Phage single-gene lysis: Finding the weak spot in the bacterial cell wall

Karthik Chamakura and Ry Young

From the Department of Biochemistry and Biophysics and the Center for Phage Technology, Texas A&M University, College Station, Texas 77843-2128

Edited by Chris Whitfield

In general, the last step in the vegetative cycle of bacterial viruses, or bacteriophages, is lysis of the host. dsDNA phages require multiple lysis proteins, including at least one enzyme that degrades the cell wall (peptidoglycan (PG)). In contrast, the lytic ssDNA and ssRNA phages have a single lysis protein that achieves cell lysis without enzymatically degrading the PG. Here, we review four “single-gene lysis” or Sgl proteins. Three of the Sgls block bacterial cell wall synthesis by binding to and inhibiting several enzymes in the PG precursor pathway. The target of the fourth Sgl, L from bacteriophage MS2, is still unknown, but we review evidence indicating that it is likely a protein involved in maintaining cell wall integrity. Although only a few phage genomes are available to date, the ssRNA Leviviridae are a rich source of novel Sgls, which may facilitate further unraveling of bacterial cell wall biosynthesis and discovery of new antibacterial agents.

Introduction to small lytic phages and “single-gene lysis”

By definition, the lytic bacteriophages encode proteins for disruption of the host envelope. The large dsDNA phages, the Caudovirales, have multiple lysis proteins, including holins, endolysins, and spanins, targeting the cytoplasmic or inner membrane (IM), peptidoglycan (PG), and outer membrane (OM), respectively, as well as multiple proteins that regulate the lysis process (1, 2). In contrast, the small lytic phages of Gram-negative hosts, comprising the ssDNA (Microviridae) and ssRNA (Leviviridae), achieve host lysis by a single gene, encoding a protein lacking any PG-degrading activity (3). This review exclusively focuses on these single-gene lysis (Sgl) proteins of small lytic phages.

Bacterial cell wall structure and biosynthesis

An exploration of Sgl mechanisms requires a brief review of the structure of the Gram-negative cell wall and its biosynthesis. The key to the structure and shape of the cell envelope is the PG layer, consisting of 2–3 layers of glycan strands made up of repeating disaccharide units of MurNac-pentapeptide and GlcNac, cross-linked by peptide bridges between pentapeptide side chains of MurNacs of adjacent strands (Fig. 1A) (4–6). The PG has considerable tensile strength (3–300 megapascals), allowing the cell to tolerate high internal osmotic pressures (3–10 atmospheres) while maintaining shape (7, 8). The entire PG of a cell can be isolated as a single complex polymer, the sacculus, studies of which have revealed that the glycan chains run almost perpendicular to the long axis of the cell (9). In most Gram-negative bacteria, the PG layer is covalently linked through >10⁵ peptide linkages to the C-terminal Lys residue of the major lipoprotein, Lpp; the PG-linked Lpp is anchored almost exclusively in the inner leaflet of the IM (5, 10, 11).

Biosynthesis of the PG can be divided into cytoplasmic, membrane, and periplasmic stages. There are seven enzymatic steps in the cytoplasm, beginning with the transfer of an enolpyruvyl moiety from PEP to UDP-GlcNac, catalyzed by MurA (Fig. S1) (12). After reduction of the enolpyruvyl moiety by MurB to create UDP-MurNac, the next five enzymes are involved in adding amino acids that form the pentapeptide (L-Ala, D-Glu, m-Dap, D-Ala, and D-Ala) to the lactyl group, resulting in the final soluble intermediate UDP-MurNac-pepS (Fig. 1A) (5, 13). The first membrane-linked step in PG synthesis begins with the transfer of this sugar nucleotide pentapeptide to the lipid carrier undecaprenyl phosphate (C55-P or UndP). This reaction is catalyzed on the cytoplasmic face of the IM by the integral membrane protein MrαY to generate a monosaccharide–lipid compound, lipid I (Fig. 1A and Fig. S2A).

MurG then catalyzes the addition of a second sugar moiety (UDP-GlcNac), resulting in the final precursor, lipid II (Fig. 1A). The last step of the membrane phase is the flipping of lipid II so that its disaccharide pentapeptide moiety is on the periplasmic face of the membrane. The enzyme “flippase” that effects this transfer has been controversial until very recently. Although FtsW was shown to flip lipid II in vitro (14, 15), several lines of evidence now support Mur if as being the essential lipid II flippase (16–21). The extracellular steps of PG biosynthesis utilize the energy stored in the phosphodiester–murmuramic acid bond of the flipped lipid II and in the D-Ala–D-Ala peptide bond to drive the glycosyltransferase and cross-linking reac-
tions, respectively (13). These steps are carried out by mono- or bifunctional penicillin-binding proteins, and recently, RodA, a member of the SEDS superfamily, was shown to catalyze glycosyltransferase reactions (22–24).

The first Sgl: Protein E from microvirus ϕX174

Famous phage, famous gene

ϕX174 is the founding member and genetic paradigm of the Microviridae, which are nearly as widespread as the Caudovirales (25). It was the first gDNA to be completely sequenced and also to be synthesized in vitro (26, 27). The 10 genes include three that are embedded out-of-frame in essential cistrons (Fig. 1B). One of these embedded genes is E, encoding the Sgl protein in the +1 reading frame of the essential scaffolding gene D (Fig. 1B). E is famous not only for being the first embedded gene to be discovered but also the first gene to be subjected to site-directed mutagenesis (26, 28). More important here is the fact that it is the only DNA virus Sgl, and it was the first Sgl gene for which the lytic mechanism was firmly established. The methods for working out its functional pathway have been replicated for all of the other Sgl systems and thus will be reviewed here in some detail.

Genetics clarifies E function

The E gene was cloned into a medium copy expression plasmid and shown to support lysis after induction (29, 30). Despite this early focus and the availability of the cloned E gene, the lysis mechanism of E remained controversial for 2 decades (3). Early transmission EM studies showed that cells infected with ϕX174 lysed as a result of septal blebs in dividing cells, generating a morphology that was remarkably similar to penicillin-mediated lysis (31, 32). Reproduction of this morphology after induction of the cloned E led to the general model that E interfered with PG biosynthesis (30, 33). However, based on physiological and scanning-EM studies, other groups proposed that E functioned either by the activation of unspecified autolytic functions (34–36) or by the formation of polymeric “transmembrane tunnels” that opened the cytoplasm directly to the medium (37). This profusion of models painted a confusing picture for the mechanism of E lysis, primarily because all lacked genetic evidence.

Mutational and deletion analysis of E revealed that the lytic function requires only the first 34 residues: lytic function was retained without the C-terminal 57 residues, a highly basic, Pro-rich segment, as long it was replaced by a stable cytoplasmic Pro-rich segment, to which E gene function requires only the first 34 residues: lytic function was retained without the C-terminal 57 residues, a highly basic, Pro-rich segment, as long it was replaced by a stable cytoplasmic Pro-rich segment, to which E function was retained without the C-terminal 57 residues, a highly basic, Pro-rich segment, as long it was replaced by a stable cytoplasmic Pro-rich segment, to which E was found to be highly unstable in a slyD background, suggesting that SlyD has a chaperone role, probably in folding the C-terminal domain (CTD), which is rich in Pro residues (42).

To continue pursuit of the E target, bypass suppressor mutations were isolated as rare plaque formers on a slyD lawn (42). These Epos (plates on slyD) were found to be missense alleles at the N terminus of E, each of which resulted in a ~10-fold increase in the biosynthesis rate, thereby compensating for the
proteolytic instability of E. This proved to be a key technological advance. Spontaneous “Eps” host mutants (E-pos sensitivity) that were resistant to induction of the plasmid-borne Epos allele were selected and cross-streaked for sensitivity to the phage. Of ~2000 survivors, all but three retained δX174 sensitivity and presumably were defective in E expression or plasmid copy number. Subsequent genetic mapping and sequencing revealed that the mutations mapped to residues (Pro-172 and Phe-288) in TMD5 and -9 of MraY (see Fig. 3). Labeling experiments with [3H]dAMP showed that E blocked cell wall synthesis ~20 min before lysis, and TLC chromatography revealed depletion of lipid-linked label and the accumulation of UDP-GlcNAc, confirming the inhibition of MraY (43). Based on the mutational data and membrane localization of both proteins, Bernhardt et al. (43, 44) proposed that E interacted with MraY through TMD–TMD interactions that were disrupted in the mutant alleles (Fig. S2B).

In vitro analysis of E-mediated inhibition of MraY

Initial attempts to overexpress full-length E failed due to its inherent lethality, but by doing inductions in the presence of the heterologous MraY from B. subtilis, a His-tagged full-length E protein was purified with a yield of 27 μg of E/150 ml of culture (45). Using this as an immunoblot standard, these workers determined that E was produced at ~500 molecules/cell at the time of lysis, in agreement with estimates from radiolabeling and E-LacZ enzyme assays (38, 46). An assay based on UDP-MurNAc-pentapeptide DNS, a fluorescent analogue of UDP-MurNAc-pentapeptide, and phytol-P, a 20-carbon analogue of C55-P, was developed to determine kinetic parameters in the presence and absence of E. For both substrates, the addition of purified E had no effect on the apparent Km but reduced the Vmax, indicating that E is a noncompetitive inhibitor of MraY (45). This was further supported by the observation that overexpression of either WT or the catalytically inactive allele of mraY, D267N, protects cells from E-mediated lysis, presumably by titrating out E (43, 47).

The E–MraY interaction

Some details of the interaction between E and MraY have been determined by assessing lysis kinetics in bulk culture in the context of mutant E alleles, both from selections and site-directed mutagenesis and both WT and catalytically inactive MraY (43–45, 47). Despite nearly saturating mutagenesis, directed mutagenesis and both WT and catalytically inactive alleles, focused on the TMD of E (49). This collection was generated. However, the efficiency of MraY recovery did not map to a hairpin structure, with one directly above TMD5 (Leu-172) and the other just above TMD9 (Pro-170). Assuming these mutations reflect critical contacts in the E binding site, the spatial arrangement indicates that the N-terminal segment of the TMD of E makes those contacts. This notion is consistent with the results from the Clemens group, who constructed a large set of alanine and leucine substitutions as well as truncation alleles, focused on the TMD of E (49). This collection was analyzed by induction in vivo, monitoring the kinetics of lysis and the accumulation of the E protein in the membrane fraction. Mutations that affected lysis without affecting accumulation mapped to a single face of the TMD, suggesting that this face interacts with the two TMDs of MraY that housed the E-resistant changes. Also, the TMD as a whole appears to be relatively insensitive to changes in its C-terminal segment, which can be replaced by a homopolymer of Leu residues. Moreover, Phe substitutions at Leu-19 and Leu-23, both of which are on the same helix face as the inactivating mutations and one of which (L19F) corresponds to an original Epos mutation, enhance lytic function without enhancing membrane insertion. Using immobilized metal-affinity chromatography and Western blotting, complexes of E and MraY could be demonstrated. However, the efficiency of MraY recovery did not correlate well with lytic function of the E allele used, indicating that simple binding of E to MraY is not sufficient for inhibition. Overall, the details of how E binds MraY and interferes with lipid I synthesis will likely require a structure of the E–MraY complex.

Overview of Sgl genes in ssRNA phages

The Sgl proteins undoubtedly evolved because of the extremely constrained size of the phage genome; compared with the Microviridae, this constraint is even worse for the ssRNA phages (the Leviviridae, or the leviviruses), which have ~4-kb gRNAs and only three core genes (Fig. 1B). Lacking tail
structures, leviviruses exploit retractable pili to initiate infection. By far the best studied leviviruses are MS2 and Qβ, both specific for the canonical F conjugation pilus; many other F-specific leviviruses that are related to these two paradigms have been studied (50–55). However, seven other distinct leviviruses are known, each targeting a different retractable pilus (Table S1) (56–61). The Sgl genes have been identified in eight of the nine distinct leviviruses by showing that, as for E, induction of a plasmid-borne clone is necessary and sufficient for lysis. Importantly, all cause disruption of the cell wall by accessing different cellular targets, suggesting that in each case, a new Sgl was evolved after radiation to a new retractable pilus, no doubt facilitated by the extremely high mutation rate of the RNA-dependent replicase (62). This means that even with the low total genomic database of less than 50,000 bases of unique Leviviridae genomes, there are multiple Sgl systems that might be exploited for probing the biosynthesis and dynamic homeostasis of the cell wall. In the following, the Sgl systems of Qβ, M, and MS2 will be reviewed. The order is not chronological but makes sense in that the targets of the first two Sgl proteins have been identified, whereas the MS2 Sgl system is an enduring mystery.

The “protein antibiotic”: A2 from Qβ

Finding rat mutants

The A2 protein has multiple functions during Qβ infection; it functions in virion assembly (63), is bound to and provides protection for the gRNA against RNase degradation (64), mediates interaction with the F-pilus, and is internalized into the host cytoplasm along with the genomic RNA (63–65). Remarkably, A2 also functions as the Sgl protein. The lytic activity of A2 was first demonstrated in 1983 by Winter and Gold (66), who showed that induction of A2 cloned on a medium-copy plasmid is necessary and sufficient to cause lysis.

To identify the target of A2, the same method was used as for ϕX174 E, selecting for host mutants that survived induction of a plasmid-borne A2 gene, followed by screening survivors by cross-streaking with Qβ phage, and the mutants that passed selection/screen were designated as rat (resistant to A-two) mutants (67). Genetics and sequence analyses revealed a single missense change in murA, L138Q. As with E, the incorporation of $[^3H]$dAP label into PG was blocked at least 20 min before the onset of lysis in cells induced for A2. Biochemical analysis of the sugar nucleotide pool from A2-inhibited cells revealed that UDP-GlcNAc was elevated, confirming MurA as the target.

In vitro inhibition of MurA by purified A2 could not be demonstrated, mainly because overexpressed A2 was insoluble. However, in what seems to be the only instance of using virions for enzyme inhibition, it was shown that the catalytic activity of MurA, but not MurA$^{L138Q}$, in crude extracts could be blocked by the addition of highly purified Qβ virions. Later experiments done with purified MurA and Qβ particles confirmed these results (68). Based on the turnover number of MurA and the number of purified virions needed to block its enzymatic activity, the MurA–A2 dissociation constant was determined to be ~10 nM (67).

The A2–MurA interaction

In addition to in vitro inhibition assays, direct protein–protein interaction between A2–MurA and A2–MurA$^{L138Q}$ was also demonstrated by yeast two-hybrid analysis, with the latter pair displaying a weaker signal, suggesting that the mutant allele weakens A2 binding (68). To probe the interaction of MurA with A2 in vitro, the well-characterized catalytic pathway of MurA was exploited. The MurA reaction is well-ordered, with UDP-GlcNAc binding in the catalytic cleft associated with a dramatic shift from open to closed conformation, which then allows PEP binding and catalysis. Interaction studies were done using a soluble MBP–A2 fusion protein and various forms of MurA, including the original rat allele, MurA$^{L138Q}$, and MurA$^{D305A}$, which is disabled for catalysis but not substrate binding, and with various combinations of the substrates and the suicide inhibitor fosfomycin. The results clearly showed that A2 preferentially binds to the UDP-GlcNAc–liganded, closed form MurA, preventing PEP from binding. Details of the binding surface were obtained by site-directed substitutions of amino acids in the area around Leu-138, yielding a cluster of new rat alleles that blocked Qβ plaque formation and clearly defined an interaction surface surrounding the catalytic loop, including the catalytic domain, CTD, and the catalytic loop (Fig. S3).

A2 differs significantly from the Mat proteins of MS2 and related Leviviridae, especially in the N terminus; deletion analysis confirmed that the lytic function is fully defined in the first 180 residues. To map the interaction domain, A2–por plates on rat suppressor alleles were isolated and mapped to three positions (Leu-28, Asp-52, and Glu-125) in the N-terminal domain (69). However, none of the por alleles were lytic when cloned and induced in the rat1 host. Immunoblot analysis revealed that A2–por mutant levels increased much more rapidly than the parental A2 during infection, resulting in early lysis and reduced yield of progeny virions in the WT host. Inspection of the sequence around the por sites confirmed that the mutations disrupt significant RNA structures that repress translational initiation in the viral RNA, thus bypassing the reduced A2–MurA$^{L138Q}$ affinity by increasing the quantity of the phage protein.

The interaction interface was recently resolved in asymmetric cryo-EM structures of Qβ particles in complex with UDP-GlcNAc–liganded MurA or fosfomycin–liganded MurA (63) (Fig. 3, A and B). The cryo-EM structures validated the interaction interface on MurA inferred from the various rat alleles and also confirmed that the NTD of A2 is in contact with MurA (Fig. 3, C–E).

lys$^M$: New target and settling a debate

The lysis gene (lys$^M$) of phage M has evolved completely embedded in the +1 reading frame of the rep gene, and it encodes a 37-amino acid protein with a single TMD (58). The functional lys$^M$-eGFP fusion suggests an N-out and C-in membrane topology (70). Early insights into the molecular mechanism of lys$^M$ lethality came from the observations that lysis proceeded through septal catastrophes, like A2 and E, suggesting that lys$^M$ might be an inhibitor of cell wall biosynthesis (31,
The identity of MurJ as the molecular target of LysM was revealed in multicyclic suppression experiments, where plasmids carrying murJ in random fragments of *Escherichia coli* genome-suppressed *lys* lethality. Furthermore, isolation of nine spontaneous *lys*–resistant mutants that mapped to two of the 14 TMDs of MurJ suggested a possible interaction interface and a plausible mechanism (Fig. 4). Additionally, it was shown that LysM was specific to MurJ, and the cells can be rescued from LysM lethality by the expression of heterologous lipid II flippase Amj from *B. subtilis*.

To address the conformational state in which LysM binds to MurJ, a substituted-cysteine accessibility method (SCAM) was used (70, 71). In the presence of LysM, SCAM analysis showed that five TMD positions showed altered SCAM patterns, which suggested that LysM binding locks MurJ into one of the two conformations proposed to constitute the lipid II–flipping cycle. The SCAM labeling pattern is consistent with MurJ being locked in a “periplasmic open” conformation, which would lead to an accumulation of lipid-linked PG precursors in the inner leaflet of IM and a corresponding decrease on the periplasmic side. Both predictions were confirmed by in vivo flippase assays, strongly indicating that LysM blocks MurJ’s activity (70). Moreover, in the presence of the LysM–resistant *murJ* alleles, the precursor levels were restored to normal. The fact that LysM only targets MurJ and causes lysis strongly suggests that MurJ is the only active lipid II flippase in *E. coli*. However, the interpretation of the LysM–resistant *murJ* alleles and the interaction interface would greatly benefit from the structure of the MurJ–LysM complex.

**MS2 lysis: To L and back**

**L: The first autolysin?**

The *L* gene was not recognized as a gene in MS2 until the isolation of a plaque-forming defective nonsense mutant that belonged to a complementation group distinct from *mat, coat*, and *rep* (72). Subsequent radioactive labeling experiments established *L* as the fourth gene of MS2, encoding a 75-amino acid polypeptide (73). As had been done with *E* and *A*₂, a plasmid clone of *L* was shown to cause lysis after induction, and the *L* protein was shown to be associated with the membrane fraction (73, 74). Opposite to *E*, it is the 39-residue CTD of *L* that accounts for membrane localization and lytic function, with the N-terminal 36 highly basic residues shown to be dispensable for lysis (76). This clearly differentiates *L* action from the *E*, *A*₂, and LysM Sgl proteins, all of which cause cessation of cell wall synthesis by interrupting the supply of lipid II to the PG machinery (43, 67, 70). To these workers, the most significant finding was that induced *L* lysis was severely compromised in acidic (pH 5.5) conditions, despite normal accumulation of *L*, raising a compelling analogy to penicillin-induced autolysis, which is also blocked under these conditions (77). This led to a general model in which *L* induced lysis by inducing autolysis, although the precise definition thereof was not provided. In immuno-EM experiments, *L* was shown to preferentially localize to apparent zones of adhesion between the IM and OM (78). This association with adhesion zones was emphasized by the fact that *L* lysis is also compromised in cells that lacked the periplasmic osmoprotectant membrane-derived oligosaccharide (79). These cells were shown to have many fewer adhesion zones and a much wider periplasm, and in this case, *L* appeared to be subject to degradation. Furthermore, a synthetic polypeptide corresponding to the C-terminal 25 amino acids of *L* was shown to permeabilize both liposomes and inverted membrane vesicles, leading the authors to invoke induction of autolysis after membrane permeabilization (80). However, these experiments lacked a negative polypeptide control, and the experiments were done at peptide/vesicle ratios in excess of 1000; moreover, permeabilization and depolarization does not result in rapid autolysis in *E. coli*, so the physiological relevance of these experiments is questionable.

**Back to L: Genetic and molecular analysis**

The consignment of MS2 *L* to the role of phage-encoded autolysin seemed to end further interest in its function, despite the likelihood that a critical component of cell wall homeostasis was targeted. However, over the next decades, a few new Levi-
viridae were characterized, many of which shared the same genetic architecture as MS2, despite no significant nucleotide sequence similarity (60, 61) (Table S1). This included not only new leviviruses specific for the F pilus but also against the conjugational pilus of several R-factor plasmids and the polar pilus of Pseudomonas (Fig. S5) (57, 58, 60, 61). We noticed that the L proteins, although unrelated in terms of sequence, shared an apparent domain organization with L: domain 1, N-terminal, highly charged; domain 2, very hydrophobic and lacking charged residues; domain 3, a central Leu-Ser dipeptide; and domain 4, a variable CTD (81). The conserved architecture suggested that L-like Sgl systems were widespread among Gram-negative bacteria were all targeting the same host function.

Two genetics-based approaches were mounted, the first aimed at identifying host factors, using, as before, inducible plasmid-based clones of L (81, 82). To avoid mutations that reduced the copy number or L transcription, a blue/white reporter plasmid was constructed, and from hundreds of colonies surviving L induction from this construct, two blue colonies were identified and designated as ill (insensitive to L lysis) mutants (82). Surprisingly, the ill mutations mapped to dnaJ, which encodes a widely conserved chaperone involved in the heat shock response (83). Analysis revealed that, in both, a P330Q missense change in dnaJ accounted for the Ill phenotype and abolished MS2 plaque formation, with both phage and survival phenotypes recessive. The Pro-330 residue is the most conserved residue in the CTD of DnaJ, which is clearly a conserved segment, although its function is unclear. The P330Q change was found to preserve the heat shock function of DnaJ but abrogates the ability of DnaJ to form complexes with L. The L suppressors, designated as Lodj (overcomes dnaJ) alleles, were isolated as mutants that allowed lysis in dnaJP330Q background; these proved to be deletions of the dispensable NTD of L. Isogenic inductions of the parental and Lodj alleles revealed that lysis was much earlier with the truncations. These results led to a model in which the NTD of L has a regulatory, lysis-delay function that blocks the interaction of L with its target; in this model, DnaJ is required for relief of this steric block (Fig. S6).

To identify key functional elements of L itself, a nearly saturated mutational analysis of L generated a collection of 103 alleles with single codon changes conferring absolute lysis defects (Fig. S4) (81). The mutational distribution validated the proposed four domain structure of the L Sgl proteins. Domain 1, comprising the dispensable, highly basic region, gave rise to only one nonlytic allele (Q33H). Domain 3 containing the LS dipeptide motif and the adjacent segments of domains 2 and 4 had the most missense changes conferring nonlytic character. All of the missense alleles tested were genetically recessive and generated membrane-associated products of parental size. In addition, several of the inactivating missense changes (i.e., L44V, F47L, F47Y, S49T, F51L, and L56F) were conservative, suggesting that the L protein makes specific heterotypic protein–protein contacts in the membrane.

Taken together, the isolation of dnaJP330Q and mutational analysis of L both suggest that L targets a host membrane protein; that the interaction is through the mutationally sensitive residues in domains 2, 3, and 4; and that, like SlyD and -E, a host chaperone is involved in regulating L function. Obviously, further investigation into the host factors involved in L lysis is needed to understand the mechanistic details of L function.


**JBC REVIEWS:** Phage Sgl proteins target PG biosynthesis

**JBC REVIEWS:** Phage Sgl proteins target PG biosynthesis
What’s next?

The premise of this review was that the study of the Sgl systems of small lytic phages would be interesting and lead to a better understanding of the bacterial cell wall biosynthesis and homeostasis. To summarize what has been discussed, four Sgl systems have been studied in depth, one from the micovirus φX174 and three from the Leviviridae. In three cases, the Sgl proteins turned out to be “protein antibiotics,” specific inhibitors of different enzymes of the highly conserved PG biosynthesis pathway (3, 70, 84). Consideration of the available genetic and biochemical data has already increased our understanding of how these important enzymes function and, in the case of the LysM, settled the identification of the flippase that exports lipid II to the periplasm. The next level of understanding will come from detailed structural information about the Sgl–enzyme inhibition complexes. The fourth case, the L protein of MS2, has not yet been fully characterized but appears to be the prototype of an Sgl type that has evolved multiple times in Leviviridae, infecting a wide range of Gram-negative bacteria (81).

Genetic and biochemical evidence was cited showing that L does not inhibit any of the steps that lead to externalized lipid II and its incorporation into existing PG and suggests that it targets a host protein. There is no conceivable answer to the L target mystery that would not be important, possibly identifying conserved proteins that are essential for proper coordination or localization of cell wall synthesis machineries or are involved in the control of powerful autolytic enzymes.

Even with the L story still incomplete, this seems like a pretty good haul of information from the study of four small genes, starting with very “low-tech,” old-fashioned and simple genetic selections. The shocking thing is that this wealth of molecular information is derived from the study of only nine distinct Leviviridae (Table S1), comprising a total of <50 kb of total genomic information. Despite the low number, these nine phages segregate into five different phage types based on where the Sgl evolved. Listed from 5’ to 3’ of the gRNA, they are as follows: AP205 (5’ of mat), Qβ (mat = A2), MS2 (overlapping end of coat and beginning of rep), phiCb5 (middle of rep), and M (near 3’ end of rep) (Fig. 2B). The diverse location of the Sgl genes suggests that they have evolved more than once and probably as a late addition to the genome after speciation to different pili or hosts (58, 61). (This includes all of the L-like genes; none of the 6 L-type Sgls has any detectable sequence identity other than the LS dipeptide sequence.) Given the diversity of 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 known step in PG biosynthesis and, if L is any indicator, possibly uncharacterized components critical for dynamic cell wall homeostasis.

Taken together, it seems obvious that it would be useful to identify new Sgl genes, and the old-fashioned “phage hunt” is a reliable approach. RNA phage hunts have so far been done for five conjugational pili, resulting in two protein antibiotic Sgls (A2 and LysM) and four unrelated L-type Sgls (LMS2, LGAL, LCl, and LPR1). Only three nonconjugational pili (Caulobacter, Acinetobacter, and Pseudomonas) have been targeted, resulting in two L-type Sgls and one, Lys of Caulobacter phage phiCBS, that does not have an L-type domain structure but does have a single N-terminal TMD, resembling both E and LysM. Considering the existence of many more retractable pili systems, there is a clear rationale to conduct RNA phage hunts in many other systems with retractable pili, especially in pathogenic bacteria.

Metagenomics is also having an impact. A recent survey of publicly available RNA-inclusive metagenomes and RNA virome studies of invertebrate species led to the identification of ~200 new ssRNA phage genomes (85, 86). Although most of the new leviviral genomes are partial, ~80 are either complete or nearly so, with all three core genes annotated. Only one (AVEO17) of these ~200 genomes had an annotated Sgl gene, being a close relative (38% sequence identity) of MS2 L (85). Given their small size, predilection for being embedded in the core genes, and extreme sequence diversity at the protein level, finding Sgl candidates in these genomes poses unique challenges to the traditional gene annotation tools. Moreover, currently, there is no direct way to sort out these leviviral genomes to a particular bacterial host or, more specifically, to a particular retractable pilus. Nevertheless, the promise of more intriguing Sgl proteins targeting novel components of the bacterial cell wall machinery surely makes our current effort, which involves identifying potential Sgl ORFs and characterizing them one by one for the ability to support lysis after induction of synthetic clones, worth doing.

Acknowledgments—The following figures and figure legends were reproduced verbatim from published references: Figs. 1, 3, and 4 and Figs. S3–S6. We thank past and current members of the Young laboratory for advice. We also thank Ing-Nang Wang, Kaspar Tars, and Rene Olsthoorn for providing ssRNA phage resources and expertise. The clerical assistance of Daisy Wilbert is highly appreciated.

References