	107	C→A	36	S→Stop	-	N.D.
3	115	T→C	39	Y→H	-	-
252	117	T→A	39	Y→Stop	-	N.D.
356	117	T→A	39	Y→Stop	-	N.D.
375	119	T→A	40	V→E	-	-
276	122	T→A	41	L→Stop	-	N.D.
285	122	T→A	41	L→Stop	-	N.D.

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
290	122	T→A	41	L→Stop	-	N.D.
211	127	T→C	43	F→I	-	+
52	130	C→G	44	L→V	-	+
234	131	T→C	44	L→P	+	+
258	131	T→C	44	L→P	+	+
31	133	G→C	45	A→P	+	+
7	137	T→A	46	I→N	-	-
34	136	A→T	46	I→F	+	+
227	137	T→A	46	I→N	-	-
133	140	T→A	47	F→Y	-	+
203	140	T→A	47	F→Y	-	+
353	140	T→A	47	F→Y	-	+
221	143	T→C	48	L→P	-	+
95	146	C→T	49	S→L	-	+
155	146	C→A	49	S→Stop	-	N.D.
160	146	C→T	49	S→L	-	+
224	145	T→A	49	S→T	-	+

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
320	146	C→A	49	S→Stop	-	N.D.
2	148	A→G	50	K→E	-	+
106	150	A→T	50	K→N	-	+
118	150	A→T	50	K→N	-	+
187	148	A→T	50	K→Stop	-	N.D.
206	149	A→T	50	K→I	-	+
229	150	A→T	50	K→N	-	+
259	148	A→C	50	K→Q	-	-
293	149	A→T	50	K→I	-	+
318	148	A→G	50	K→E	-	+
342	150	A→T	50	K→N	-	+
378	148	A→T	50	K→Stop	-	N.D.
383	150	A→T	50	K→N	-	+
396	148	A→T	50	K→Stop	-	N.D.
41	152	T→C	51	F→S	-	+
83	152	T→C	51	F→S	-	+
314	155	C→A	52	T→N	-	-

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
19	158	A→G	53	N→S	-	+
114	157	A→G	53	N→D	-	+
161	157	A→C	53	N→H	-	+
165	158	A→G	53	N→S	-	+
250	158	A→T	53	N→I	-	-
334	159	T→A	53	N→Q	-	-
355	159	T→A	53	N→K	-	-
370	159	T→A	53	N→Q	-	-
107	164	T→A	55	L→Stop	-	N.D.
306	164	T→A	55	L→Stop	-	N.D.
22	167	T→A	56	L→H	-	+
55	167	T→A	56	L→H	-	+
242	167	T→A	56	L→H	-	+
277	167	T→C	56	L→P	-	-
296	167	T→C	56	L→P	-	-
385	167	T→A	56	L→H	-	+
350	170	T→C	57	L→P	-	-

Table 3.3 Continued

Mutant number[§]	Position of the mutation in *L*	Base pair(s) changed	Amino acid(s) position	Amino acid(s) changed	Lysis*	Protein levels[#]
108	179	T→C	60	L→P	-	-
247	179	T→A	60	L→Q	-	-
304	179	T→C	60	L→P	-	-
372	179	T→A	60	L→Q	-	-
216	188	T→A	63	V→E	-	+
371	188	T→A	63	V→E	-	+
103	197	C→A	66	T→K	-	+
215	197	C→G	66	T→R	-	-
171	205	A→T	69	T→S	-	-
138	211	C→T	71	Q→Stop	-	N.D.
104	214	C→T	72	Q→Stop	-	N.D.
84	218	T→A	73	L→Stop	-	N.D.
191	218	T→A	73	L→Stop	-	N.D.
202	218	T→A	73	L→Stop	-	N.D.

§ The number refers to the mutant clone isolate from the selection plates.

* "+", indicates lysis in liquid culture, "-", indicates no lysis in liquid culture.

"+", indicates expression levels of the *L* allele comparable to WT, "-", indicates less than WT expression levels.

N.D. Not determined.

Near-saturating mutational analysis for L lytic function

To interrogate the new L domain structure for functional significance, a mutational analysis involving selection for non-lethal *L* alleles was implemented. The general approach was to create a mutagenized library of PCR-mutagenized *L* alleles in a plasmid in which the lysis gene was cloned under a tightly regulated arabinose-inducible pBAD promoter (149). The transformants carrying this library were plated for surviving cells under inducing conditions. The plasmid DNAs from 396 survivors were extracted and sequenced. Several false positives also survived on the inducer plates, probably due to the lower concentration of arabinose (0.2% w/v) in the inducer plates. Among the false positives 19 were wild-type *L*, 3 were silent changes, and 34 mutant alleles had no lysis defect in liquid culture. All multiple mutation, frameshift and deletion alleles were discarded, resulting in a total of 102 lysis-defective, single mutation alleles (Table 3.3). A summary of this analysis based on a total of 49 unique single mutation alleles obtained from the lysis-defect selection, as well as a mutant allele obtained by site-directed mutagenesis, is depicted in Figure 3.2. The mutant hunt was close to saturation, as could be judged from the frequency of repeat alleles that were identified late in the procedure. Moreover, of 29 possible nonsense mutations that could be accessed by single base changes in the *L* gene, 23 were obtained, distributed throughout the cistron. Generally, the mutational analysis supports the domain concept proposed above. Since Domain 1 has been previously reported to be non-essential, missense alleles in this region conferring a lysis defect were not expected.

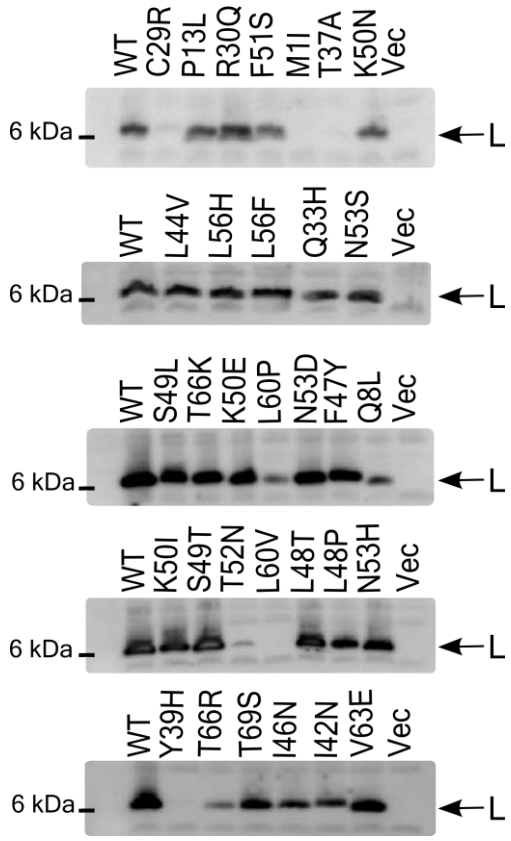

Figure 3.3 Immunoblots of mutant alleles of *L*.
The protein **s**amples from various *L* mutants were collected 5 min prior to the onset of lysis, normalized to A_{550} ~0.4 units, and immunoblotted with anti-L antibody. The wild type (WT), empty vector (Vec), and the mutant alleles are indicated directly above each blot. The molecular mass standard in kDa is represented on the left. The bands corresponding to L are indicated by an arrow on the right.

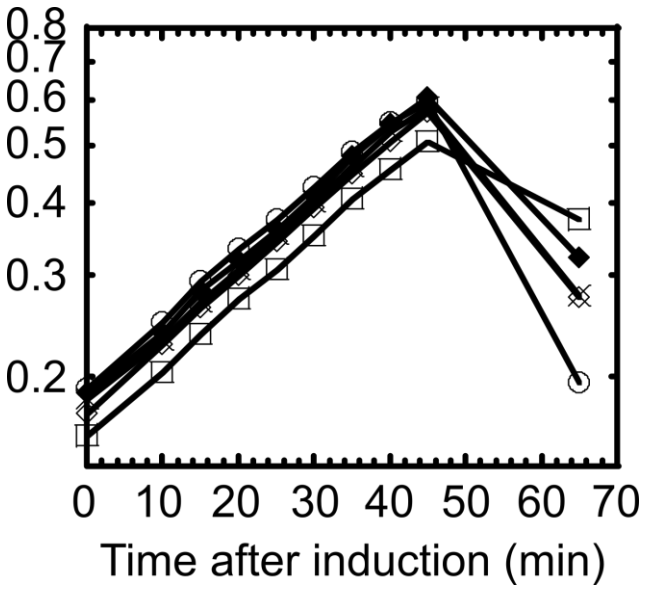

Figure 3.4 *L* mutant alleles are recessive.
The lysis profiles of pBAD24 *L* alleles and pKC17 (*his6-L*) co-transformants in TB28 background. The culture growth was monitored by measuring A_{550}, cultures were induced at time 0, and A_{550} of cultures were at indicated time points. ○, pBAD24 *L* and pKC17; □, pBAD24 L_{K50E} and pKC17; ◊, pBAD24 L_{L44V} and pKC17; X, pBAD24 L_{F47Y} and pKC17; ♦, pBAD24 L_{S49T} and pKC17.

Accordingly, three of the missense alleles we did obtain in Domain 1 were changes to the start codon (M1I, M1T), which ablate L production entirely, and T3I, T3S, P6L, Q8L, and C29R, which exhibited a severe accumulation defect (Table 3.3). Since the C29S allele is fully functional, we suspect that C29R is also lysis-defective due to proteolytic instability, but because this position is within the epitope used for raising the antibody, this could not be confirmed. Only one allele, Q33H, was lysis-defective and exhibited normal accumulation (Table 3.3 and Figure 3.3). The simplest interpretation is that even in the dispensable N-terminal Domain 1, missense changes can occur that block function; e.g., by preventing DnaJ from activating L through interaction with domain 1. In any case, by far the highest incidence of missense changes with lysis-defective phenotypes were obtained in the LS dipeptide motif, and in the regions of Domains 2 and 4 surrounding the LS sequence, supporting the importance of the motif and its context. Moreover, a significant proportion of the inactivating missense changes were conservative, especially in terms of hydrophobic and polar character: e.g., L44V, F47L, F47Y, S49T (similar to S49C), F51L, and L56F). The inactivating effect of such substitutions, coupled with their undiminished levels of protein accumulation (Figure 3.3) suggests that L makes a specific protein-protein contact that depends on these mutationally sensitive residues, rather than acting as a membrane-disrupting peptide as previously suggested (136). Most of the other mutations in Domains 2 and 4 more distal to the key LS motif represent more drastic changes in side change character. In addition, all of the alleles tested were recessive to wild type *L*, even when the mutant allele was expressed from a higher-copy plasmid (Figure 3.4), suggesting that L lytic function does

not require homo-oligomerization but instead requires a heterotypic interaction with a host protein. Finally, all of the defective alleles tested showed unperturbed association with the membrane fraction (Figure 3.5). Taken together, the results of the mutational analysis indicate that L has a host membrane protein target and interacts with it through the LS motif and the nearby residues in essential Domains 2 and 4.

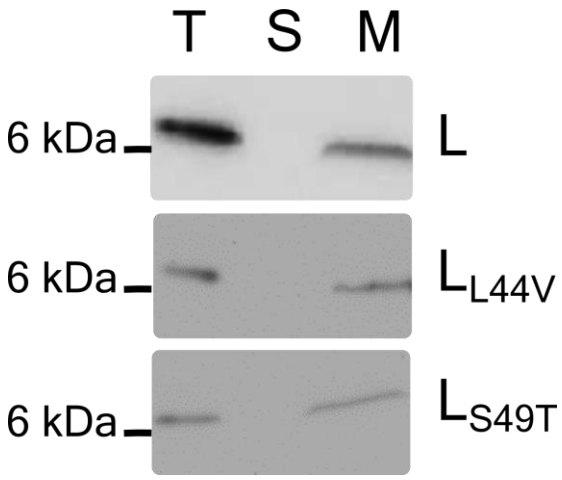

Figure 3.5 Both wildtype and mutant alleles of L are membrane-associated.
The cell lysates from wildtype and mutant alleles (L44V and S49T) were fractionated into soluble and membrane fractions by high speed ultracentrifugation. The normalized fractions were analyzed by immunoblotting with anti-L antibody. The molecular mass standard in kDa is represented on the left. Total (T), soluble (S), and membrane (M) fractions are indicated on top.

Discussion

New perspective on L mode of action

Despite its early identification as the lysis protein in one of the most intensively studied phages of all time, MS2 L has long eluded operational or mechanistic characterization at the molecular level. Until recently, the extant literature on L focused on a consensus in which its essential C-terminal domain acted as a membrane-disrupting polypeptide that subverted the energization of the bilayer and somehow induced an autolytic response involving unknown host proteins (136). We have interrogated L function by genetic analysis based on analysis of survivors of L expression from a plasmid vector. First, this approach revealed that the host DnaJ chaperone was required for L function, interacting with the N-terminal, highly basic domain of L. Both this domain and thus the DnaJ interaction were shown to be dispensable, and a model was proposed in which the N-terminus acted as a regulatory domain that inhibited the interaction of L with a target protein. In this scenario, then, DnaJ acts not as the target but instead binds the regulatory domain and displaces it from its inhibitory position (Chapter II). Here, in the next phase of our genetic approach, we have probed the full length of the L protein by mutational analysis, revealing a distinct domain structure with differential mutational sensitivity and highlighting a dipeptide motif, Leu48-Ser49 that is conserved in heterologous amurins. In addition, many of the missense changes found to abolish lytic function without affecting accumulation or membrane-localization of L are very conservative and indicative of a protein-protein interface. Taken together, the data support a model in which the LS motif and its neighboring subdomains form the core of

an essential heterotypic protein-protein interaction domain. Moreover, the fact that all the inactivating mutations are recessive suggests that L does not act as an oligomeric lytic factor, as would be the case for models in which L is proposed to form membrane-permeating lesions.

Evolution of L-like amurins

The study of ssRNA phages began with the F-pilus specific RNA phages of *E. coli*, and indeed the original classification of these Leviviridae into two genera, the Leviviruses and Allolevivirus. This was largely dependent on the fact that the former had an independent lysis gene L, overlapping the *coat* and *rep* genes, whereas the latter did not, instead effecting lysis with a secondary function of the Mat protein (A_2 in Qβ) (62, 130). Many Leviviridae, either MS2-like and Qβ-like based on the arrangement of genes, were sequenced, and in all the MS2-like phages, the L sequences were related (Figure 3.1). However, more recently the genome sequences of 7 Leviviridae specific for other retractable pili have become available, four (PRR1, Hgal1, M and C-1) for conjugational pili of R-factor plasmids (141, 162, 163), two for polar pili (PP7 for Pseudomonas; phiCb5 for *C. crescentus*) (140) and one for a Type IV pilus (AP205 for *A. baumannii*) (83). Although there is no significant sequence similarity, the arrangements of the genes in PRR1, Hgal1, C-1 and PP7 all resemble that of MS2, with a small ORF overlapping the *coat-rep* gene junction. In each case, the small gene was identified as the lysis gene by testing the induction of a plasmid-based clone. However, in the other three phages, AP205, M, and phiCb5, there was no ORF at the coat-rep junction (83, 140, 141). In AP205 and M, the lysis gene was identified by cloning and

testing for inducible lysis, and in both cases, its position was novel: in AP205, it was a separate gene at the 5' end of the genome, whereas in M, it was a gene that is embedded out of frame at a different position within *rep*. Very recently, the lysis protein of the latter, Lys^M, was shown to be an inhibitor of MurJ, the Lipid II flippase of *E. coli*, and thus, operationally, resembles the A_2 (Mat) protein of Qβ, a specific inhibitor of MurA, the first enzyme in the peptidoglycan biosynthesis pathway; in both cases, lysis is effected by inhibition of cell wall synthesis. The work we have reported here, with the identification of the LS motif and the domain structure of L, clearly groups all the lysis proteins found encoded at the *coat-rep* junction and the AP205 lysis protein with MS2 L. Moreover, the implication is that the L-mediated pathway to lysis, which operates without affecting net cell wall biosynthesis, interacts with a conserved target common to these diverse Gram-negative genera. In addition, since the *L* gene in AP205 is located in a completely difference place in the genome compared to the *L* genes straddling the *coat-rep* boundary, it seems certain that the conservation of the LS motif and the domain structure derives from convergent evolution that occurred after speciation of a progenitor Leviviridae to the respective retractable pili. This may also be true for the L proteins in PRR1, Hgal1, C-1 and PP7, which, although occupying the same coat-rep junction region of their respective genomes, share no sequence homology. In support of this notion, other work has shown that for MS2, the region of *rep* within which the key hydrophobic and LS elements are encoded in the out of frame *L* cistron in MS2 is not conserved, so the emergence of *L*-like genes in this region may simply reflect the limited availability of genomic space for sequence variation.

Comparative analysis of the domain structure of L proteins

Based on our mutational analysis of MS2 L, we have postulated four domains as significant for lytic function of L-like amurins, including the dipeptide sequence LS as a completely conserved motif (Figure 3.2). Assuming the LS sequence has to function from the same molecular context in all of the L-like proteins, then a comparative analysis of the putative domain structured in unrelated L-like amurins may make it possible to infer properties of subcellular localization, membrane topology and potential target proteins. We have chosen as a comparison set six L sequences from F-specific *Leviviridae*, which we designate as the homologous set, and the five L-like proteins from *Leviviridae* specific for other retractable pili, here designated as the heterologous set (Figure 3.1). Confidence for the meaningfulness of this comparison stems from the fact that the L sequences from four of the five heterologous phages (Hgal1, C1, AP205 and PRR1) have been shown to necessary and sufficient for lysis if induced from a plasmid clone in *E. coli*. Preliminary results from this laboratory have demonstrated similar lytic capacity for the cloned *L* gene of PP7 (Unpublished results, Chamakura, K.R., Tran, J.S., and Young, R.F.).

The simplest conclusions from the alignment pertain to the non-essential Domain 1, where there are clearly common characteristics for all the L proteins. In all cases, Domain 1 is marked by a high content of charged residues, biased towards basic residues. Despite a wide variation in size, from 7 residues in AP205 to 36 in MS2 L, Domain 1 in all L proteins has a net predicted charge of +3 to +6. The simplest notion is that, although L is firmly associated with the membrane, Domain 1 is disposed in the

cytosol, consistent with its ability to bind DnaJ. Unstructured basic domains have been identified in other amurins (124, 127) and may serve in initial localization to the surface of the membrane.

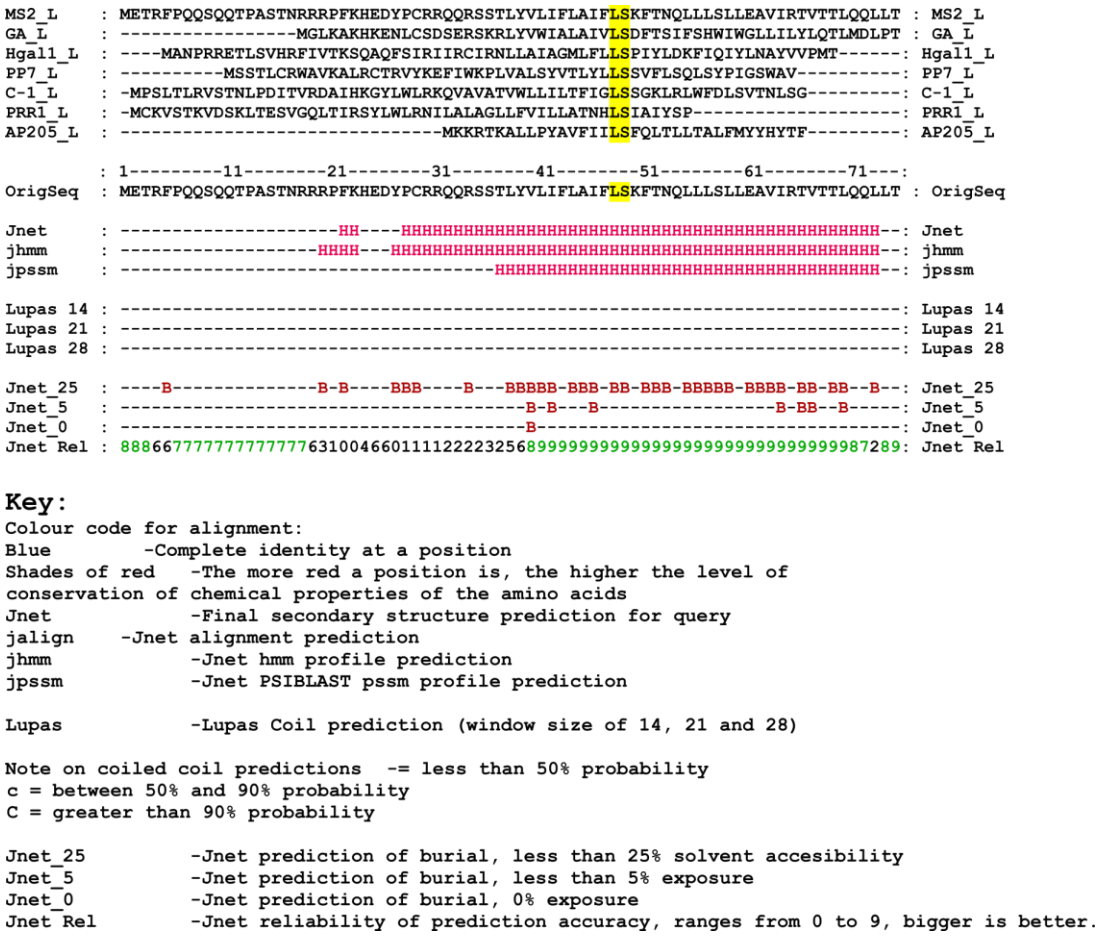

Figure 3.6 Secondary structure analysis of L and heterologous L sequences.
A representative from each of the L sub-groups were chosen and aligned with respect to the LS motif (highlighted in yellow). The aligned sequences were then used as input for secondary structure prediction on Jpred 4 server (164).

The simplest approach to a comparative analysis requires the assumption that Domain 1 will prove to be non-essential for the heterologous L proteins. On this basis, the alignment of L with the five heterologous L sequences was used for seeding secondary structure analysis of L using the Jpred algorithm (164), resulting in a high-confidence prediction for alpha-helical structure from Leu37 to the C-terminus of the molecule (Figure 3.6). This predicted alpha-helical domain coincides exactly with Domains 2-4, the essential domains for L function. Moreover, the TMHMM algorithm, widely-used for TMD prediction, predicts with uniformly high confidence that L and its five functional homologs are integral membrane proteins, which agrees with the behavior of L in subcellular fractionation experiments, even when its N-terminal domain is in complex with DnaJ (Chapter II) In each case the predicted membrane-embedded domain begins at the start of Domain 2 and extends through the LS motif and into Domain 4. In addition, a by-product of the *L* mutagenesis was the isolation of two Pro substitution alleles, L44P and A45P, with normal lysis function (Figure 3.2). Although proline residues are severe helix-breakers in solvent-exposed structures, they are often tolerated at the ends of TMDs (165), suggesting that this domain is embedded in the bilayer. Taken together, these considerations suggest a structural model in which all of the essential residues in L are in a helical structure that spans the bilayer. However, the existence of a TMD is not easily reconciled with the detailed primary structure of the L proteins from the F-specific homologous set. In each of these cases, there is a charged residue (Lys50 in L) adjacent to the LS motif. Unbalanced charged residues are rarely found in TMDs in bacterial integral membrane proteins, except in some transport

proteins where there is solvent exposure deep in the membrane-embedded helices (166).

Thus, although MS2 L, and, by extension, the other L proteins are undoubtedly membrane proteins, canonical requirements for a simple TMD cannot be met. Moreover, even for the heterologous L sequences, where it is not necessary to embed a charged residue to accommodate a TMD, the predicted TMD would also place the LS motif within the bilayer, where specific interactions presumably including the H-bonding capacity of the serine hydroxyl group would be hard to envision unless the membrane is unusually distorted. However, recent findings suggest that membrane distortion might be integral to normal cell wall growth. To account for homeostatic growth of the PG sacculus, models for the recruitment of the PG biosynthetic complexes to sites of negative membrane curvature have been proposed and supported by in vivo and in silico experiments (167, 168). Disruption of such a system by L interacting with a membrane-embedded target protein would account for the ability of L proteins to effect lysis without inhibiting net incorporation of PG into the existing wall, as has been reported for MS2 L (133, 144). Thus interaction of L with its target, perhaps via attraction to distorted membrane, would lead to a lytic catastrophe. To address this general hypothesis for L lysis, experiments are underway to assess the distribution of L molecules relative to the sites of PG biosynthesis.

CHAPTER IV

A VIRAL PROTEIN ANTIBIOTIC INHIBITS LIPID II FLIPPASE

Introduction

For bacteriophage infections, the strong cell walls of bacteria, consisting of a single highly polymeric molecule of peptidoglycan (PG), pose a formidable barrier to the release of progeny virions. Phage lysis proteins evolved to overcome this barrier can point the way to new antibacterial strategies (169). Double-stranded DNA phages of Gram-negative hosts employ up to six lysis proteins including the endolysins that degrade the PG, but small lytic single-strand DNA (the microviruses) and RNA phages (the leviviruses) effect lysis of the bacterial host by expressing a single non-enzymatic protein (4, 12, 170). The two best-studied examples are the A_2 protein of levivirus (ssRNA phage) Qβ and the E protein of the canonical microvirus (ssDNA phage) ϕX174 (119, 130). Because these proteins inhibit specific enzymes in the conserved bacterial cell wall synthesis pathway (Figure 4.1) and cause lysis through a septal catastrophe that is indistinguishable from that caused by antibiotics directed against these enzymes, A_2 and E have been called "protein antibiotics" (14). The leviviruses, ssRNA phages of Gram-negative bacteria, are particularly attractive systems for looking for more of these molecules because of their simple three-gene core structure (*mat*, *coat*, *rep*) and extremely high mutation rate, leading to a great diversity (Figure 4.1).

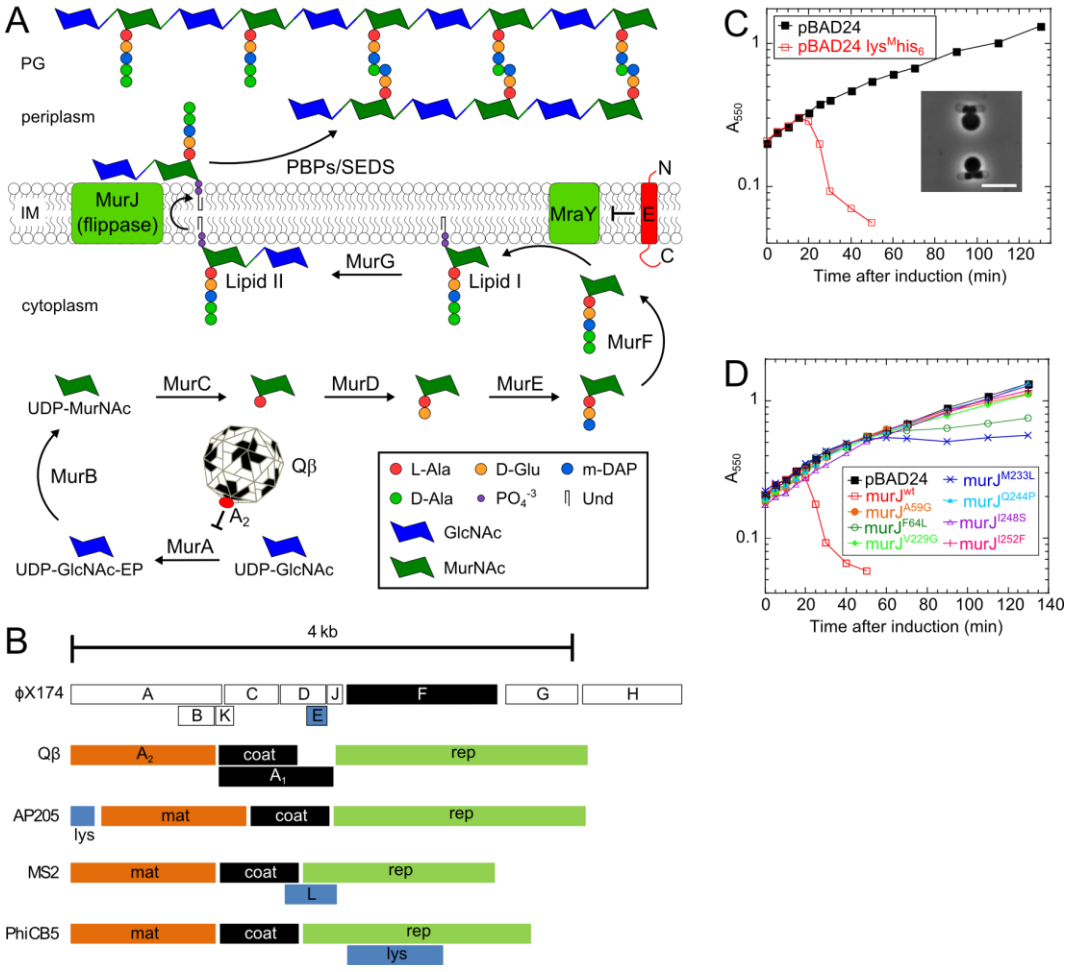

Figure 4.1 The lysis protein of phage M targets Lipid II flippase.
(A) The PG precursor pathway, from cytoplasmic UDP-GlcNAc to periplasmic Lipid II, with known targets of single gene lysis systems indicated. (B) Genome organization in small ssDNA (microvirus ϕX174) and ssRNA (leviviruses Qβ, MS2, AP205, PhiCb5 and M) phages. In the ssRNA phages, the order of essential genes is always *mat coat*, and *rep*. *mat* encodes the maturation or attachment protein, which is present in single copy on each virion and is responsible for adsorption to the receptor pilus.

Figure 4. 1 Continued. The *coat* encodes the capsid protein, present in 178 copies and encapsidating the single genomic RNA, and *rep* encodes the viral subunit of the RNA-dependent RNA polymerase. In Qβ, the *mat* gene is named A_2 and has the additional function of host lysis. Lysis genes have evolved in a variety of sites, including within essential genes in a different reading frame. (C) Cells with either pBAD24 or pKC3 (*lys-his*$_6$ clone) were induced with 0.4% (w/v) arabinose and growth was monitored by turbidity. Inset: phase contrast image of cells at the time of lysis (20 min post-induction) (scale bar = 5 μm). (D) Same as **(C),** except in hosts with insensitive to M lysis (Iml) alleles.

Here we show that the lysis protein of levivirus M of *E. coli* is a specific inhibitor of MurJ, which has recently been shown to be the essential flippase that presents the final cytoplasmic, lipid-linked PG precursor to the extracellular cell wall biosynthesis machinery (114). Since Lys^M is encoded in an alternate reading frame of the essential *rep* replicase gene (141), this finding indicates that the ssRNA phages have previously unappreciated capacity for evolving "protein antibiotics" specific for diverse steps in the cell envelope biosynthesis and maintenance pathway.

Table 4.1 Strains and plasmids

Strain	Relevant Genotype	Reference
XL1-Blue	*recA endA1 gyrA96 thi hsdR17 supE44 relA1 lac* [F'::Tn10 *proA+B+lacIq Δ(lacZ)M15]*	Stratagene
MC4100	e14- *araD139* Δ*(argF-lac)U169 rspL*150(strR) *relA*1 *flhD*5301 Δ(*fruK-yeiR*)725(*fruA25*) *rbsR22 deoC1* Δ(*fimB-fimE*)632(::*IS*1)	(171)
MG1655	*ilvG- rfb*-50 *rph*-1	(146)
TB28	MG1655 *lacIZYA <>frt*	(147)
NR754	*araD*$^{+}$ revertant of MC4100	(172)
T4112	TB28 *lysA763 <>frt*	This study
TU276	TB28 *lysA763 <>frt*	This study
RY34350	T4112 Δ*pyrC::kan*	This study
RY34351	RY34350 *pyrC+murJ*^{F64L}	This study
RY34352	RY34350 *pyrC+murJ*^{M233L}	This study
RY34363	RY34350 *pyrC+murJ*^{I248S}	This study
RY34366	RY34350 *pyrC+murJ*^{A59G}	This study
RY34376	RY34350 *pyrC+murJ*^{Q244P}	This study
RY34380	RY34350 *pyrC+murJ*^{V229G}	This study
RY34381	RY34350 *pyrC+murJ*^{I252F}	This study

Table 4.1 Continued

Strain	Relevant Genotype	Reference
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2184	(pFLAGMurJΔCys/T14C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2185	(pFLAGMurJΔCys/A26C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2202	(pFLAGMurJΔCys/L3C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2208	(pFLAGMurJΔCys/E70C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2214	(pFLAGMurJΔCys/A37C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2225	(pFLAGMurJΔCys/V215C)	(112)
	NR754 Δ*murJ* Δ*hsdR::ka*n	
NR2285	(pFLAGMurJΔCys/V281C)	(112)
NR2449	NR754 (pFLAGMurJΔCys)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2481	(pFLAGMurJΔCys/A268C)	(112)
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2503	(pFLAGMurJΔCys/L47C)	(112)

Table 4.1 Continued

Strain	Relevant Genotype	Reference
	NR754 Δ*murJ* Δ*hsdR::kan*	
NR2504	(pFLAGMurJΔCys/I54C)	(112)
	NR754 ΔmurJ::FRT	
NR3925	(pFLAGMurJΔCys/F64C)	This study
	NR754 Δ*murJ*::FRT	
NR3926	(pFLAGMurJΔCys/F64L)	This study
	NR754 Δ*murJ*::FRT	
NR3927	(pFLAGMurJΔCys/V229G)	This study
	NR754 Δ*murJ*::FRT	
NR3928	(pFLAGMurJΔCys/M233L)	This study
	NR754 Δ*murJ*::FRT	
NR3929	(pFLAGMurJΔCys/I248S)	This study
	NR754 Δ*murJ*::FRT	
NR3930	(pFLAGMurJΔCys/I252F)	This study
	NR754 Δ*murJ*::FRT	
NR3931	(pFLAGMurJΔCys/Q244P)	This study
	NR754 Δ*murJ*::FRT	
NR3932	(pFLAGMurJΔCys/M233L/I248S)	This study
NR3942	NR754 (pFLAGMurJΔCys/V229G/I248S)	This study

Table 4.1 Continued

Strain	Relevant Genotype	Reference
NR3945	NR754 Δ*murJ*::FRT Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/L3C)	This study
NR3946	NR754 Δ*murJ*::FRT Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/A37C)	This study
NR3947	NR754 Δ*murJ*::FRT Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/V281C)	This study
NR3948	NR754 Δ*murJ*::FRT Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/E70C)	This study
NR3949	NR754 Δ*murJ*::FRT Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/V215C)	This study
NR3957	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/T14C)	This study
NR3958	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/A26C)	This study
NR3959	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/L47C)	This study
NR3960	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/I54C)	This study
NR3961	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/S259C)	This study

Table 4.1 Continued

Strain	Relevant Genotype	Reference
NR3962	NR754 Δ*murJ* Δ*hsdR*::kan (pKC6; pFLAGMurJΔCys/A268C)	This study

Materials and methods

Bacterial strains, plasmids and growth conditions

Primers are listed in Table 4.1. Cultures were grown with aeration at 37°C in lysogeny broth (LB) supplemented with ampicillin (100 μg/ml), chloramphenicol (10 or 25 μg/ml), L-arabinose (0.4% w/v), when indicated isopropyl-β-D-thiogalactopyranoside (IPTG; RPI) at 25 μM or 50 μM final concentration were added, and 20 μg/ml final 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-Gal) when indicated. All chemicals were supplied by Sigma Aldrich unless otherwise indicated. The Δ*pyrC::kanR* allele was obtained from Keio collection (148) and was transduced into appropriate strains by P1 transduction (150).

Table 4.2 Primers

Primer	
KC30	cggcagaaaagtccacattg
KC31	cacactaccatcggcgcta
KC163	cctttggtggataccttcaccaagggcagcagccatcaccatcatc accactgataactcgagaagcttggctgttttggcgg
KC210	cctcctctcgagttatcagtggtg
KC230	gaataaaccctgtaacatctggcgg
KC234	accgcatcaggaaccgacgctatatg
KC238	ggggcattttcccaggcacttgtaccgattctggcggaat ataaaag
KC247	ggggcattttcccaggcatgtgtaccgattctggcggaat ataaaag
KC248	gcaatgcgcgtggtgaaacagttgggaccggcgatccttg
KC249	gccggagcaatgcgcggggtgaaacagatgggaccggc
KC252	ccttggcgtctctgtgagcccgatctccttaatcatcaacac cattttgc
KC253	cgtctctgtgagccagatctccttaagcatcaacaccattttt gcctcgtttc
KC254	gccagatctccttaatcatcaacacctttttgcctcgtttct tgcttccg

Table 4.2 Continued

Primer	Sequence 5'-3'
KC263	gccgtatctttgccgaaggggattttcccaggcatttgtacc gattctg
KC264	gccgtatctttgccgaaggggattttcccaggcattggtacc gattctg

Plasmid pKC3 (*bla araC* P_{ara}*::lys-his*$_6$ *lacZα*) encoding a hexahistidine-tagged Lys^M was constructed in multiple steps. First a synthetic construct was obtained (Genscript) containing the DNA sequence from phage M (Accession number: JX625144) encompassing the *lys* gene and *lacZα*, flanked by restriction sites for EcoRI and XhoI/HindIII upstream and downstream, respectively, to facilitate cloning, and with its start codon changed from TTG to ATG. This *lys* construct was sub-cloned into pBAD24 between the EcoRI and HindIII sites to yield pKC1. Codons encoding a hexa-histidine tag were inserted at the 3' end of *lys* in pKC1 through site-directed mutagenesis with the primer KC163, to generate pKC2. The resulting *lys-his6* gene was amplified from pKC2 by PCR with primers KC30 and KC210 and Pfu DNA polymerase. The PCR product was digested with EcoRI and XhoI and cloned between the corresponding sites in pKC1 to yield pKC3. The cloned fragments at each step were verified by sequencing.

A *lys* gene modified at its 3' end with codons for the c-myc tag (GEQKLISEEDLNSAVD) was obtained as part of a gBlock® (IDT) containing flanking KpnI and HindIII sites. Plasmid pKC4 (*bla araC* p_{ara}*::lys-cmyc)* was constructed by digesting the gBlock with KpnI and HindIII and ligating the product into the corresponding sites in the vector pBAD30.

Plasmid pKC5 (*cat araC* p_{ara}*::lys-cmyc)* was constructed by sub-cloning *lys-cmyc* from pKC4 to pBAD33 with KpnI and HindIII.

Plasmid pKC6 (*bla araC* p_{ara}*::lysC10S-cmyc)* was constructed similar to pKC4 where codon for cysteine at position 10 was changed to serine.

Mutant alleles of *murJ* were constructed by site-directed mutagenesis into the plasmid pFLAGMurJΔCys (112) using Phusion® high-fidelity DNA polymerase (New England Biolabs) and primers KC238, KC247, KC248, KC249, KC252, KC253, KC254, KC263, and KC264.

Strains RY34351, RY34352, RY34363, RY34366, RY34376, RY34380, and RY34381 carrying Iml mutations were constructed by P1 transduction with Iml mutants as donors and RY34350 (*lysA*:: *frt*, Δ*pyrC*::*kan*) and plated on M9 minimal medium supplemented with 0.2% (w/v) glucose and lysine (5 μg/ml). The *murJ* loci from transductants were amplified by PCR with primers KC230 and KC234 and the mutations confirmed by sequencing.

Multi-copy suppression

The *E. coli* multi-copy (MC) library containing 1- 4 kb fragments of chromosomal DNA under p_{ara} control used in this study was constructed for an earlier

work (173). Approximately 200 ng of pooled MC library and ~100 ng of pKC3 plasmid were co-transformed into XL1-Blue by electroporation. The transformants were plated on LB-Ara-Amp-IPTG-X-gal plates to select and screen for survivors. The blue colonies that appeared on the selection plates were picked and the plasmid DNA was extracted using a Qiagen® mini-prep kit. The extracted DNA was digested with HindIII and analyzed on 0.8% agarose for 1 h at 95 V to separate pKC3 from MC library plasmid. The slower migrating band was gel-purified using a Qiagen® gel-extraction kit, self-ligated, transformed into XL1-Blue, and selected on LB-chloramphenicol (10 μg/ml) plates. The plasmid DNA was extracted with Qiagen® mini-prep kit and the junctions were sequenced at Eton Bioscience Inc. with primers KC30 and KC31. The sequence of the inserts was mapped to the *E. coli* MG1655 genome (NC_000913.3) to identify the possible suppressors. To rule out the suppressors that interfere with arabinose uptake and utilization, the isolates were streaked on M9 minimal agar with arabinose (0.2% w/v) as the sole carbon source.

Selection and screening for insensitive to M lysis (Iml) mutants

Cultures of XL1-Blue pKC3 were grown to A_{550} ~ 0.2 and *lys* was induced with arabinose (0.4% w/v). At ~2 hours post-induction, cells were collected from 10 ml of culture through centrifugation at 6,000 x g for 10 min, resuspended in 100 μl LB, and plated on LB-Ara-Amp-IPTG-X-gal plates. The colonies that turned blue after overnight incubation at 37°C were picked and streak-purified on the same selection media. The Iml mutants thus isolated were further characterized in liquid cultures by following their growth post-induction of *lys* from pKC3. The expression levels of Lys^{M}-his_6 in Iml

mutants were compared with that in wild-type by Western blotting with anti-His (Sigma-Aldrich). The *murJ* locus in Iml mutants was amplified by PCR using Phusion® high-fidelity DNA polymerase (New England Biolabs) with primers KC230 and KC234. The amplified PCR product was gel-purified and sequenced with primers KC230, KC231, KC234, and KC262.

Microscopy

Strain TB28 pKC3 was grown to A_{550}~0.2 and induced with L-arabinose 0.4% (w/v). At 20 min post-induction, 1 ml of culture was spun down at 14,000 x g for 30 sec. The pellet was resuspended in 20 μl, of which 5 μl was spotted on a glass slide with cover slip on top. Phase contrast images of the cells were taken with a Zeiss AX10 microscope with 100X oil immersion objective and 100 ms exposure.

Plasmid retention assay

TB28 cells carrying plasmids with different mutant alleles of *murJ* were grown to A_{550} ~ 0.2 and induced with 0.4% arabinose. The cell lysates 1 h post-induction were harvested, filter sterilized, and passed through Qiagen® DNA binding columns. The bound DNA was eluted with water and transformed into TB28 and the number of transformants were counted. The *murJ* alleles that gave least number of transfromants were scored as tight mutants.

Detecting Lipid II flippase activity with ColM.

Lipid II flippase assays were done with slight modifications from previously described[8]. Strain RY34351/pKC4, TU276/pFLAGMurJ/pKC4, and TU276/pFLAGMurJ$^{M233L,\ I248S}$/pKC4 were grown in LB supplemented with 0.2% (w/v)

glucose, 25 μg/ml of chloramphenicol and 15 μg/ml ampicillin at 37°C with shaking overnight. Cultures were then diluted 1:100 into M9 medium supplemented with 0.2% maltose and 100 μg of methionine, lysine and threonine and incubated at 37°C with shaking. When A_{600}~ 0.3, 15μl of [^{3}H]-mDAP (1μCi/μl, ARC) was added to 10 ml of culture and arabinose was added to a final concentration of 0.4% (w/v) as indicated. After 15 minutes, 10 μl of colicin M (500 μg/ml) was then added to the culture. Cells were immediately chilled on ice for 10 min and harvested by centrifugation at 10,000 x *g* for 2 minutes at 4°C. Spent supernatant was discarded carefully, and cell pellets were resuspended in 1 ml of hot water and further incubated at 100°C for 30 minutes. Heat-killed cells and hot water extracts were separated by centrifugation at 100,000 x *g* for 20 minutes at 4°C. The supernatant fractions were lyophilized, dissolved in 300μl of buffer A (50mM triethylammonium formate pH4.6, 6% (v/v) methanol), followed by HPLC with 70 minutes isocratic elution of buffer A on a Nucleosil C18 column (Agilent A0119250X046). Radiolabeled soluble PG precursors and ColM products in the eluate were detected by inline radioflow detector (Berthold). The hot water extracted cell pellets were boiled for 3 minutes in 1ml of water to remove residual water soluble radiolabeled compounds. The washed pellets were then resuspended in 100μl of 10mM Tris-HCl pH 7.4. Lipid-linked precursors were extracted from the pellet suspension twice by adding 100μl of 6M pyridinium acetate/1-butanol (1:2, v/v) followed by vortexing. The top butanol fractions were carefully transferred to a new tube after centrifugation at 3,000 x *g* for 30 seconds and washed once with 100μl butanol saturated water. Labeled lipid-linked precursors were then quantitated from 100μl of top butanol

fractions by scintillation counting in 10 ml of Ecolite scintillation fluid (882745, MP biochemicals).

Complementation test of pFLAGMurJΔCys derivatives

The ability of pFLAGMurJΔCys-derived plasmids to complement a chromosomal Δ*murJ*::*frt* allele was tested as previously described (112) using strain NR3267 (NR754 Δ*murJ*::*frt* pRC7KanMurJ)(174). Haploid strains carrying complementing pFLAGMurJΔCys-derived plasmids and merodiplod strains NR3942 (NR754 pFLAGMurJΔCys$^{V229G,\ I248S}$) and its wild-type derivative NR2449 (NR754 pFLAGMurJΔCys$^{V229G,\ I248S}$) were processed for western blotting using anti-FLAG M2 (Sigma-Aldrich) and anti-mouse-HRP (GE Amersham). For immunoblots, cells were grown overnight, normalized by OD_{600}, pelleted, and lysed with 50 µl of SCAM lysis buffer (5 mM Tris-HCl pH 7.4, 1 % SDS, 6 M urea). Prior to loading onto a 10 % SDS-polyacrylamide gel, samples were mixed with 50 µl of 2X AB buffer (6.84 mM Na_2HPO_4, 3.16 mM NaH_2PO_4, 50 mM Tris-HCl pH 6.8, 6 M urea, 1% β-mercaptoethanol, 3% SDS, 10% glycerol, 0.1% bromophenol blue).

SCAM

murJ haploid strains (Table 4.1) carrying Δ*murJ*::*frt* pFLAGMurJΔCys-derivatives with Cys substitutions at various positions and pKC6 were grown to A_{600} ~1.0. Production of Lys$^{M(C10S)}$-myc was induced by the addition of 0.2 % L-arabinose (w/v). After 10 min at 37ºC, cells were pelleted and processed for SCAM as previously described (112).

In silico MurJ structural modelling

Renderings of MurJ structural model (112) were generated as previously described (114).

Table 4.3 Multi-copy suppressor plasmids

Multicopy suppressor	5' junction start site	3' junction end site	Full length genes encompassed within the cloned fragment
1	1126096	1129782	*yceH-yceM-murJ*
2	1127429	1129782	*murJ*
3	112749	1129732	*murJ*
4	1616946	1618333	*ydeA*

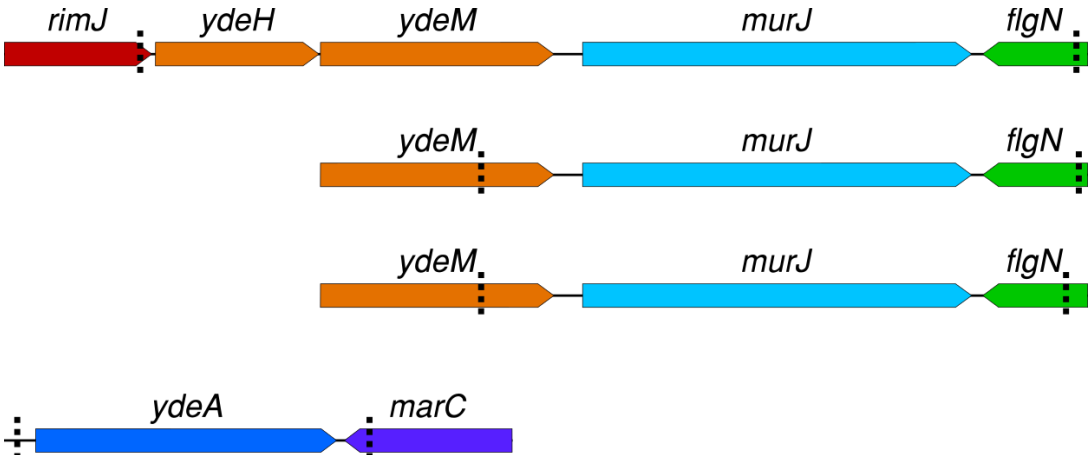

Figure 4.2 Lys^{M} lethality is suppressed by multi-copy plasmids carrying the Lipid II flippase gene *murJ*.
DNA segments that were present on the multi-copy plasmids that suppressed Lys^{M} lethality are demarcated by vertical dotted lines. *rimJ* = ribosomal-protein-S5-alanine N-acetyltransferase; *ydeH* = diguanylate cyclase; *ydeM* = predicted anaerobic sulfatase maturation enzyme; *murJ* = lipid II flippase; *flgN* = flagellar biosynthesis protein; *ydeA* = arabinose exporter; *marC* = inner membrane protein of unknown function.

Results

Since many leviviruses infect via conjugation pili, they usually can propagate in, and thus effect lysis of, diverse bacterial species. Recently, Rumnieks and Tars (2012) surveyed the few available ssRNA phage genomes and found that the lysis genes appear to have evolved at different sites in the core leviviral genome after speciation of the phage to new pili receptors. In particular, they found that levivirus M, which infects via the conjugation pilus of IncM antibiotic-resistance plasmids, has its *lys* gene embedded in the +1 reading frame of *rep* (141). We observed that lysis caused by expression of *lys* cloned in a multicopy plasmid proceeded mostly through septal catastrophes, like A_2 and E (Figure 4.1), suggesting that the product of *lys*, Lys^M, might be an inhibitor of cell wall biosynthesis. Next, we used a multicopy library of *E. coli* genome fragments to look for clones that conferred protection against the lethal expression of *lys* (Figure 4.2, Table 4.3). Four plasmids were identified, three of which were shown to carry *murJ*, which has been recently shown to encode the long-sought "flippase" that translocate Lipid II across the inner membrane in cell wall synthesis (110, 111, 114) (Figure 4.1). A fourth plasmid carried *ydeA*, which encodes a pump that exports cytoplasmic arabinose, the inducer for this vector construct (175).

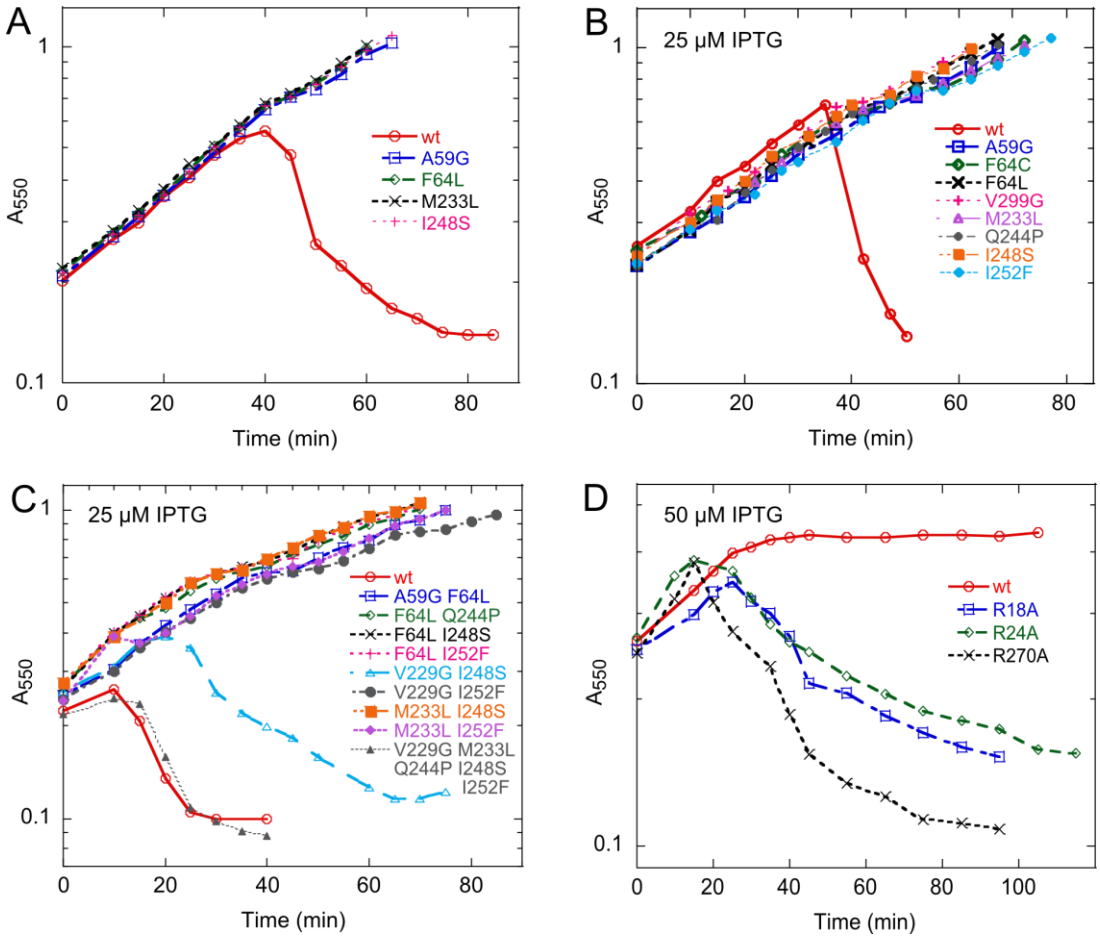

Figure 4.3 Only functional *murJ* alleles rescue cells from Lys^{M} lethality.
TB28 pKC4 (*lys-cmyc* clone) cells and various pFLAG *murJ* alleles were induced with IPTG (25 μM or 50 μM) at the beginning of culture growth and 0.4% (w/v) arabinose was added at time 0. (A) Chromosomal and (B) plasmid expression of *murJ* Iml alleles rescues cells from lysis. (C**)** Plasmid expression of *murJ* Iml double mutants rescues cells from lysis except in the case of V229G I248S, which partially rescues cells. Note that an allele in which 5 Iml mutations spanning the entire length of TMD7, fails to rescue cells from lysis; this allele is non-functional (not shown). (D) Over-expression of wildtype *murJ*, but not various nonfunctional alleles, rescues cells from lysis.

To isolate MurJ variants that are resistant to Lys^{M}, spontaneous chromosomal "Iml" mutants (conferring *insensitivity to M lysis*) were selected using the same *lys* vector and mapped to *murJ* by P1 transduction. Each of these *murJ* mutants survived induction of *lys* expression with no defect in logarithmic growth, whereas the parental host underwent complete lysis at ~20 min after induction (Figure 4.1, Figure 4.3).

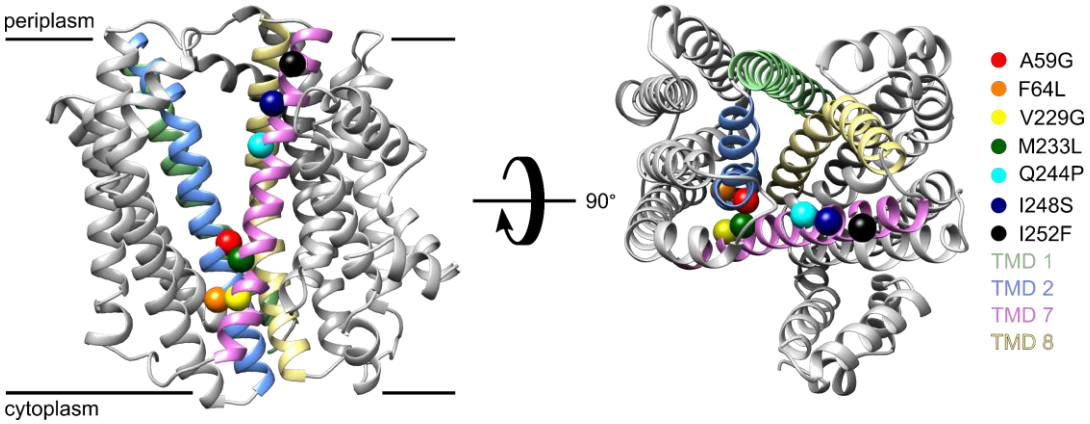

Figure 4.4 The Iml mutations mapped to TMD2 and TMD7 of MurJ.
Structural model of MurJ generated by I-TASSER (112). The TMDs that line the central hydrophilic cavity are colored. The seven Iml mutants identified map to TMDs 2 (light blue) and 7 (magenta). Lateral view (left) and periplasmic view (right).

Table 4.4 Iml alleles

murJ Iml mutant	Nucleotide change	Type of change	Codon change	Amino acid change
1	T→G	Transversion	GTG→GGG	V229G
2	A→T	Transversion	ATG→TTG	M233L
3	T→C	Transition	TTT→CTT	F64L
4	C→G	Transversion	GCA→GGA	A59G
5	T→G	Transversion	ATC→AGC	I248S
6	A→T	Transversion	ATT→TTT	I252F
7	A→C	Transversion	CAG→CCG	Q244P
8	T→G	Transversion	ATC→AGC	I248S
9	T→G	Transversion	ATC→AGC	I248S

Sequencing of the *murJ* locus in these 9 mutants revealed seven different single missense changes, clustered in two of the fourteen transmembrane domains (TMDs) of MurJ, two in TMD2 and 5 in TMD7 (Figure 4.4, Table 4.4). None of these mutations affected accumulation of the MurJ protein (Figure 4.5).

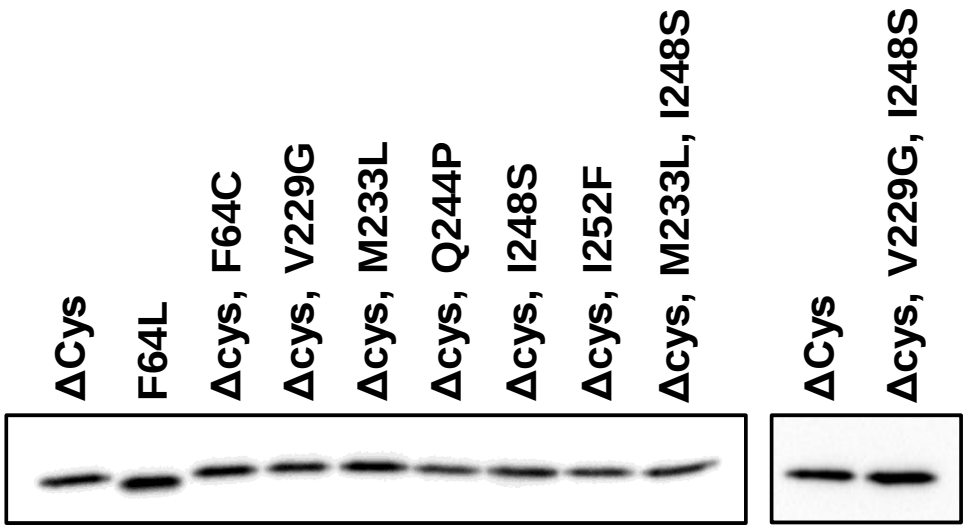

Figure 4.5 Iml MurJ variants are stably produced.
Anti-FLAG immunoblot was used to determine the effect of Iml mutations on MurJ protein levels in whole cells. All Iml MurJ variants were produced from plasmids derived from pFLAGMurJΔCys. Functional, plasmid-encoded Iml *murJ* alleles were expressed in cells where the native chromosomal *murJ* allele was deleted (indicated as Δ*murJ* below the immunoblot). Non-complementing $murJ^{V229G,\ I248S}$ allele was expressed in a strain with a wild-type *murJ* chromosomal allele (indicated as $murJ^{+}$ below the immunoblot). ΔCys refers to $MurJ^{\Delta Cys}$, a functional variant where the two native cysteine residues are substituted by serine residues and which is produced from pFLAGMurJΔCys. FLAG-$MurJ^{F64L}$, as previously described for FLAG-MurJ (112), has higher mobility than $MurJ^{\Delta Cys}$.

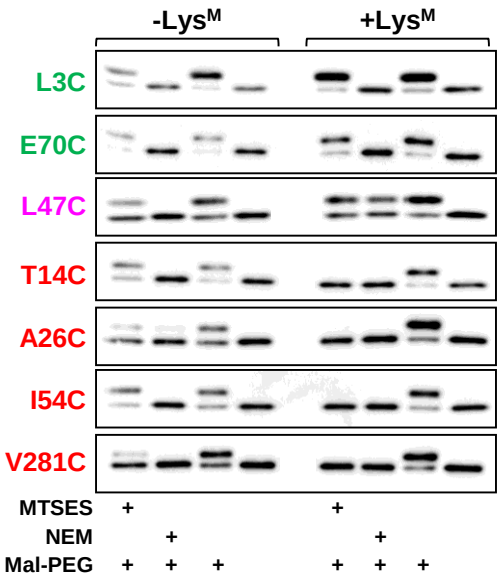

Figure 4.6 Lys^M induces conformational changes in MurJ.
Substituted-cysteine accessibility method (SCAM) was used to determine whether the induction of Lys^M production changes the conformation of MurJ in the IM. Conformational changes were reflected as changes in the exposure of cysteine residues introduced at specific positions in MurJ (indicated on the left) to the aqueous periplasm or cytoplasm. FLAG tag immunoblotting was used to monitor the mass shift in MurJ that the thiol-modifying agent Mal-PEG causes unless cysteines are previously modified by IM-permeable NEM (reacts with cysteines exposed to either the periplasm or the cytoplasm) or IM-impermeable MTSES (only reacts with periplasmic cysteines). Substitutions in green showed no change in SCAM pattern; substitutions in pink showed less modification by MTSES and NEM (and therefore more Mal-PEG modification), indicating less exposure to the aqueous environment; and substitutions in red showed more modification by MTSES and NEM (and therefore less Mal-PEG modification), indicating more exposure to the aqueous environment.

Chemical probing of MurJ in situ using thiol accessibility (SCAM) provided evidence that TMDs 2 and 7, along with TMDs 1 and 8, line a solvent-exposed cavity (112), validating a structural model for MurJ generated by I-TASSER (112, 176). In this model, the Iml mutations either map along the same cavity-exposed face of TMD7 or to TMD2 near the closed, cytoplasmic interface end of the cavity (Figure 4.4). We used SCAM to address whether Lys^M affected MurJ conformation in the membrane. Cys-substitutions in 11 positions in MurJ were tested for reactivity to the thiol reagents MTSES (membrane-impermeable) and NEM (membrane-permeable) in cells in which a plasmid clone encoding Lys^M-cmyc was induced (Figure 4.6). The SCAM patterns of the five positions in aqueous domains (N-terminus, two cytoplasmic loops, two periplasmic loops) were unaffected by *lys* induction, but 5 of the 6 TMD positions exhibited altered MTSES sensitivity, with four of the positions being converted from partial to complete reactivity. These results suggest that Lys^M binds to MurJ and causes a conformational change; the prevalence of the pattern change showing conversion from partial to complete reactivity (positions T14, A26, I54, and V281) suggests that Lys^M binding locks MurJ into one of the two conformations proposed to constitute the Lipid II-flipping cycle, with the solvent-filled cavity facing either the periplasm or cytoplasm (112, 113) (Figure 4.7). Further support for this notion derives from inductions of *lys* in the presence of multicopy plasmids carrying non-functional alleles of *murJ*; despite producing normal levels of protein (113), these alleles do not protect the host from Lys^M-dependent lysis, whereas the wild-type allele does (Figure 4.3).

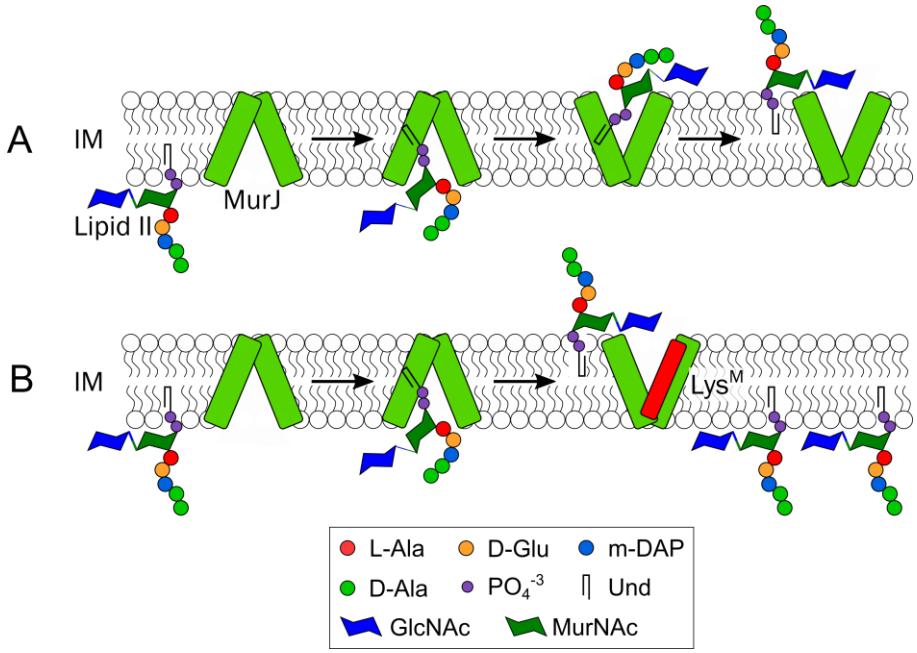

Figure 4.7 Model for inhibition of MurJ by Lys^{M}.
(A) Four TMDs (1, 2, 7, and 8) that line the solvent accessible cavity of MurJ[13] are represented as "V-shaped" rectangle blocks (green) that alternate between cytoplasmic and periplasmic exposed conformations. MurJ is proposed to bind to Lipid II in the inner leaflet of IM, switch conformations and release Lipid II in the outer leaflet of IM. (B) Lys^{M} (red) blocks Lipid II flipping activity of MurJ by interacting with TMDs 2 and 7. This leads to an accumulation of Lipid II in the inner leaflet of IM, blockage of cell wall synthesis, and cell lysis.

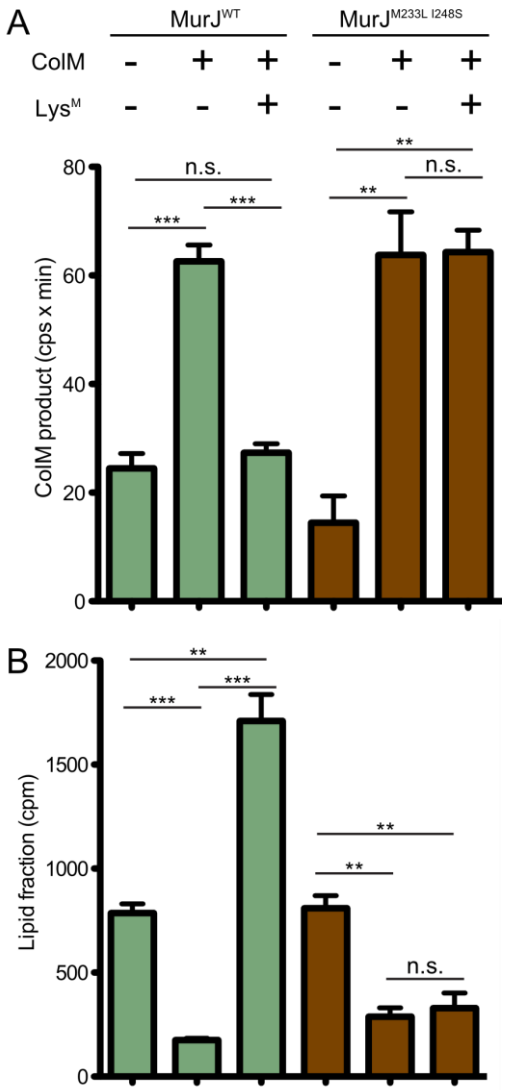

Figure 4.8 Lys^{M} blocks Lipid II flipping.
Cells of Δ*lysA* strains producing FLAG-tagged MurJ (MurJ^{WT}, TU192/pFLAGMurJ/pKC4) or its derivative (MurJ$^{M233L\ I248S}$, TU192/pFLAGMurJ$^{M233L\ I248S}$/pKC4) were labeled with [^{3}H]-mDAP. Arabinose was added at the same time to induce the production of Lys^{M} as indicated. Colicin M was then added 15 minutes after labeling, and growth continued for another 10 minutes.

Figure 4.8 Continued. Samples were then withdrawn and either extracted with hot water, or sequentially with hot water then butanol. The amount of ColM product was measured from hot water extracts by HPLC with an in-line radioflow detector (A). Scintillation counting was used to quantify [^{3}H]-mDAP label in the butanol fractions (B). Shown are the means ± SEM from three experiments. *P* values were determined by Student's *t* test. Chromatograms of the HPLC results were shown in Figure 4.10. cps, counts per second. ***, $p<0.001$; **, $p<0.01$; n.s., not significant.

Previously, Sham et al., used a colicin M-based assay to show that MurJ inhibition blocked Lipid II flipping (114). When added to the medium of growing cells, colicin M invades the periplasm to cleave the undecaprenol carrier from flipped Lipid II, generating a soluble disaccharide pentapeptide product that is subsequently converted to disaccharide tetrapeptide (ColM product) by periplasmic carboxypeptidases. Thus MurJ flippase function can be monitored by the degree to which radiolabeled precursors accumulate in the lipid fraction (unflipped Lipid II) and the water-soluble fraction (flipped, colicin-cleaved, carboxypeptidase-processed disaccharide tetrapeptide), respectively. Using this assay on growing cells carrying the inducible *lys* construct, we showed that expression of *lys* caused a dramatic accumulation of label in the lipid fraction and a corresponding reduction of the soluble ColM product (Figure 4.8 and Figure 4.9).

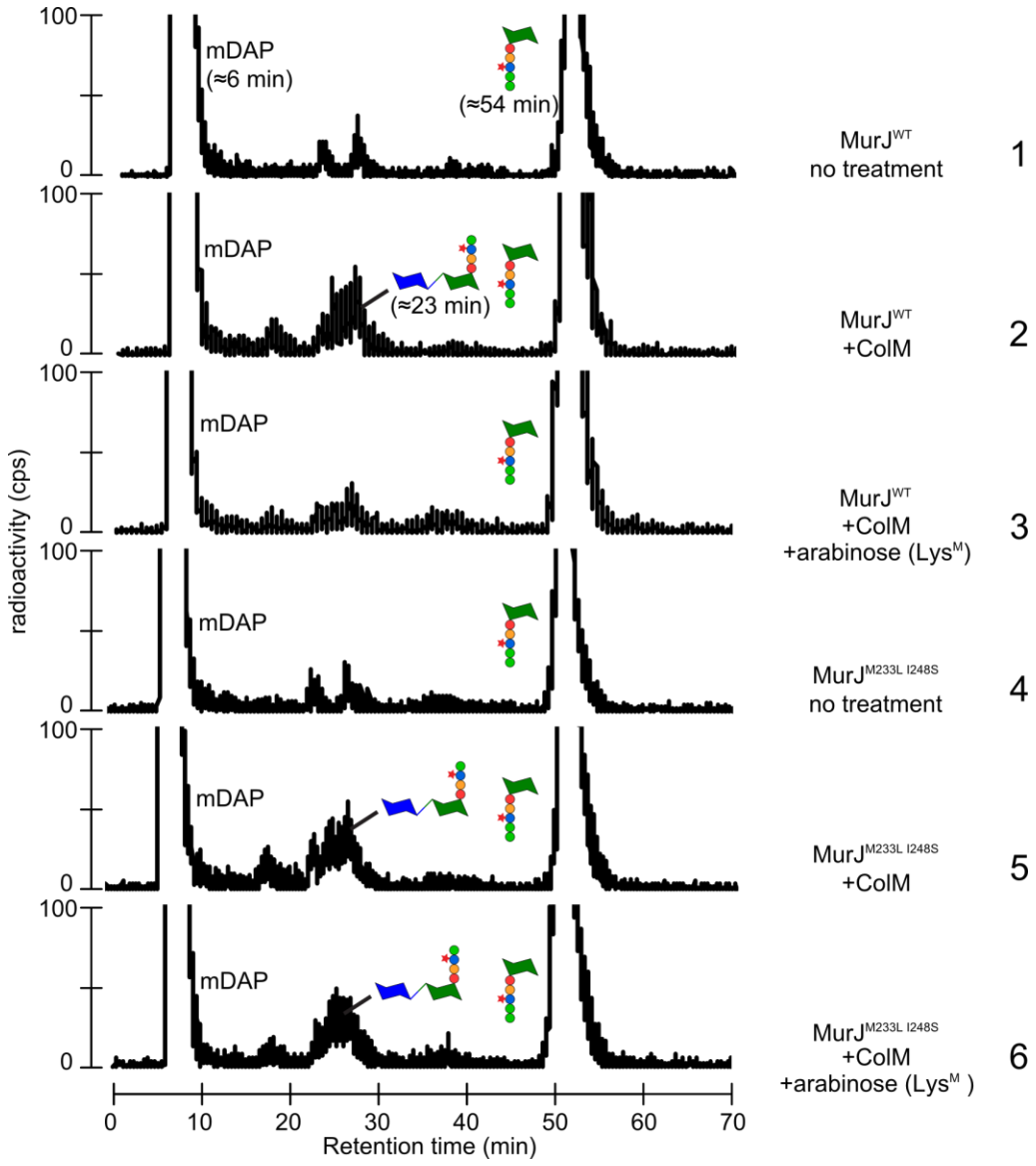

Figure 4.9 HPLC chromatograms for detection of ColM degradation products. Hot water extracts were prepared from the labeling experiments described in **Figure 4.8** ColM product and other soluble PG precursors were separated from the hot water extracts by reverse-phase (RP) HPLC using isocratic elution with 50mM

(**Figure 4.9 Continued**) triethylammonium formate pH 4.6, 6% (v/v) methanol on a Nucleosil C18 column followed by detection with inline radioflow detector (Berthold). The MurJ variants produced and the indicated treatments were labeled at the right of the corresponding chromatograms. Retention times of the labeled peaks were given by parentheses. Quantitation of ColM product (1-pyrophospho-MurNAc(pep4)-GlcNAc) was done by measuring the area under the ColM peak, and shown in **Figure 4.8**.

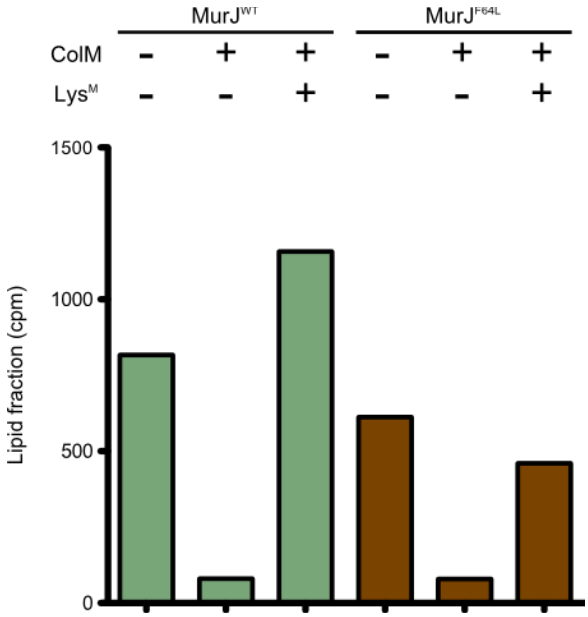

Figure 4.10 MurJF64L confers partial resistance to LysM.
Although Lys^M can still inhibit $MurJ^{F64L}$ activity, cells expressing $MurJ^{F64L}$ have more ColM sensitive Lipid II compared to the cells expressing wildtype MurJ. See **Figure 4.8** and Methods for experimental details.

Thus, like chemical inactivation of MurJ, Lys^M blocks Lipid II flipping without affecting its biosynthesis. When this assay was repeated with the Iml missense mutant *murJ*^{F64L}, partial Lys^M-mediated inhibition of Lipid II flipping was observed (Figure 4.10). To further increase resistance to Lys^M, various double Iml mutations were combined. Iml allele *murJ*$^{M233L, I248S}$ was ultimately selected for reduced lysis upon Lys^M induction by plasmid retention assay (see Methods) and by lysis profiles (Figure 4.3). When this allele was tested in the colicin M assay, Lipid II flipping was detected even when Lys^M was induced (Figure 4.8). Thus *murJ* mutations that block lysis after *lys* induction also prevent Lys^M from inhibiting the MurJ flippase activity. In addition, consistent with the hypothesis that Lys^M inhibits MurJ flippase activity, *E. coli* cells expressing an unrelated Lipid II flippase Amj from *Bacillis subtilis* (118) are also resistant to Lys^M (Figure 4.11). We therefore conclude that the phage M lysis gene encodes a specific inhibitor of MurJ.

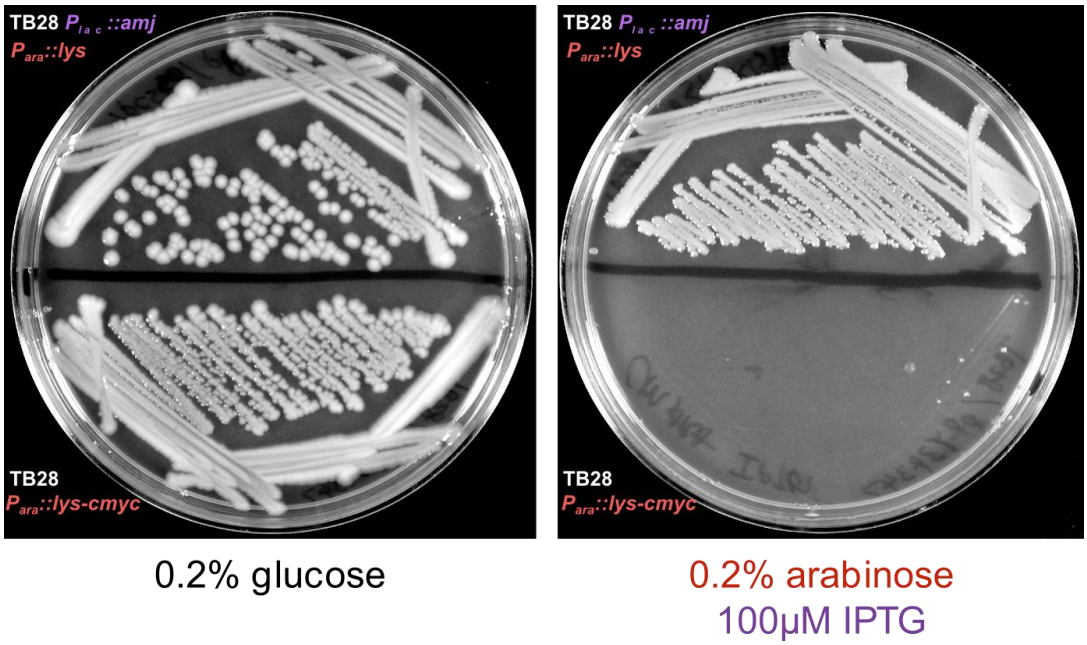

Figure 4.11 Amj from *Bacillus subtilis* rescues *E. coli* from Lys^{M}.
TB28 cells carrying either pKC5 (P_{ara}-*lys-cmyc cat*) alone or with pCS83 (P_{lac}-*amj spec*) were streaked on LB agar plates supplemented with 0.2 % (w/v) glucose (left) or 0.2% (w/v) arabinose and 100 μM IPTG (right) and incubated at 37°C overnight.

Discussion

The targeting of MurJ by Lys^M is the third instance of a single gene lysis system targeting an essential step in the peptidoglycan biosynthesis pathway; the other two are the ϕX174 E protein, which is a specific, non-competitive inhibitor of MraY (122), and Qβ A_2, which inhibits MurA (130, 132). In each case, no genomic space is dedicated to the lysis function. ϕX174 E is encoded in the +1 reading frame of an essential morphogenesis gene (Figure 4.1). In Qβ, A_2 is the Mat protein, which is present as a single molecule on the capsid of all leviviruses, serving to recognize and bind to the F pilus. However, in Qβ, A_2 has evolved an additional ability to bind to and inhibit MurA. In phage M, *lys* has evolved at an entirely new site within the *rep* gene, again out of frame. Thus, like E and A_2, Lys^M, the third phage-encoded "protein antibiotic", occupies no dedicated genetic space. Recent analysis of publicly available metagenomes and transcriptomes has revealed that the ssRNA leviviruses are much more ubiquitous and diverse than previously appreciated (55, 142), at least in part because many DNA-centric metagenomics protocols involve extensive use of RNase during virome sampling. Some estimates suggest that in gut samples, RNA phages can represent 10% or more of the total virome (177), constituting a deep resource of genomic "dark matter". Our results suggest that the leviviruses, empowered by the extremely error-prone character of their RNA-dependent RNA replicase (178), is capable of generating a vast environmental library of small proteins that can effect rapid lysis of bacterial hosts by targeting proteins involved in building the cell envelope. The long-neglected genetics of ssRNA phages

may be thus turn out to be powerful tools for development of new antibiotics effective against resistant bacteria.

CHAPTER V

CONCLUSIONS AND DISCUSSION

The results presented in this dissertation demonstrate that lysis proteins of small lytic phages, especially leviviruses, target different steps in the universally conserved PG biosynthesis pathway. Previous work on classic SGL systems, A_2 (Qβ) and E (ϕX174), shows them to be specific non-competitive inhibitors of PG biosynthesis enzymes MurA and MraY, respectively. In contrast, the molecular understanding of the lysis mechanism of L from MS2 has remained elusive, although earlier physiological and biochemical experiments indicated that lysis is likely elicited by triggering host autolysis without affecting PG biosynthesis (138, 179). The first genetic evidence of a host factor involved in the L lysis pathway was provided by the isolation of a *dnaJ* allele (*dnaJ*$_{P330Q}$) that conferred resistance to L and the demonstration that DnaJ interacted with the N-terminal half of L (Chapter II). In the known prokaryotic DnaJ proteins, the proline residue at position 330 is the most conserved residue in the C-terminal chaperone and dimerization domain of DnaJ. The loss of the N-terminal half of L in *L*$_{odJ}$ (*L overcomes dnaJ*) alleles abrogates the dependence of L lysis on the presence of DnaJ. The studies presented in Chapter II therefore indicate that DnaJ does not play an essential role in the lytic mechanism but it might be involved stabilizing the interaction between L and its unknown target. The genetic analysis presented in Chapter II is consistent with the finding that C-terminal half of L is the lytic determinant, while the N-terminal half functions as a regulatory domain controlling the timing of lysis via its interaction with DnaJ. Additionally, genetic analysis of L (Chapter III) has revealed that most critical

residues for lytic activity are clustered around an LS-dipeptide sequence that includes the highly-conserved serine residue at position 49. Even highly conservative changes around this region, such as L44V, F47Y, and S49T abolish the lytic activity of L. In addition, all the non-functional alleles isolated were recessive; suggesting that L functions as a monomer and possibly forms a tight helix-helix interaction with the target in the membrane. The LS dipeptide sequence can be found in other SGL proteins irrespective of sequence similarity with L, and in all of these systems the LS sequence demarcates the C-terminus of a hydrophobic oligopeptide sequence.

To determine if other ssRNA phages target PG biosynthesis enzymes other than MurA, we performed genetic and biochemical analysis of the mechanism of Lys^{M} (*E. coli* phage M)-induced lysis (Chapter IV). In these studies, multi-copy suppressor plasmids carrying *E. coli* chromosomal fragments encompassing *murJ*, the lipid II flippase, gene were isolated. Additionally, one isolate carried a chromosomal fragment encompassing the gene *ydeA*, an arabinose efflux pump. Furthermore, nine *Insensitive to M lysis* (Iml) mutants were isolated, of which, seven were unique changes that mapped to 2 out of 14 TMDs of MurJ. Of the seven, two mapped to TMD2 and five to TMD7. Lys^{M}-mediated lysis can be blocked by overexpression of functional MurJ as variants $murJ_{R18A}$, $murJ_{R24A}$, and $murJ_{R270A}$ were unable to rescue cells from lysis. Lys^{M} inhibition of MurJ was demonstrated *in vivo* by my collaborator L-T. Sham using a colicin M-based Lipid II flippase assay, providing further evidence to support the claim that MurJ is the sole Lipid II flippase in *E. coli*. Thus, in addition to A_2 and E, Lys^{M} also functions

as an inhibitor of cell wall biosynthesis, but instead of targeting the first committed step or first lipid-linked step, it targets the last lipid-linked cytoplasmic step of PG synthesis.

Current and future work on SGLs and ssRNA phages

L-mediated lysis

The molecular mechanism of L-mediated lysis is still not known. The prevailing operational model, that L activates a mysterious host autolytic response, was proposed by Holtje and van Duin in 1989 and still stands. However, neither the identity of the host target nor molecular details of the putative autolysis has been determined, despite multiple attempts with genetic and biochemical approaches. In this study, two genetic approaches were employed. First, spontaneous host mutants that are insensitive to LodJ were selected and two mutant candidates were isolated from 50 independent experiments. However, in both candidates the *LodJ* gene on the plasmid was disrupted by a nonsense mutation or by the insertion of a transposon. The fact that these extremely rare plasmid segregants were isolated, rather than mutations in the L target, points to the idea that the L interaction site on the target is either essential or is very small, possibly one or two residues. The second approach involved selecting suppressors of L-mediated lysis from a library of *E. coli* chromosomal fragments in a multi-copy plasmid platform. All the suppressors that were isolated carried chromosomal fragments encompassing the gene *ydeA*, an arabinose efflux pump. YdeA was probably pumping out the inducer and conferred resistance to L lysis. The genetic analysis of L combined with the difficulty in obtaining host mutants resistant to L suggest that either the target or the target-L interaction site is over-determined. One approach to circumvent this problem is to

identify evolutionarily distant L-like proteins that are functional in *E. coli* and employ such genes in genetic selections to identify the target. The L-like candidates from *Pseudomonas* (PP7 and PRR1) and *Acinetobacter* (AP205) phages could provide a promising start in this regard. The lysis protein from AP205, a 35 aa peptide, is particularly interesting as it evolved in a completely new genomic location at the 5' end of the viral genome, rather than at the usual *coat-rep* junction. Another approach is to block L-mediated lysis with multi-copy suppressor plasmids harboring chromosomal fragments from an evolutionarily distant host. Along similar lines a genetic approach that may be worth pursuing is multi-copy suppression with environmental DNA (eDNA), DNA isolated from microbes in the soil. In this scenario, heterologous expression of an evolutionarily distant target from eDNA cosmid library would protect *E. coli* from L-mediated lysis and thus aid in identifying the homologous target in *E. coli*.

phiCb5 lysis gene

The lysis gene of *Caulobacter crescentus* ssRNA phage phiCb5 has evolved as an embedded gene in the +1 reading frame of *rep* (Figure 1.7). The genomic location of the lysis gene mirrors that of phage M yet the protein shares many features with E from ϕX174. There is no sequence similarity between Lys^{phiCb5} and E, but the two proteins share features such as N-out, C-in membrane topology, a highly basic and proline-rich C-terminus, and a short N-terminus (Figure 5.1). The striking similarity in domain architecture despite unrelated primary structures between the two suggests the possibility that Lys^{phiCb5} has independently evolved to target MraY. The high mutation rates in

ssRNA phages, lack of a DNA stage in RNA replication, and lack of RNA recombination mechanisms all point to an independent evolution event. However, expression of Lys^{phiCb5} in *E. coli* does not cause lysis.

To determine if Lys^{phiCb5} targets *Caulobacter* MraY, an *E. coli* strain with its native *mraY* gene replaced with *Caulobacter crescentus mraY* is currently being constructed. Lys^{phiCb5}-mediated lysis in this strain would suggest that the target is MraY. Additionally, this strain could be used to isolate mutations that confer resistance to Lys^{phiCb5} and thus map the interface between the two proteins. If Lys^{phiCb5} targets MraY it would be interesting to see if the TMDs targeted by are the same as E (TMDs 5 and 9). The striking similarity with E also raises some interesting questions; for example, is the C-terminal domain required for lysis? Additionally, it would be interesting to assess whether Lys^{phiCb5} requires the *C. crescentus* equivalent of SlyD, as a potential regulatory interaction.

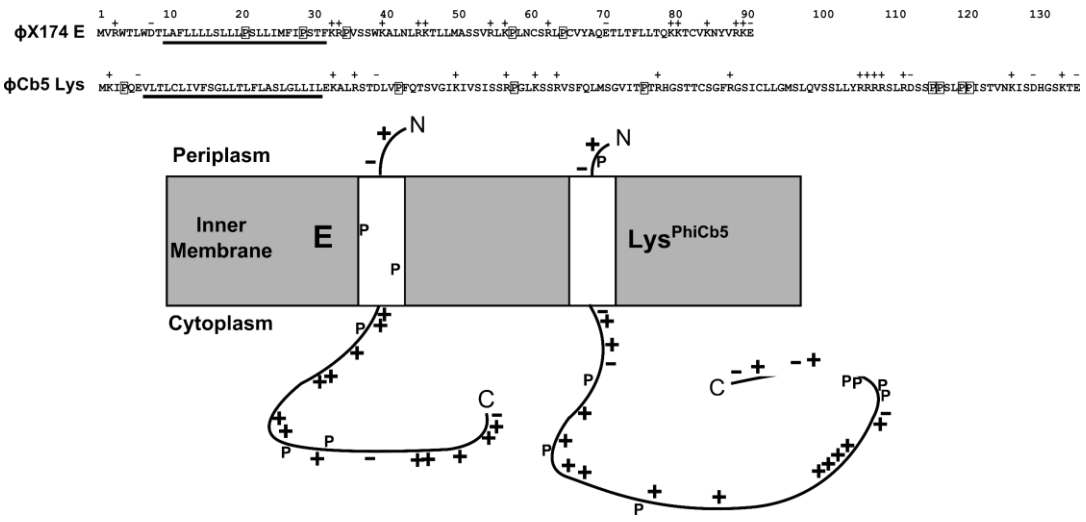

Figure 5.1 The similarity between ϕX174 E and Lys^{PhiCb5} lysis proteins.
The primary structures of E and Lys^{PhiCb5} are shown (top). The basic and acidic residues are indicated above the primary structure by + and – symbols respectively. Proline residues are boxed and residues corresponding to predicted TMDs are underlined. Membrane topologies of both E and Lys ^{PhiCb5} are shown (bottom). The N- and C-termini of the proteins are indicated by N and C respectively. Proline, basic, and acidic residues are shown as P, +, and – respectively.

Lysis genes from viral dark matter

Recently, two separate groups have identified a total of 186 novel ssRNA genomes (55, 142). Of the 122 novel genomes identified by Krishnamurthy et al only 20 genomes were either complete or nearly complete. The rest were partial genomes containing either *mat* or *rep* genes. Among 186 total genomes, only one genome (AVE017) had an annotated lysis gene, which was identified based on weak (<20%) aa similarity with the MS2 L primary structure. Thus, the genomic sequences of these new phages could be considered as "viral dark matter", which serves as a great source for mining new lysis gene candidates and novel mechanisms of lysis. However, the lysis genes of ssRNA phages pose a different set of challenges for existing bioinformatic tools. For example, lysis genes are often small and are embedded within other genes, both features that confound most gene-calling tools. Additionally, due to the high error-prone nature of Rep, the lysis proteins generally lack any significant sequence similarity and thus are often missed in homology based searches. To overcome these difficulties, a set of criteria should be defined and then algorithms designed to identify the likely lysis gene candidates. Since many of the known SGL proteins have a single TMD, three simple criteria can be established to identify lysis gene candidates with a TMD:

1. An ORF with a Shine-Dalgarno sequence and any of the four start codons (ATG, GTG, TTG, and CTG).
2. The ORF should be at least 20 codons (as a helix requires a minimum of 16 hydrophobic aa to span the lipid bilayer).
3. The ORF should have at least one predicted TMD.

Additionally, another set of criteria could be set to identify candidates that have the characteristic features of L-like proteins:

1. An ORF (>20 codons) with a Shine-Dalgarno sequence upstream of any of the 4 start codons (ATG, GTG, TTG, and CTG).
2. A minimum of two positive charged residues at the N-terminus (K or R)
3. A stretch of at minimum 10 mainly hydrophobic amino acids (AGVLIPF).
4. The presence of the LS motif at the C-terminus of hydrophobic stretch.

Although the host organism for these new ssRNA phages is not known, the candidates identified through the above criteria could be tested for lytic activity in *E. coli*, as the PG biosynthesis enzymes are highly conserved. The gene candidates could be codon-optimized for expression in *E. coli*, chemically synthesized, cloned in *E. coli* expression plasmids, and screened for lytic activity.

New frontiers in ssRNA phage biology

Isolation of new ssRNA phages

It is clear from recent RNA virome studies that the diversity of ssRNA phages has been vastly underappreciated. The diversity of lysis mechanisms among the known ssRNA phages and the discovery of new genes of unknown function warrant a concerted effort by phage biologists to isolate new ssRNA phages. Since the available ssRNA phage genomes are over-represented by F-specific phages, isolation of new ssRNA phages should focus on phages that are specific for retractable pili other than F-like pili. In addition to differential plating on pili (+) and pili (-) strains and RNase sensitivity assays, steps that aid in selectively enriching ssRNA phages should be included in phage isolation protocols. For example, adding a size selection step after initial phage enrichment stage eliminates vast number of large dsDNA phages and thus reduce the number phages that have to be differentially plated.

Revisiting the scanning model for translation of embedded genes

It is believed that translation of the lysis gene of MS2 is dependent on translation reinitiation by ribosomes that reach the *coat* stop codon (180, 181). It is thought that the ribosomes that finished translation of *coat* randomly drift a short distance and scan for nearby start codons without ever leaving the mRNA. It is estimated that about 5% of the ribosomes that terminate at the coat stop codon backtrack to the *L* start codon and initiate translation of L. Though it is an attractive model for translation of an embedded gene, it does raise some interesting questions. How does the backtracking happen? Why do ribosomes do not dissociate? What conformational changes in ribosomes are required to backtrack and reinitiate? Some of these questions could be answered by cryo-electron microscopy and asymmetric reconstruction of translating ribosomes.

De novo evolution of a gene

Ever since the publication of The Origin of Species by Charles Darwin in 1859, different aspects of evolutionary theory have been thoroughly investigated. The advent of molecular biology techniques, advances in sequencing technology, and computational capabilities have accelerated our understanding of the molecular underpinnings of evolution. However, *de novo* evolution of a new gene with a novel function within a limited genomic space has not been investigated. Because of their limited genomic space, short life cycle, high mutation rates, and a relatively high progeny per infection cycle, ssRNA phages are ideal candidates to study *de novo* evolution of genes. Lysis genes in these phages are usually found embedded in other genes and take up minimal coding space, which indicates that they evolved recently, probably after speciation to a

new retractable pilus. The most common lysis proteins of *Leviviridae* are small membrane proteins that target membrane proteins of PG biosynthesis. This simplifies the problem of protein evolution from a three-dimensional problem to a two-dimensional problem. One could hypothesize that the first event is the evolution of a stretch of hydrophobic residues that can form a TMD and the subsequent events are aimed at refining target recognition and timing of lysis. Thus, lysis genes make them an ideal candidate to study *de novo* evolution. This simple ssRNA phage system could be engineered to go a step back in the evolutionary pathway by abolishing the lysis gene and then evolving the phages that restore the lysis function. Deep sequencing the evolving ssRNA populations would provide insights into the genetic variations that lead to the origin of a new gene.

REFERENCES

1. Weinbauer MG, Rassoulzadegan F. 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol 6**:**1-11.
2. Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. MicrobiolMolBiolRev 64**:**69-114.
3. Molineux IJ, Panja D. 2013. Popping the cork: mechanisms of phage genome ejection. Nat Rev Microbiol 11**:**194-204.
4. Young R. 2013. Phage lysis: do we have the hole story yet? Curr Opin Microbiol 16**:**790-797.
5. Katsura I, Hendrix RW, Roberts JW, Stahl FW, Weisberg RA. 1983. Tail assembly and injection, p 331-346, Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
6. Goldberg EB, Grinius L, Letellier L. 1994. Recognition, Attachment and Injection, p 347-357. *In* Karam JD (ed), Molecular biology of bacteriophage T4, vol 1. American Society for Microbiology, Washington D.C.
7. Sawant P, Eissenberger K, Karier L, Mascher T, Bramkamp M. 2016. A dynamin-like protein involved in bacterial cell membrane surveillance under environmental stress. Environ Microbiol 18**:**2705-2720.
8. Abedon ST, Hyman P, Thomas C. 2003. Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. Appl Environ Microbiol 69**:**7499-7506.
9. Wang IN. 2006. Lysis timing and bacteriophage fitness. Genetics 172**:**17-26.

10. Reader RW, Siminovitch L. 1971. Lysis defective mutants of bacteriophage lambda - genetics and physiology of S cistron mutants. Virology 43**:**607-622.

11. Hutchison CA, III, Sinsheimer RL. 1966. The process of infection with bacteriophage ϕX174. X. Mutations in a ϕX17 lysis gene. J Mol Biol 18**:**429-447.

12. Young R. 2014. Phage lysis: three steps, three choices, one outcome. J Microbiol 52**:**243-258.

13. Berry JD, Rajaure M, Pang T, Young R. 2012. The spanin complex is essential for lambda lysis. J Bacteriol 194**:**5667-5674.

14. Bernhardt TG, Wang IN, Struck DK, Young R. 2002. Breaking free: "protein antibiotics" and phage lysis. Res Microbiol 153**:**493-501.

15. Ackermann H-W. 2006. Classification of bacteriophages, p 8-16. *In* Calendar R (ed), The Bacteriophages. Oxford, Oxford.

16. van Duin J, Tsareva N. 2006. Single-stranded RNA phages, p 179-196. *In* Calendar R (ed), The Bacteriophages. Oxford University Press, New York.

17. Fane BA, Brentlinger KL, Burch AD, Chen M, Hafenstein S, Moore E, Novak CR, Uchiyama A. 2006. ϕX174 et al., the *Microviridae*, p 129-145. *In* Calendar R (ed), The Bacteriophages, 2nd ed. Oxford University Press, Oxford.

18. Labonte JM, Suttle CA. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J 7**:**2169-2177.

19. Tucker KP, Parsons R, Symonds EM, Breitbart M. 2011. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. ISME J 5**:**822-830.

20. Garner SA, Everson JS, Lambden PR, Fane BA, Clarke IN. 2004. Isolation, Molecular Characterisation and Genome Sequence of a Bacteriophage (Chp3) from *Chlamydophila pecorum*. Virus Genes 28**:**207-214.

21. Liu BL, Everson JS, Fane B, Giannikopoulou P, Vretou E, Lambden PR, Clarke IN. 2000. Molecular characterization of a bacteriophage (Chp2) from *Chlamydia psittaci* 2. Journal of Virology 74**:**3464-3469.

22. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, Fiddes JC, Hutchison CA, 3rd, Slocombe PM, Smith M. 1978. The nucleotide sequence of bacteriophage ϕX174. J Mol Biol 125**:**225-246.

23. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison CA, III, Slocombe PM, Smith M. 1977. Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265**:**687-695.

24. Hutchison CA, III, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M. 1978. Mutagenesis at a specific position in a DNA sequence. J Biol Chem 253**:**6551-6560.

25. Smith HO, Hutchison CA, 3rd, Pfannkoch C, Venter JC. 2003. Generating a synthetic genome by whole genome assembly: ϕX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A 100**:**15440-15445.

26. Hayashi M. 1978. Morphogenesis of the isometric phages, p 531-547. *In* Denhardt DT, Dressler D, Ray DS (ed), The single-stranded DNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

27. Chen M, Uchiyama A, Fane BA. 2007. Eliminating the requirement of an essential gene product in an already very small virus: scaffolding protein B-free ϕX174, B-free. J Mol Biol 373**:**308-314.

28. Incardona NL, Selvidge L. 1973. Mechanism of adsorption and eclipse of bacteriophage ϕX174. II. Attachment and eclipse with isolated *Escherichia coli* cell wall lipopolysaccharide. J Virol 11**:**775.

29. Ohkawa T. 1980. On the structure of the lipopolysaccharide core in the cell wal of *Escherichia coli* K12 W2252-llu^{-} and its Ter-mutant cells. Biochem Biophys Res Commun 95**:**938.

30. Godson GN, Fiddes JC, Barrell BG, Sanger F, Denhardt DT, Dressler D, Ray DS. 1978. Comparative DNA sequence analysis of the G4 and ϕX174 genomes, p 51-86, The Single-stranded DNA Phages. Cold Spring Harbor Laboratory, Cold Spring Harbor.

31. Kodaira KI, Nakano K, Okada S, Taketo A. 1992. Nucleotide sequence of the genome of bacteriophage α3: Interrelationship of the genome structure of the gene products with those of the phages ϕX174, G4 and ϕK. BiochimBiophysActa 1130**:**277-288.

32. Ilag LL, McKenna R, Yadav MP, BeMiller JN, Incardona NL, Rossmann MG. 1994. Calcium ion-induced structural changes in bacteriophage ϕX174. J Mol Biol 244**:**291-300.

33. Incardona NL. 1983. A kinetic model for virus binding which involves release of cell-bound virus-receptor complexes. J Theor Biol 105**:**631-645.

34. Incardona NL, Tuech JK, Murti G. 1985. Irreversible binding of phage ϕX174 to cell-bound lipopolysaccharide receptors and release of virus-receptor complexes. Biochemistry 24**:**6439-6446.

35. Jazwinski SM, Lindberg AA, Kornberg A. 1975. The gene H spike protein of bacteriophages ϕX174 and S13. I. Functions in phage-receptor recognition and in transfection. Virology 66**:**283.

36. Sun L, Young LN, Zhang X, Boudko SP, Fokine A, Zbornik E, Roznowski AP, Molineux IJ, Rossmann MG, Fane BA. 2014. Icosahedral bacteriophage ϕX174 forms a tail for DNA transport during infection. Nature 505**:**432-435.

37. Shlomai J, Polder L, Arai K, Kornberg A. 1981. Replication of ϕX174 DNA with purified enzymes. I. Conversion of viral DNA to a supercoiled, biologically active duplex. J Biol Chem 256**:**5233-5238.

38. Eisenberg S, Griffith J, Kornberg A. 1977. ϕX174 cistron A protein is a multifunctional enzyme in DNA replication. Proc Natl Acad Sci U S A 74**:**3198-3202.

39. Hayashi M, Aoyama A, Richardson DL, Hayashi MN. 1988. Biology of the Bacteriophage ϕX174, p 1-71. *In* Calendar R (ed), The Bacteriophages. Plenum Press, New York.

40. Wang IN, Smith DL, Young R. 2000. Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol 54**:**799-825.

41. Young KD, Young R. 1982. Lytic action of cloned ϕX174 gene *E* J Virol 44**:**993-1002.

42. Maratea D, Young K, Young R. 1985. Deletion and fusion analysis of the ϕX174 lysis gene *E* Gene 40**:**39-46.

43. Vidaver AK, Koski RK, Ven Etten JL. 1973. Bacteriophage phi6: a lipid-containing virus of *Pseudomonas phaseolicola* Journal of Virology 11**:**799-805.

44. Semancik JS, Vidaver AK, Van Etten JL. 1973. Characterization of segmented double-helical RNA from bacteriophage phi6. J Mol Biol 78**:**617-625.

45. Day LA, Mindich L. 1980. The molecular weight of bacteriophage phi 6 and its nucleocapsid. Virology 103**:**376-385.

46. Mindich L. 2006. Phages with Segmented Double-Stranded RNA Genomes, p 197-207. *In* Calendar R (ed), The Bacteriophages. Oxford University Press, New York

47. Poranen MM, Paatero AO, Tuma R, Bamford DH. 2001. Self-assembly of a viral molecular machine from purified protein and RNA constituents. Mol Cell 7**:**845-854.

48. de Haas F, Paatero AO, Mindich L, Bamford DH, Fuller SD. 1999. A symmetry mismatch at the site of RNA packaging in the polymerase complex of dsRNA bacteriophage phi6. J Mol Biol 294**:**357-372.

49. Kainov DE, Pirttimaa M, Tuma R, Butcher SJ, Thomas GJ, Jr., Bamford DH, Makeyev EV. 2003. RNA packaging device of double-stranded RNA bacteriophages, possibly as simple as hexamer of P4 protein. J Biol Chem 278**:**48084-48091.

50. Juuti JT, Bamford DH, Tuma R, Thomas GJ, Jr. 1998. Structure and NTPase activity of the RNA-translocating protein (P4) of bacteriophage phi 6. J Mol Biol 279**:**347-359.

51. Olkkonen VM, Ojala PM, Bamford DH. 1991. Generation of infectious nucleocapsids by in vitro assembly of the shell protein on to the polymerase complex of the dsRNA bacteriophage phi 6. J Mol Biol 218**:**569-581.

52. Mindich L, Lehman J. 1979. Cell wall lysin as a component of the bacteriophage phi6 virion. Journal of Virology 30**:**489-496.

53. Johnson MD, III, Mindich L. 1994. Isolation and characterization of nonsense mutations in gene 10 of bacteriophage phi 6. Journal of Virology 68**:**2331-2338.

54. Mindich L, Qiao X, Qiao J, Onodera S, Romantschuk M, Hoogstraten D. 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J Bacteriol 181**:**4505-4508.

55. Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D. 2016. Hyperexpansion of RNA Bacteriophage Diversity. PLoS Biol 14**:**e1002409.

56. Loeb T, Zinder ND. 1961. A bacteriophage containing RNA. Proc Natl Acad Sci U S A 47**:**282-289.

57. Boedtker H, and Gesteland, RF. 1975. Physical properties of RNA bacteriophages and their RNA, p 1-28. *In* Zinder ND (ed), RNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor.

58. Valegard K, Liljas L, Fridborg K, Unge T. 1990. The three-dimensional structure of the bacterial virus MS2. Nature 345**:**36-41.

59. Capecchi M, and Webster, RE. 1975. Bacteriophage RNA as template for in vitro protein synthesis, p 279-299. *In* Zinder ND (ed), RNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor.

60. Furuse K, Hirashima A, Harigai H, Ando A, Watanabe K, Kurosawa K, Inokuchi Y, Watanabe I. 1979. Grouping of RNA coliphages based on analysis of the sizes of their RNAs and proteins. Virology 97**:**328-341.

61. Watanabe I, Miyake, T., Sakurai, T., Shiba, T., and , Ohno T. 1967. Isolation and grouping of RNA phages. Proc Jpn Acad**:**204-209.

62. Murphy FA, Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., Summers, M.D. 1995. Virus taxonomy : classification and nomenclature of viruses : sixth report of the International Committee on Taxonomy of Viruses Springer-Verlag, Vienna.

63. Crawford EM, Gesteland RF. 1964. The adsorption of bacteriophage R-17. Virology 22**:**165-167.

64. Brinton CC, Jr., Gemski P, Jr., Carnahan J. 1964. A New Type of Bacterial Pilus Genetically Controlled by the Fertility Factor of *E. coli* K 12 and Its Role in Chromosome Transfer. Proc Natl Acad Sci U S A 52**:**776-783.

65. Brinton CC. 1965. The structure, function, synthesis and genetic control of bacterial pili and a molecular model for DNA and rna transport in gram negative bacteria. Transactions of the New York Academy of Sciences 27**:**1003-1054.

66. Brinton CC, Jr. 1971. The properties of sex pili, the viral nature of "conjugal" genetic transfer systems, and some possible approaches to the control of bacterial drug resistance. CRC Crit Rev Microbiol 1**:**105-160.

67. Lawley TD, Klimke WA, Gubbins MJ, Frost LS. 2003. F factor conjugation is a true type IV secretion system. FEMS Microbiol Lett 224**:**1-15.

68. Ottow JC. 1975. Ecology, physiology, and genetics of fimbriae and pili. Annu Rev Microbiol 29**:**79-108.

69. Lawn AM. 1966. Morphological Features of the Pili Associated with *Escherichia coli* K12 Carrying R Factors or the F Factor. Microbiology 45**:**377-383.

70. Harrington LC, Rogerson AC. 1990. The F pilus of *Escherichia coli* appears to support stable DNA transfer in the absence of wall-to-wall contact between cells. J Bacteriol 172**:**7263-7264.

71. Valentine RC, Strand M. 1965. Complexes of F-pili and RNA Bacteriophage. Science 148**:**511-513.

72. Valentine RC, Wedel H. 1965. The extracellular stages of RNA bacteriophage infection. Biochem Biophys Res Commun 21**:**106-112.

73. Danziger RE, Paranchych W. 1970. Stages in phage R17 infection. II. Ionic requirements for phage R17 attachment to F-pili. Virology 40**:**547-553.

74. Manchak J, Anthony KG, Frost LS. 2002. Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. Mol Microbiol 43**:**195-205.

75. Kozak M, Nathans D. 1971. Fate of maturation protein during infection by coliphage MS2. NatNew Biol 234**:**209-211.

76. Krahn PM, O'Callaghan RJ, Paranchych W. 1972. Stages in phage R17 infection. VI. Injection of A protein and RNA into the host cell. Virology 47**:**628-637.

77. Paranchych W. 1966. Stages in phage R17 infection: The role of divalent cations. Virology 28**:**90-99.

78. Curtiss R, 3rd. 1969. Bacterial conjugation. Annu Rev Microbiol 23**:**69-136.

79. Bradley DE. 1972. Shortening of *Pseudomonas aeruginosa* pili after RNA-phage adsorption. J Gen Microbiol 72**:**303-319.

80. Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature 541**:**112-116.

81. Witherell GW, Gott JM, Uhlenbeck OC. 1991. Specific interaction between RNA phage coat proteins and RNA. Prog Nucleic Acid Res Mol Biol 40**:**185-220.

82. Beremand MN, Blumenthal T. 1979. Overlapping genes in RNA phage: a new protein implicated in lysis. Cell 18**:**257-266.

83. Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, van Duin J. 2002. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. J Gen Virol 83**:**1523-1533.

84. Bradley DE. 1966. The Structure and Infective Process of a *Pseudomonas Aeruginosa* Bacteriophage Containing Ribonucleic Acid. Microbiology 45**:**83-96.

85. Schmidt JM, Stanier RY. 1965. Isolation and characterization of bacteriophages active against stalked bacteria. J Gen Microbiol 39**:**95-107.

86. Shapiro L, Bendis, I. 1975. RNA phages of bacteria other than *E. coli*, p 397-410. *In* Zinder ND (ed), RNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor.

87. Gorzelnik KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J. 2016. Asymmetric cryo-EM structure of the canonical Allolevivirus Qβ reveals a single maturation protein and the genomic ssRNA *in situ*. Proc Natl Acad Sci U S A 113**:**11519-11524.

88. Webber K, Konigsberg, W. 1975. Proteins of the RNA phages, p 15-84. *In* Zinder ND (ed), RNA phages. Cold Spring Harbor Laboratory, Cold Spring Harbor.

89. Höltje JV. 1998. Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. Microbiology and Molecular Biology Reviews 62**:**181-203.

90. Vollmer W, Blanot D, De Pedro MA. 2008. Peptidoglycan structure and architecture. FEMS Microbiology Reviews 32**:**149-167.

91. Typas A, Banzhaf M, Gross CA, Vollmer W. 2012. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. Nature Reviews Microbiology 10**:**123-136.

92. Silhavy TJ, Kahne D, Walker S. 2010. The Bacterial Cell Envelope. Cold Spring Harbor Perspectives in Biology 2**:**a000414.

93. Gan L, Chen S, Jensen GJ. 2008. Molecular organization of Gram-negative peptidoglycan. Proc Natl Acad Sci U S A 105**:**18953-18957.

94. Brown ED, Vivas EI, Walsh CT, Kolter R. 1995. MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. J Bacteriol 177**:**4194-4197.

95. Yoon HJ, Lee SJ, Mikami B, Park HJ, Yoo J, Suh SW. 2008. Crystal structure of UDP-N-acetylglucosamine enolpyruvyl transferase from *Haemophilus influenzae* in complex with UDP-N-acetylglucosamine and fosfomycin. Proteins 71**:**1032-1037.

96. Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K. 1996. Structure of UDP-N-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-N-acetylglucosamine and the drug fosfomycin. Structure 4**:**1465-1474.

97. Schonbrunn E, Sack S, Eschenburg S, Perrakis A, Krekel F, Amrhein N, Mandelkow E. 1996. Crystal structure of UDP-N-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin. Structure 4**:**1065-1075.

98. Lovering AL, Safadi SS, Strynadka NC. 2012. Structural perspective of peptidoglycan biosynthesis and assembly. Annu Rev Biochem 81**:**451-478.

99. Teo AC, Roper DI. 2015. Core Steps of Membrane-Bound Peptidoglycan Biosynthesis: Recent Advances, Insight and Opportunities. Antibiotics (Basel) 4**:**495-520.

100. Chung BC, Zhao J, Gillespie RA, Kwon DY, Guan Z, Hong J, Zhou P, Lee SY. 2013. Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. Science 341**:**1012-1016.

101. Meeske AJ, Riley EP, Robins WP, Uehara T, Mekalanos JJ, Kahne D, Walker S, Kruse AC, Bernhardt TG, Rudner DZ. 2016. SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature 537**:**634-638.

102. Cho H, Wivagg CN, Kapoor M, Barry Z, Rohs PD, Suh H, Marto JA, Garner EC, Bernhardt TG. 2016. Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. Nat Microbiol doi:10.1038/nmicrobiol.2016.172**:**16172.

103. Barreteau H, Magnet S, El Ghachi M, Touze T, Arthur M, Mengin-Lecreulx D, Blanot D. 2009. Quantitative high-performance liquid chromatography analysis

of the pool levels of undecaprenyl phosphate and its derivatives in bacterial membranes. J Chromatogr B Analyt Technol Biomed Life Sci 877**:**213-220.

104. Bouhss A, Trunkfield AE, Bugg TD, Mengin-Lecreulx D. 2008. The biosynthesis of peptidoglycan lipid-linked intermediates. FEMS Microbiol Rev 32**:**208-233.

105. Brown S, Santa Maria JP, Walker S. 2013. Wall Teichoic Acids of Gram-Positive Bacteria. Annual review of microbiology 67**:**10.1146/annurev-micro-092412-155620.

106. Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res 338**:**2503-2519.

107. Weissborn AC, Rumley MK, Kennedy EP. 1991. Biosynthesis of membrane-derived oligosaccharides. Membrane-bound glucosyltransferase system from *Escherichia coli* requires polyprenyl phosphate. J Biol Chem 266**:**8062-8067.

108. Mohammadi T, van Dam V, Sijbrandi R, Vernet T, Zapun A, Bouhss A, Diepeveen-de Bruin M, Nguyen-Disteche M, de Kruijff B, Breukink E. 2011. Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. EMBO J 30**:**1425-1432.

109. Mohammadi T, Sijbrandi R, Lutters M, Verheul J, Martin NI, den Blaauwen T, de Kruijff B, Breukink E. 2014. Specificity of the transport of lipid II by FtsW in *Escherichia coli*. J Biol Chem 289**:**14707-14718.

110. Ruiz N. 2008. Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in Escherichia coli. Proc Natl Acad Sci U S A 105**:**15553-15557.

111. Inoue A, Murata Y, Takahashi H, Tsuji N, Fujisaki S, Kato J. 2008. Involvement of an essential gene, *mviN*, in murein synthesis in *Escherichia coli*. J Bacteriol 190**:**7298-7301.

112. Butler EK, Davis RM, Bari V, Nicholson PA, Ruiz N. 2013. Structure-function analysis of MurJ reveals a solvent-exposed cavity containing residues essential for peptidoglycan biogenesis in *Escherichia coli*. J Bacteriol 195**:**4639-4649.

113. Butler EK, Tan WB, Joseph H, Ruiz N. 2014. Charge requirements of lipid II flippase activity in *Escherichia coli*. J Bacteriol 196**:**4111-4119.

114. Sham LT, Butler EK, Lebar MD, Kahne D, Bernhardt TG, Ruiz N. 2014. Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. Science 345**:**220-222.

115. El Ghachi M, Bouhss A, Barreteau H, Touze T, Auger G, Blanot D, Mengin-Lecreulx D. 2006. Colicin M exerts its bacteriolytic effect via enzymatic degradation of undecaprenyl phosphate-linked peptidoglycan precursors. J Biol Chem 281**:**22761-22772.

116. Fay A, Dworkin J. 2009. Bacillus subtilis homologs of MviN (MurJ), the putative Escherichia coli lipid II flippase, are not essential for growth. J Bacteriol 191**:**6020-6028.

117. Vasudevan P, McElligott J, Attkisson C, Betteken M, Popham DL. 2009. Homologues of the Bacillus subtilis SpoVB protein are involved in cell wall metabolism. J Bacteriol 191**:**6012-6019.

118. Meeske AJ, Sham LT, Kimsey H, Koo BM, Gross CA, Bernhardt TG, Rudner DZ. 2015. MurJ and a novel lipid II flippase are required for cell wall biogenesis in Bacillus subtilis. Proc Natl Acad Sci U S A 112**:**6437-6442.

119. Bernhardt TG, Roof WD, Young R. 2000. Genetic evidence that the bacteriophage ϕX174 lysis protein inhibits cell wall synthesis. Proc Natl Acad Sci U S A 97**:**4297-4302.

120. Witte A, Wanner G, Bläsi U, Halfmann G, Szostak M, Lubitz W. 1990. Endogenous transmembrane tunnel formation mediated by ϕX174 lysis protein E. J Bacteriol 172**:**4109-4114.

121. Bouhss A, Mengin-Lecreulx D, Le Beller D, Van Heijenoort J. 1999. Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis. Mol Microbiol 34**:**576-585.

122. Zheng Y, Struck DK, Young R. 2009. Purification and functional characterization of ϕX174 lysis protein E. Biochemistry 48**:**4999-5006.

123. Buckley KJ, Hayashi M. 1986. Lytic activity localized to membrane-spanning region of ϕX174 E protein. Mol Gen Genet 204**:**120-125.

124. Tanaka S, Clemons WM, Jr. 2012. Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ϕX174. Mol Microbiol 85**:**975-985.

125. Roof WD, Young R. 1995. ϕ X174 lysis requires *slyD* , a host gene which is related to the FKBP family of peptidyl-prolyl cis - trans isomerases. FEMS Microbiol Rev 17**:**213-216.

126. Hottenrott S, Schumann T, Pl• ckthun A, Fischer G, Rahfeld JU. 1997. The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl cis/trans - isomerase. Journal of Biological Chemistry 272**:**15697-15701.

127. Bernhardt TG, Roof WD, Young R. 2002. The *Escherichia coli* FKBP-type PPIase SlyD is required for the stabilization of the E lysis protein of bacteriophage ϕX174. Mol Microbiol 45**:**99-108.

128. Roof WD, Horne SM, Young KD, Young R. 1994. *slyD*, a host gene required for ϕX174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. J Biol Chem 269**:**2902-2910.

129. Bernhardt TG, Roof WD, Young R. 2002. The *Escherichia coli* FKBP-type PPIase SlyD is required for the stabilization of the E lysis protein of bacteriophage ϕX174. Mol Microbiol 45**:**99-108.

130. Bernhardt TG, Wang IN, Struck DK, Young R. 2001. A protein antibiotic in the phage Qβ virion: diversity in lysis targets. Science 292**:**2326-2329.

131. Bernhardt TG, Struck DK, Young R. 2001. The lysis protein E of ϕX174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. J Biol Chem 276**:**6093-6097.

132. Reed CA, Langlais C, Kuznetsov V, Young R. 2012. Inhibitory mechanism of the Qβ lysis protein A_2. Mol Microbiol 86**:**836-844.

133. Bernhardt TG. 2001. Ph.D thesis. Breaking free: small phages inhibit murein synthesis to lyse their hostTexas A&M University, College Station.

134. McIntosh BK. 2008. Ph.D thesis. Bacteriophage MS2 L protein: Genetic and biochemical charecterizationTexas A&M University, College Station.

135. Berkhout B, de Smit MH, Spanjaard RA, Blom T, van Duin J. 1985. The amino terminal half of the MS2-coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. EMBO J 4**:**3315-3320.

136. Goessens WHF, Driessen AJM, Wilschut J, van Duin J. 1988. A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2-coded lysis protein dissipates the proton-motive force in *Escherichia coli* membrane vesicles by generating hydrophilic pores. EMBO J 7**:**867-873.

137. Walderich B, Holtje JV. 1989. Specific localization of the lysis protein of bacteriophage MS2 in membrane adhesion sites of *Escherichia coli.* J Bacteriol 171**:**3331-3336.

138. Walderich B, Ursinus-Wosner A, van Duin J, Holtje JV. 1988. Induction of the autolytic system of *Escherichia coli* by specific insertion of bacteriophage MS2 lysis protein into the bacterial cell envelope. J Bacteriol 170**:**5027-5033.

139. Kellenberger E. 1990. The 'Bayer bridges' confronted with results from improved electron microscopy methods. Mol Microbiol 4**:**697-705.

140. Kazaks A, Voronkova T, Rumnieks J, Dishlers A, Tars K. 2011. Genome structure of *Caulobacter* phage phiCb5. J Virol 85**:**4628-4631.

141. Rumnieks J, Tars K. 2012. Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. BMC Microbiol 12**:**277.

142. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. 2016. Redefining the invertebrate RNA virosphere. Nature doi:10.1038/nature20167.

143. Atkins JF, Steitz JA, Anderson CW, Model P. 1979. Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function. Cell 18**:**247-256.

144. Holtje JV, van Duin J. 1984. MS2 phage induced lysis of *E. coli* depends upon the activity of the bacterial autolysins, p 195-199. *In* Nombela C (ed), Microbial Cell Wall Synthesis and Autolysis. Elsevier Science Publishers, New York.

145. Smith DL, Chang CY, Young R. 1998. The λ holin accumulates beyond the lethal triggering concentration under hyper-expression conditions. Gene Expr 7**:**39-52.

146. Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site in *E. coli* as gamma delta. Cold Spring Harb Symp Quant Biol 45**:**135-140.

147. Bernhardt TG, de Boer PA. 2003. The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. Mol Microbiol 48**:**1171-1182.

148. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-

frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2**:**2006 0008.

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

150. Miller JH. 1972. Generalized transduction; use of P1 in strain construction, p 201-205, Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

151. Miller JH. 1972. M9 minimal medium, p 431, Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

152. Berry J, Savva C, Holzenburg A, Young R. 2010. The lambda spanin components Rz and Rz1 undergo tertiary and quaternary rearrangements upon complex formation. Protein Sci 19**:**1967-1977.

153. Moussa SH, Kuznetsov V, Tran TA, Sacchettini JC, Young R. 2012. Protein determinants of phage T4 lysis inhibition. Protein Sci 21**:**571-582.

154. Lopilato J, Bortner S, Beckwith J. 1986. Mutation in a new chromosonal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. Mol Gen Genet 205**:**285-290.

155. Bardwell JC, Tilly K, Craig E, King J, Zylicz M, Georgopoulos C. 1986. The nucleotide sequence of the *Escherichia coli* K12 *dnaJ*$^{+}$ gene. A gene that encodes a heat shock protein. J Biol Chem 261**:**1782-1785.

156. Holtje JV, Fiedler W, Rotering H, Walderich B, van Duin J. 1988. Lysis induction of *Escherichia coli* by the cloned lysis protein of the phage MS2 depends on the presence of osmoregulatory membrane-derived oligosaccharides. J Biol Chem 263**:**3539-3541.

157. Lubitz W, Halfmann G, Plapp R. 1984. Lysis of *Escherichia coli* after infection with ϕX174 depends on the regulation of the cellular autolytic system. J Gen Microbiol 130**:**1079-1087.

158. Shi YY, Hong XG, Wang CC. 2005. The C-terminal (331-376) sequence of *Escherichia coli* DnaJ is essential for dimerization and chaperone activity: a small angle X-ray scattering study in solution. J Biol Chem 280**:**22761-22768.

159. Puvirajesinghe TM, Elantak L, Lignon S, Franche N, Ilbert M, Ansaldi M. 2012. DnaJ (Hsp40 protein) binding to folded substrate impacts KplE1 prophage excision efficiency. J Biol Chem 287**:**14169-14177.

160. Winter RB, Gold L. 1983. Overproduction of bacteriophage Qβ maturation (A_2) protein leads to cell lysis. Cell 33**:**877-885.

161. Zheng Y, Struck DK, Bernhardt TG, Young R. 2008. Genetic analysis of MraY inhibition by the ϕX174 protein E. Genetics 180**:**1459-1466.

162. Ruokoranta TM, Grahn AM, Ravantti JJ, Poranen MM, Bamford DH. 2006. Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it in the Levivirus genus with characteristics shared with Allolevivirus. J Virol 80**:**9326-9330.

163. Kannoly S, Shao Y, Wang IN. 2012. Rethinking the evolution of single-stranded RNA (ssRNA) bacteriophages based on genomic sequences and characterizations of two R-plasmid-dependent ssRNA phages, C-1 and Hgal1. J Bacteriol 194**:**5073-5079.

164. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43**:**W389-394.

165. Yohannan S, Yang D, Faham S, Boulting G, Whitelegge J, Bowie JU. 2004. Proline substitutions are not easily accommodated in a membrane protein. J Mol Biol 341**:**1-6.

166. Kuk AC, Mashalidis EH, Lee SY. 2017. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. Nat Struct Mol Biol 24**:**171-176.

167. Ursell TS, Nguyen J, Monds RD, Colavin A, Billings G, Ouzounov N, Gitai Z, Shaevitz JW, Huang KC. 2014. Rod-like bacterial shape is maintained by feedback between cell curvature and cytoskeletal localization. Proc Natl Acad Sci U S A 111**:**E1025-1034.

168. Morgenstein RM, Bratton BP, Nguyen JP, Ouzounov N, Shaevitz JW, Gitai Z. 2015. RodZ links MreB to cell wall synthesis to mediate MreB rotation and robust morphogenesis. Proc Natl Acad Sci U S A 112**:**12510-12515.

169. Fischetti VA. 2008. Bacteriophage lysins as effective antibacterials. Curr Opin Microbiol 11**:**393-400.

170. Young R, Wang IN. 2006. Phage lysis, p 104-126. *In* Calendar R (ed), The Bacteriophages, 2nd ed. Oxford University Press, Oxford.

171. Casadaban MJ. 1976. Transposition and fusion of the lac genes to selected promoters in Escherichia coli using bacteriophage lambda and Mu. J Mol Biol 104**:**541-555.

172. Ruiz N, Gronenberg LS, Kahne D, Silhavy TJ. 2008. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. Proc Natl Acad Sci U S A 105**:**5537-5542.

173. Yunck R, Cho H, Bernhardt TG. 2016. Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. Mol Microbiol 99**:**700-718.

174. Elhenawy W, Davis RM, Fero J, Salama NR, Felman MF, Ruiz N. 2016. The O-Antigen Flippase Wzk Can Substitute for MurJ in Peptidoglycan Synthesis in *Helicobacter pylori* and *Escherichia coli*. PLoS One 11**:**e0161587.

175. Bost S, Silva F, Belin D. 1999. Transcriptional activation of *ydeA*, which encodes a member of the major facilitator superfamily, interferes with arabinose accumulation and induction of the *Escherichia coli* arabinose PBAD promoter. J Bacteriol 181**:**2185-2191.

176. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9**:**40.

177. Osawa S, Furuse K, Watanabe I. 1981. Distribution of ribonucleic acid coliphages in animals. Appl Environ Microbiol 41**:**164-168.

178. Domingo E, Holland JJ. 1997. RNA virus mutations and fitness for survival. Annu Rev Microbiol 51**:**151-178.

179. Ursinus-Woosner A, Holtje JV. 1989. Functioning of the cloned phage MS2 lysis protein in Escherichia coli impaired in murein synthesis. FEMS Microbiology Immunology 48**:**75-79.

180. Berkhout B, Schmidt BF, van Strien A, van Boom J, van Westrenen J, van Duin J. 1987. Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. J Molelcular Biology 195**:**517-524.

181. van Duin J, Calendar R. 1988. Single-stranded RNA bacteriophages, p 117-167. *In* Fraenkel-Conrat H, Wagner RR (ed), The Bacteriophages, vol 1. Plenum Press, New York.