# High-throughput suppressor screens unveil functional convergence of single-gene lysis proteins

Benjamin A. Adler[1,2,3], Karthik Chamakura[4], Heloise Carion[2], Jonathan Krog[2], Adam M. Deutschbauer[5], Ryland F Young[4], Vivek K. Mutalik[3,5*], Adam P. Arkin[2,3,5*]

1. The UC Berkeley-UCSF Graduate Program in Bioengineering, Berkeley, CA, USA
2. Department of Bioengineering, University of California, Berkeley, Berkeley, California, USA
3. Innovative Genomics Institute, University of California, Berkeley, Berkeley, United States
4. Department of Biochemistry and Biophysics, Center of phage Technology, Texas A&M University, College Station,TX, United States
5. Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

*vkmutalik@lbl.gov (VKM); aparkin@lbl.gov (APA)

**Abstract:**

In contrast to canonical dsDNA phages, lysis in ssRNA *Fiersviridae* and ssDNA *Microviridae* phages is encoded by a minimal single gene (*sgl*), to meet the size constraints of some of the smallest genomes on earth. To achieve lysis, Sgl proteins exploit evolutionary “weak spots” in the bacterial cell wall by inhibiting specific steps in cell wall synthesis. Although a handful of these proteins have been characterized, the potentially diverse “weak spots” targeted by most Sgls remains enigmatic. Here, we repurpose Dub-seq for genome-wide assessment of host suppressors of Sgl activity and apply to eight diverse Sgls awaiting molecular target characterization. In addition to known molecular mechanisms, we discover a complex network of suppressors across the spectrum of genes involved in cell wall biogenesis. We highlight the first case of Sgl convergent evolution between unrelated proteins by determining the molecular target of $Sgl^{PP7}$ as MurJ, suggesting the existence of universal “weak spots” in the bacterial cell wall.

## Introduction

Lysis of the bacterial host is the last step in the bacteriophage (phage) life cycle, determined by the lytic program encoded on the phage's genome. In canonical double-strand DNA (dsDNA) phages, lysis is mediated by multiprotein systems that disrupt the cytoplasmic/inner membrane, degrade the peptidoglycan/cell wall, compromise the outer membrane, and regulate the lytic process [1]. In contrast, single strand RNA (ssRNA) phages of the family *Fiersviridae* (e.g., MS2 and Qβ) and single-strand DNA (ssDNA) phages of the subfamily *Bullavirinae* (e.g., φX174) use the product of a single gene (Sgl) to carry out host lysis [2,3]. Decades of research on a handful of ssRNA and ssDNA phage has highlighted the sequence diversity, genomic context and target pathway specificity of Sgls for eliciting a successful host lysis program. Some Sgls seem to be potent inhibitors of bacterial cell wall biogenesis, analogous to a class of mechanisms of several major antibiotics classes, earning them the name "protein antibiotics (PA)" [4–6]. Greater capacity for screening Sgls offers new opportunities for finding and characterizing natural "weak spots" in the bacterial cell wall, offering new opportunities for antibiotic development.

Characterization of the PA-type Sgls has highlighted the importance of peptidoglycan (PG) biosynthesis for bacterial growth and survival. For example, protein $Sgl^{\varphi X174}$ inhibits MraY, blocking the synthesis of Lipid I, the first lipid-linked precursor in the PG synthesis pathway [7,8]. $Sgl^{QB}$ inhibits the first step of cytoplasmic PG precursor biosynthesis through MurA inhibition [9]. More recently, $Sgl^{M}$ was found to inhibit the essential lipid II flippase, MurJ [6]. Thus all three of these Sgls block the primary PG biosynthesis pathway. All three cause lysis marked by septal catastrophes, indistinguishable from the lytic morphology associated with penicillin and other PG poisons [6,7,9]. In contrast, the lytic mechanism of some Sgls, including $Sgl^{MS2}$ - the first ssRNA lysis gene identified - has remained elusive [10,11]. Early studies have indicated that, in contrast to the PA-type Sgls, $Sgl^{MS2}$ does not block net precursor incorporation into PG but imparts deleterious changes to the fine structure of the murein [11–14].

Identifying and characterizing Sgl mechanisms of action remains a key bottleneck. Early studies on model ssDNA and ssRNA phages showed that native *sgl* genes are hard to identify through genomic approaches [12,15]. Sgls are typically small, embedded within other coding regions, and bear little sequence similarity to each other [3,16–18]. Linking a gene to its mechanism of action has most often been accomplished using classical genetic selections [7,9]. Most recently, a multicopy-suppression screen was used to identify the target of $Sgl^{M}$. In this approach, random fragments of the host-genome are complemented on plasmids to screen for proteins able to suppress lysis during infection or co-expression with $Sgl^{M}$ [6]. These types of approaches were manageable when the

pace of new ssRNA phage discovery was slow. As of 2016, the total genomic space for ssRNA phages was less than 100kb, of which about half was dedicated to close relatives of the classic phages MS2 and Qβ. However, in the last five years, metagenomic and metatranscriptomic analyses have identified more than 10,000 new ssRNA phage genomes [19–21]. Given these dramatic advances, unbiased and scalable screening platforms are needed to realize the promise and diversity of candidate Sgls [18–21].

We recently reported a dual DNA barcoded overexpression library approach (Dub-seq) that generates shotgun cloning of randomly sheared DNA fragments of host genome on a replicable plasmid and uses next-generation sequencing to map the barcodes onto cloned genomic regions [22]. Once characterized, such a library can be used innumerable times for carrying out competitive fitness assays to uncover traits and phenotypes encoded on the cloned genomic regions. This decoupling of competitive fitness assays from library characterization to assess the phenotypic importance of the genes contained on those fragments enables parallelized gain-of-function screens. We have demonstrated the utility, scalability and barcode standardization of fitness assays by studying the tolerance phenotypes against diverse antibiotics, stressors and metals, and most recently used to characterize genetic barriers in phage-host interactions at a scale that was not possible before [22,23]. Potentially, Dub-seq could scalably identify host genes and pathways targeted by phage encoded Sgl proteins.

In this study, we repurpose Dub-seq for genome-wide assessment of host suppressors of Sgl activity and apply to eight diverse Sgls awaiting molecular target characterization. We establish the screening platform by recapitulating the known molecular target of $Sgl^{M}$ and discover several new host targets and mediators for others that are general across diverse Sgls. Our high-throughput genetic screen identifies MurJ as the target for $Sgl^{PP7}$, uncovers several inner membrane proteases as Sgl-specific suppressors and provides an insight into disparate putative molecular targets of $Sgl^{MS2}$ involved in cell-wall biogenesis and PG recycling. We validate several of these findings including target specificity of $Sgl^{PP7}$ and its interaction with MurJ in mutation mapping experiments. Finally, we discuss the implications and extensibility of our approach in characterizing hundreds of putative Sgl in genomic databases.

## Results

### *Devising rapid Sgl suppressor identification screens*

A common outcome of expressing an Sgl protein from a plasmid is the intense toxicity associated with host lysis, establishing a selection pressure leading to plasmid-knockdown, expression defects, or mutations in the lysis protein itself [11,14]. Transcriptional fusion of reporters such as *lacZ*$_{\alpha}$ to the *sgl* of interest has provided a

simple readout for reducing false positives in both spontaneous and multi-copy suppressor selection/screens, but this practice ultimately constrains analysis to a very limited number of suppressor mutants [6,11,14]. We hypothesized that this inherent limitation in the Sgl-suppressor screens could be mitigated by expressing a single toxic Sgl in the presence of a barcoded shotgun expression library of the host (Dub-seq library) which enables the quick identification of survivors using barcode sequencing as a readout (Figs. 1a,b). In contrast to traditional overexpression libraries, Dub-seq libraries enable economical and high-throughput assessment of gene overexpression or dosage effects that might alleviate/suppress the toxicity phenotype imparted by Sgl expression. With careful analysis, we could decouple fitness effects conferred by single mutations due to Sgl toxicity from positive fitness conferred by having more copies of a functional allele (i.e., a multicopy suppressor).

As a proof-of-principle, we adapted our Dub-seq platform for screening suppressors against the expression lethality of $Sgl^M$, an inhibitor of the lipid II flippase, MurJ. (Figs. 1c,d) [6]. Briefly, we first cloned *sgl*$^M$ into a standard, low copy plasmid under an anhydrotetracycline (aTc)-inducible promoter and verified *sgl*$^M$ toxicity induced a growth defect in *E. coli* K-12 (Fig. 1d, Methods). We then transformed a previously characterized *E. coli* Dub-seq plasmid library (pFAB5516), consisting of *E.coli* genomic DNA fragments cloned between two 20bp random DNA barcodes (for more details, see: [22]). This process resulted in a library of 17,007 unique members, (BA1320L), with each strain harboring an inducible *sgl*$^M$ vector and a unique member of the pFAB5516 library (as in Fig. 1a) (Supplementary Table 1). We then subjected this BA1320L library to a range of $Sgl^M$ induction levels in liquid culture and collected survivors. From samples collected before and at the end of the experiments, plasmid DNA was recovered and subjected to BarSeq PCR and sequencing on a Hiseq4000 platform (Methods). We computed the gene fitness scores from protein-coding genes and RNA genes as reported earlier (Methods) [22]. We classified genes with a fitness score > 2 as high confidence hits if they have sufficient read coverage (>25 reads/barcode for both t=0 and the experiment) and these fitness effects were consistent across multiple fragments that cover the genes and across replicate experiments (Methods). We observed low induction of $Sgl^M$ yielded reproducible data (N=2 fitness experiments; Supplementary Fig. 5). As expected, *murJ* emerged as the only consistently enriched gene in our screen, covered by multiple fragments (Figs. 1e,2, Supplementary Fig. 5). These experiments indicated that the Sgl-dependent growth defect coupled with Dub-seq suppressor screen could correctly identify host factors that, when over-expressed, overcome the toxicity of the Sgl protein and thus could map the Sgl target to host pathways.

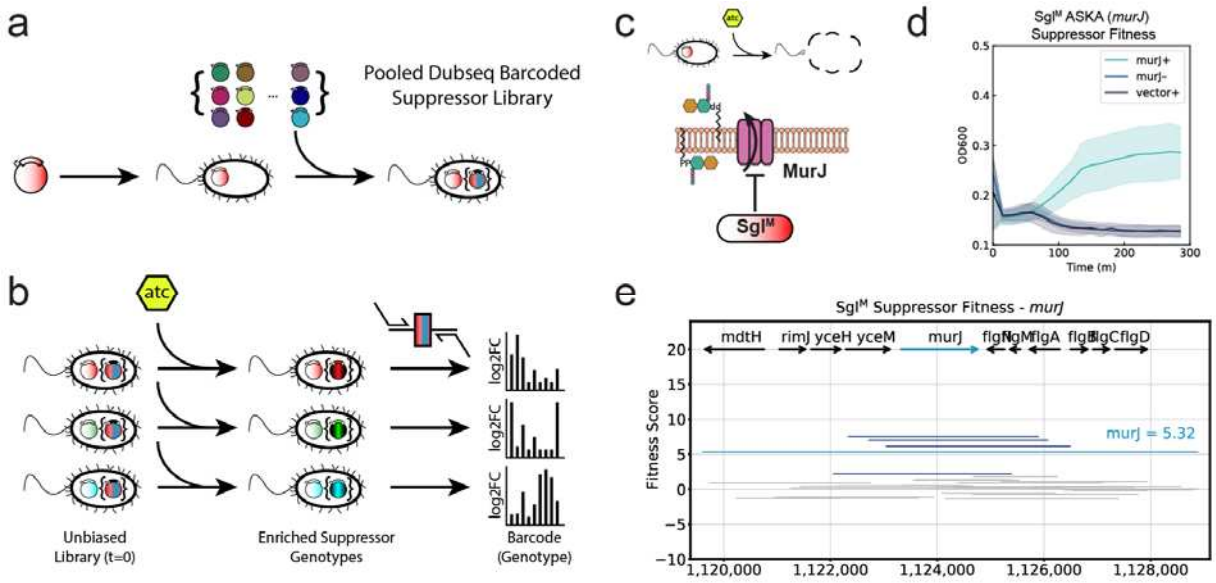


**Fig. 1: Genome-wide screen to identify host suppressors of phage-encoded single gene lysis systems. (a)** Cartoon description of each suppressor library. A toxin gene (in this paper, an *sgl* gene) is cloned into an atc-inducible vector (red gradient vector). A previously characterized Dub-seq library containing barcoded, random genomic fragments from BW25113 (pFAB5516) is transformed into the strain carrying the cloned *sgl* gene [22], creating a pooled Dub-seq suppressor library consisting of tens of thousands of independent transformants. This is schematically represented by a strain containing the toxic gene and one member of the pFAB5516 library (rainbow plasmid). **(b)** To assay for suppressor mutants, we performed experiments similar to barseq assays previously described [22,24,25]. Because the source Dub-seq library employed in this study was previously characterized, no additional library characterization beyond barseq was needed. In brief, the suppressor library from (A) or a variant containing a different *sgl* gene (blue or green gradient vector, left) was grown to OD600 ~1.0, where a Time=0 sample, representing a comparatively unbiased distribution of genomic fragments, was collected. We then subjected the cultures to different concentrations of aTc, biasing the library towards suppressing fragments (dark-gradient plasmids). Strains were tracked by quantifying the abundance of DNA barcodes associated with each strain by Illumina sequencing. Sgl-specific gene fitness profiles were calculated by taking the log2-fold-change of barcode abundances post- (t) to pre- (t=0) induction of *sgl*. High fitness scores indicate that genetic content of genome fragments confers fitness against *sgl* expression. **(c)** Schematic representation of the results anticipated based on Chamakura et al., 2017, $Sgl^M$, binds and inhibits essential lipid II flippase, MurJ, leading to lysis [6]. **(d)** Growth curves show heterologous expression of wildtype MurJ can suppress $Sgl^M$ lytic activity. Teal represents $sgl^M$ and *murJ* co-overexpression using 16.1 ng/µL aTc and 25µM IPTG respectively. Blue represents $sgl^M$ expression in the absence of *murJ* induction using 16.1 ng/µL aTc and uninduced IPTG. Black represents $sgl^M$ expression in the presence of an empty ASKA vector using 16.1 ng/µL aTc and 25µM IPTG. All growth curves represent 3 biological replicates (i.e. 3 independent experiments across 3 days). **(e)** Representative experiment for our suppressor screening approach against $Sgl^M$. Multiple barcodes representing fragments containing MurJ were specifically enriched in our screens.

### *Extending Dub-seq Suppressor Screens to Additional sgl Genes*

To demonstrate the scalability of our approach to other Sgls, we extended our suppressor screen library approach to 8 additional, diverse Sgl proteins sourced from an earlier study for which molecular target information is limited/non-existent (Fig. 2,

Supplementary Table 1) [11]. These Sgl candidates included $Sgl^{\alpha 3}$ and $Sgl^{St1}$ from *Bullavirinae* phages that share sequence similarity to $Sgl^{\varphi X174}$ (Figs. 2a,c) and $Sgl^{AP205}$, $Sgl^{MS2}$, $Sgl^{Hgal1}$, $Sgl^{PP7}$, $Sgl^{Ku1}$, and $Sgl^{PRR1}$ from *Fiersviridae* phages that bear little sequence-identity to each other [3] (Fig. 2d). Except for $Sgl^{MS2}$, studies addressing the molecular target of these Sgl proteins have not been reported. In total, we performed 20 genome-wide suppressor screens, collected suppressor candidates, and processed BarSeq PCR samples for deep sequencing (experimental and library overviews are described in Extended Data 1 and Extended Data 2 respectively). After careful curation of the dataset for sufficient read coverage and consistency, we identified 133 high-confidence hits across 7 suppressor Dub-seq screens, with nearly every Sgl experiment yielding a suppressor phenotype. Some of these suppressor candidates seem to provide fitness benefits across multiple Sgl protein toxicity effects. Specifically, 14 genes-- *dacC*, *galE*, *micF*, *mltF*, *mpl*, *rbsR*, *rnd*, *rraB*, *rstA*, *treA*, *treB*, *sdiA*, *waaQ*, with another Y-gene of unknown function (*ytfP*)-- showed high-confidence fitness across more than 3 Sgl proteins. These common hits broadly encode diverse functions such as PG turnover or recycling (*dacC*, *mltF*, *yceG*, *mpl*), cell division regulation (*sdiA*), transcript regulation (*rnd*, *rraB*), carbon transport and catabolism (*rbsR*, *treA*, *treB*) and tolerance to envelope stress (*micF*, *galE*, *rstA*, *sdiA*, waaQ, *treA*, *treB*), implicating their importance in membrane homeostasis and role in compensating for Sgl-induced toxicity. Some of these experiments identified dozens of high-confidence suppressors per Sgl protein, while some had only a couple of hits as seen with $Sgl^{M}$. (Figs. 1, 2). For example, $sgl^{\alpha 3}$, $sgl^{St1}$, $sgl^{PRR1}$, $sgl^{Hgal1}$, and $sgl^{MS2}$ appeared to have a diversity of suppressors emerge during our screens, while sgls $sgl^{PP7}$, $sgl^{AP205}$, and $sgl^{M}$ revealed relatively few high-confidence suppressor hits, indicating a combination of specific and general genetic routes to suppress Sgl lysis function.

We wondered whether Sgl-homology-based molecular target prediction is possible for Sgl proteins. In particular, the *Microviridae*-derived candidates $Sgl^{\alpha 3}$ and $Sgl^{St1}$ share substantial sequence similarity to each other and to MraY-targeting $Sgl^{\varphi X174}$ (Fig. 2c). However, only $Sgl^{St1}$ showed enrichment of *mraY*-encoding fragments, indicative of MraY-mediated suppression (Fig. 2e). While we find it unlikely for $Sgl^{\alpha 3}$ to not target MraY, it is possible that the more acute toxicity of $Sgl^{\alpha 3}$ overwhelmed the toxicity-suppressive effects of MraY expression from the Dub-seq library. Potentially mutants in *mraY*, instead of increased copy number conferred by Dub-seq expression, are necessary to overcome $Sgl^{\alpha 3}$ toxicity similar to prior work on $Sgl^{\varphi X174}$ [7,8]. Additionally, Sgl proteins $Sgl^{PP7}$ and $Sgl^{M}$ shared little sequence similarity with each other (15.6% sequence identity, MUSCLE BLOSUM62) but yielded high scoring *murJ*-containing

fragments (Figs. 2d,e). This indicated MurJ as a potential target for Sgl$^{PP7}$ lethality. Previously Sgl$^{PP7}$ was classified with other Sgls as a having homologous function with Sgl$^{MS2}$, the original Sgl, based on the presence of an N-terminal basic domain and an intervening hydrophobic domain, followed by a Leu-Ser dipeptide motif [14]. Below, we confirm that, as for Sgl$^{M}$, MurJ is the molecular target of Sgl$^{PP7}$.

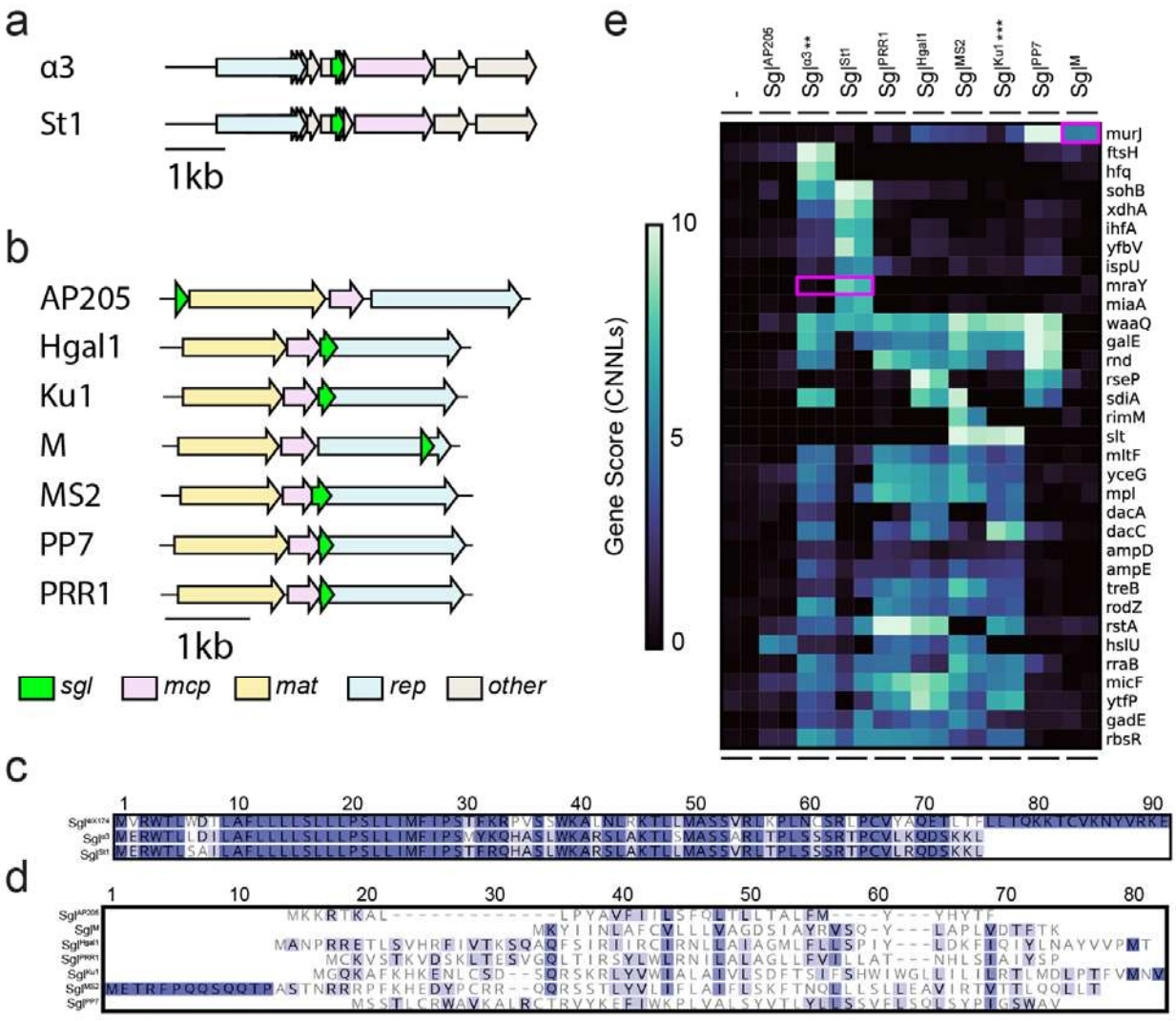


**Fig. 2: Sgl diversity in genomic context, sequence identity, and suppressor genotypes. (a)** All *Microviridae*- and **(b)** *Fiersviridae*-derived Sgls investigated in this study are shown within their native genomic context. With the exception of Sgl$^{AP205}$, all lysis genes (green) occur in sequences overlapping with 1 or more additional genes. **(c)** Sequence alignment of *Microviridae*-derived lysis proteins investigated here. The well-characterized lysis protein Sgl$^{\varphi X174}$ is shown for reference. Lysis proteins derived from phages Sgl$^{\alpha 3}$ and Sgl$^{St1}$ bear substantial sequence similarity to Sgl$^{\varphi X174}$. **(d)** Sequence alignment of *Fiersviridae* lysis proteins investigated here. Unlike those from Microviridae, *Fiersviridae* lysis proteins bear little resemblance to each other. In both (c) and (d), sequence alignments were performed with MUSCLE using a BLOSUM62 matrix [26] and shaded with increasing sequence similarity. **(e)** Multicopy suppressors of lysis proteins as identified through high throughput gain of function screening. A curated

selection of top fit genes are shown for visualization purposes. Multicopy suppressors identified- or inferred from prior work are boxed in pink. All data shown are from *sgl* induction at 15.6ng/µL atc unless noted otherwise (see ** and ***).

* $Sgl^{\varphi X174}$ is shown as a reference but not screened in this study. ** $sgl^{\alpha 3}$ experiments performed at 7.8ng/µL atc. *** $sgl^{Ku1}$ experiments performed at 3.9 ng/µL atc.

## Systems-Level Genotypic Profiles Yields Clues to Additional Mechanisms to overcome Sgl-induced lysis

In contrast to $Sgl^{M}$, $Sgl^{St1}$ and $Sgl^{PP7}$, where each of the strongest suppressor hits (*murJ* and *mraY*) played a direct role in PG biosynthesis and maturation, the top scoring candidate suppressors against $Sgl^{MS2}$, $Sgl^{PRR1}$, $Sgl^{Hgal1}$, $Sgl^{KU1}$ and $Sgl^{\alpha 3}$ encoded functions critical to PG turnover and recycling. Potentially, these suppressors encode capacity to compensate for Lipid-I starvation. The top scoring suppressor for $Sgl^{KU1}$ and well-studied $Sgl^{MS2}$ corresponded to *slt* (Fig. 2e). Its product, Slt, alongside 6 other lytic transglycosylases (LTs) (MltA, MltB, MltC, MltD, MltE (EmtA), MltF, and MltG (YceG)), cleave glycosidic PG bonds to facilitate cell wall remodeling and recycling, shown to be important during growth [5,27,28]. In addition to Slt, fragments encoding LTs such as MltF and YceG, as well as fragments encoding Mpl, a murein peptide ligase catalyzing the last step in PG precursor recycling, provided tolerance to lytic effects of multiple Sgls (Figs. 2e, 3).

Additional ties to cell wall structure and integrity emerged as major suppressor classes. Enriched Dub-seq fragments encoding penicillin binding proteins (PBPs) such as DacC (PBP6) and DacA (PBP5), the antibiotic stress potentiator AmpE, and cytoskeleton proteins such as RodZ and its partner MreB provided tolerance against more than 3 Sgls. These cross-fitness patterns highlight the complex network of PG biosynthesis and cell-wall maturation processes through interaction with PBPs and cytoskeleton proteins (Fig. 3c) [5,29,30]. The emergence of cell-elongation components alongside *slt*, *mltF*, and *mltG* as suppressors is a curious new similarity between Sgls and the potentiating effects of cell-wall recycling of specific beta-lactam antibiotics, warranting further investigation [25,31–33].

In contrast, multiple inner-membrane proteases conferred high-confidence *sgl*-specific suppressor phenotypes (Fig. 3a). For example, multiple genomic fragments encoding inner-membrane protease *sohB* appeared specifically enriched against both overexpression of $sgl^{\alpha 3}$ and $sgl^{St1}$ while *ftsH* appeared to provide fitness solely against

*sgl*$^{\alpha 3}$. Dub-seq fragments encoding the protease *rseP* suppressed lethality of Sgl$^{PP7}$ and Sgl$^{Hgal1}$ while protease complex HslVU[1] encoding fragments gave fitness against Sgl$^{AP205}$, Sgl$^{PRR1}$, and Sgl$^{MS2}$ (Fig. 3a). Most Sgl proteins presumably get localized to the inner membrane, where it is conceivable that the increased presence of proteases may degrade the Sgl proteins [2]. Alternatively, suppressor protease expression may impact the abundance, availability, or phenotypic consequence of the native Sgl target. While some of the proteolytic targets of RseP, HslVU, FtsH, and SohB are known, evidence suggests that there may be additional inner membrane substrates to be discovered [34–36].

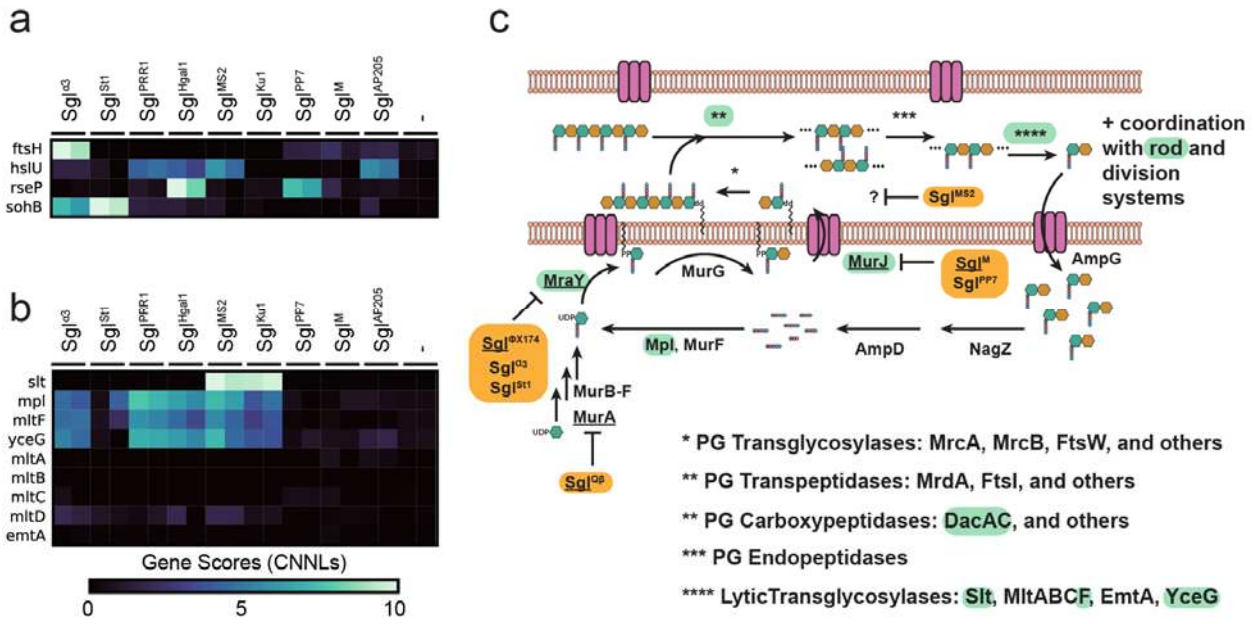


**Fig. 3: Systems-Level Suppressor Phenotypes Observed in this Study. (a)** Heatmap of inner-membrane proteases identified during Dub-seq screening against Sgl toxicity. **(b)** Heatmap of all known lytic transglycosylases in *E.coli*. Genes for the lytic transglycosylases *mltF* and *yceG (mltG)* and the murein peptide recycling protein, *mpl* conferred broad fitness to multiple lysis proteins. **(c)** Schematic representation of PG maturation and recycling in *E.coli* adapted from Johnson et al. [27] with a focus on genetic interactions observed in the lysis protein literature and this study. Highlighted in teal are genes that emerged as repeatable hits in our screens. Underlined are proteins previously implicated in Sgl function, shown with their inhibitors highlighted in orange. In brief, the PG precursor Lipid II is synthesized through the serial activity of enzymes MurA-F, MraY, and MurG, and then translocated to the periplasm via the flippase MurJ. The disaccharide oligopeptide moiety of Lipid II is then added to the PG, crosslinked, and tailored with other PG polymers through coordinated activity of PG-transglycosylases (*), PG-transpeptidases (**), and PG-carboxypeptidases (**) to form a mature PG. Mature PG is broken down into recyclable monomers through endopeptidase (***) and lytic transglycosylase (****) activities. Such monomers can be recycled through AmpG import, further breakdown by NagZ and AmpD (and other

[1] Only fragments containing the HslVU operon were enriched. Fragments containing either *hslV* or *hslU* alone remained unenriched. For more detail, see Supplementary Figs. 2a, 6g, 8c.

enzymes not shown), and re-ligation to the substrate for MurF by Mpl. Additional regulation on PG maturation is imposed by the rod and division networks and discussed in detail elsewhere [5].

## Confirming the suppression phenotypes observed in genetic screens

To validate some of the top Sgl-specific suppressor candidates observed in our genome-wide screens, we transferred plasmids from the ASKA collection into *E. coli* cells expressing individual Sgl candidates and studied their lysis suppression activity (Fig. 4, Methods) [37]. As *sgl* genes and cognate suppressors are encoded on compatible plasmid systems and are induced by aTc and IPTG respectively, we can monitor Sgl-induced lysis and its rescue by target gene expression at different induction levels. In particular, we focused on the top hit for Sgl$^{PP7}$, *murJ*, already shown to be the target of Sgl$^{M}$ [6]. We also investigated a pair of Sgl-specific protease suppressors: RseP with Sgl$^{PP7}$ and FtsH with Sgl$^{\alpha 3}$. These coexpression of *sgl* and its target suppressor gene experiments confirmed that induction of individual suppressor genes at elevated levels can overcome the growth inhibition induced by Sgls. MurJ and RseP for Sgl$^{PP7}$ and FtsH for Sgl$^{\alpha 3}$ all showed suppressor activity in agreement with fitness phenotypes observed in our genetic screens (Figs. 4a-f).

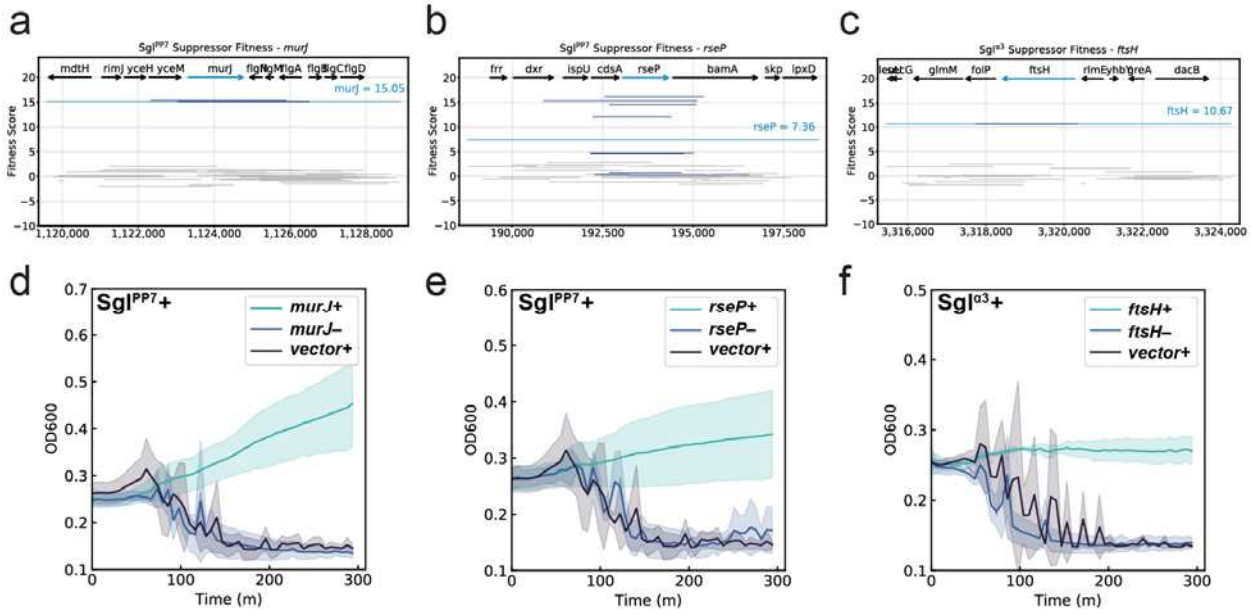


**Fig. 4: Sampling of validated multicopy suppressors against Sgl protein expression. (a,d)** Panels refer to suppressor activity of *murJ* expression against Sgl$^{PP7}$-mediated lysis. **(b,e)** Panels refer to suppressor activity of *rseP* expression against Sgl$^{PP7}$-mediated lysis. **(c,f)** Panels refer to suppressor activity of *ftsH* expression against Sgl$^{\alpha 3}$-mediated lysis. Dub-seq fragment plots for the highlighted suppressor locus are shown in panels (a,b,c). Dark blue lines correspond to fragments covering the gene of interest. Gray lines correspond to fragments not covering the gene of interest. Teal line corresponds to gene fitness score. (d,e,f) Panels show lysis inhibition growth effects from a 96-well microplate reader assay. In all panels the lysis protein was expressed from the same vector as in the corresponding

experiments in (a,b,c) at either 31.25 ($Sgl^{PP7}$) or 62.5 ($Sgl^{\alpha 3}$) ng/µL atc. Multicopy suppressors or empty vector controls were expressed from the corresponding ASKA mutant collection plasmid under IPTG control. Teal curves correspond to suppressor induction at 50 (*murJ*, *rseP*) or 200 (*ftsH*) µM IPTG induction. Black curves correspond to the empty ASKA vector negative control. Blue curves correspond to the uninduced suppressor plasmid. All curves are plotted as mean of three biological replicates across three days with interpolated ± standard deviation confidence intervals. Large variations in lysis conditions are likely caused by visible cell debris accumulation in the course of the microplate reader experiment.

Among the confirmed suppressor candidates for $Sgl^{PP7}$, MurJ, a lipid II flippase, an essential enzyme in PG synthesis that carry out flipping of Lipid II from cytoplasmic side to periplasmic side, was recently shown to be also targeted by $Sgl^{M}$. To further investigate the activity and specificity of MurJ suppression of $Sgl^{PP7}$ induced lysis, we coexpressed the heterologous lipid II flippases $MurJ_{TA}$ from *Thermsipho africanus* and Amj from *Bacillus subtilis* with $Sgl^{PP7}$ and compared the suppression activity by $Sgl^{PP7}$ with $Sgl^{M}$ (Fig. 5b). Our results indicate that $Sgl^{PP7}$ lethality can be rescued by expression of heterologous lipid II flippases strongly indicating that $Sgl^{PP7}$ targets MurJ.

Finally, to uncover greater details on the molecular interaction between $Sgl^{PP7}$ and MurJ, we isolated spontaneous mutants in *E.coli* MurJ after inducing *gfp-sgl*$^{PP7}$, a hybrid fusion that exhibited enhanced lytic function (Fig. 5b). After analyzing survivors, we discovered a single amino acid substitution in MurJ conferring $Sgl^{PP7}$-resistance: Q244P. Q244P is localized to transmembrane domain 7 (TMD7), one of the 14 transmembrane domains that is part of the solvent-exposed cavity of MurJ and, specifically, undergoes a major conformation shift as MurJ shifts between cytoplasmic- and periplasmic-open states [6]. Considering this amino acid change was previously observed to confer resistance to $Sgl^{M}$, these two dissimilar proteins target the same molecular interface of MurJ [6].

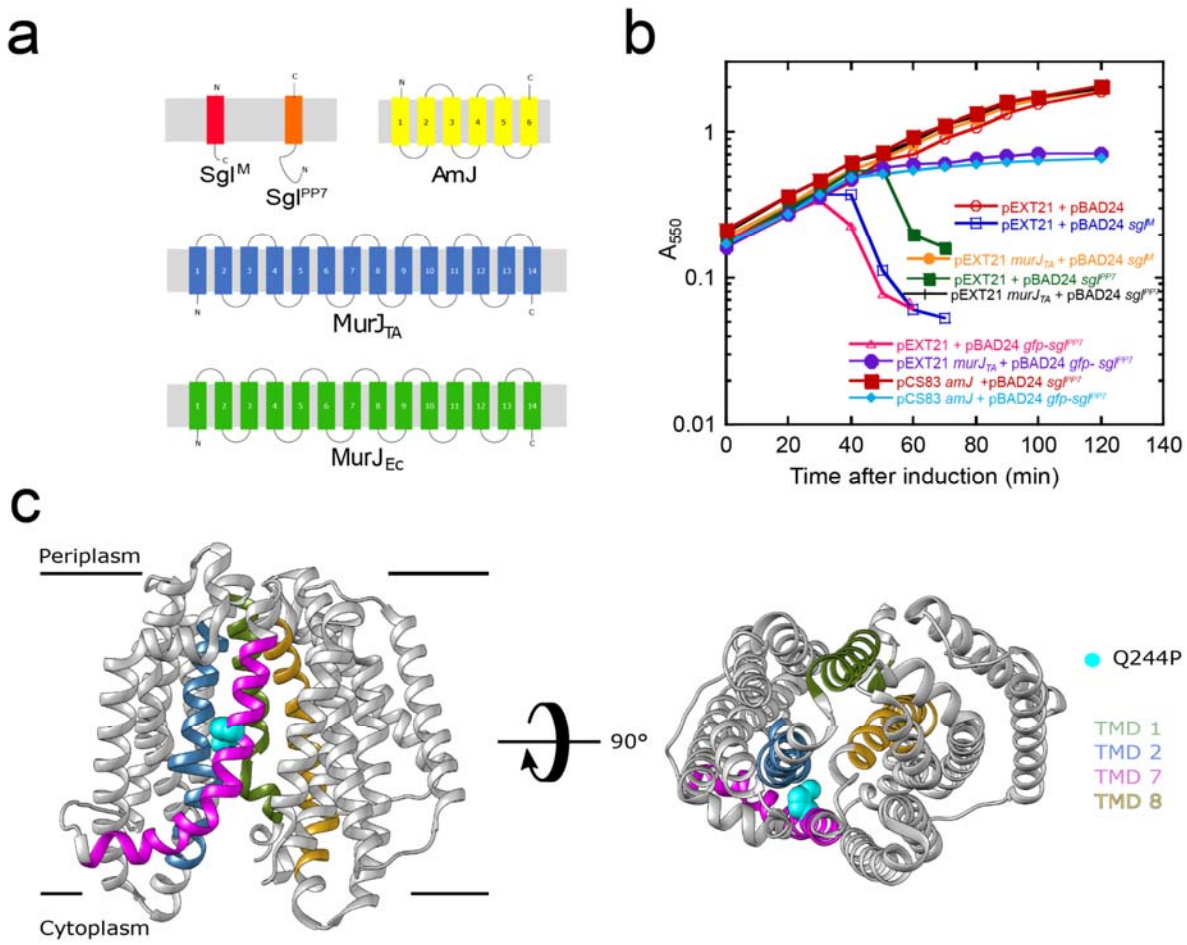


**Fig. 5: Sgl$^{PP7}$-MurJ interaction. (a)** Predicted membrane topologies of Sgl$^{M}$ (red), Sgl$^{PP7}$ (orange), AmJ (yellow), MurJ$_{TA}$ (*Thermosipho africanus*) (blue), and MurJ$_{EC}$ (*Escherichia coli*) (green) are shown in the context of bacterial cytoplasmic membrane (gray rectangle) with periplasmic side and cytoplasmic side represented above and below the gray rectangle, respectively. The N- and C- termini of the respective proteins are indicated with "N" or "C". **(b)** Lysis profiles of TB28 strain co-transformed with pEXT21 + pBAD24 (red open circle), pEXT21 + pBAD24 *sgl$^{M}$* (dark blue open square), pEXT2-*murJ$_{TA}$* + pBAD24 *sgl$^{M}$* (light orange filled circle), pEXT21 + pBAD24 *sgl$^{PP7}$* (dark green filled square), pEXT2-*murJ$_{TA}$* + pBAD24 *sgl$^{PP7}$* (black cross), pEXT21 + pBAD24 *gfp-sgl$^{PP7}$* (pink triangle), pEXT2-*murJ$_{TA}$* + pBAD24 *gfp-sgl$^{PP7}$* (purple filled circle), pCS83 *amJ* + pBAD24 *sgl$^{PP7}$* (red filled square), pCS83 *amJ* + pBAD24 *gfp-sgl$^{PP7}$* (light blue diamond). **(c)** The amino acid substitutions in *E. coli murJ* (MurJ$_{EC}$) that confer resistance to *gfp-sgl$^{PP7}$* are shown on the crystal structure of an inward open conformation of MurJ$_{EC}$ (PDB 6CC4). The TMDs that line the central hydrophilic cavity are colored as follows TMD1 (olive darb), TMD (steel blue), TMD7 (magenta), and TMD8 (gold rod). The substituted amino acid is highlighted as cyan spheres on TMD7 (magenta). Lateral view (left) and periplasmic view (right).

## Discussion

In this study we applied an unbiased genome-wide genetic screen to identify multicopy suppressors of 8 Sgl lysis proteins - a diverse group of lysis genes from *Fiersviridae* and *Microviridae* phages. In total, we observed 133 Sgl x suppressor combinations across $sgl^{\alpha 3}$, $sgl^{St1}$, $sgl^{M}$, $sgl^{PP7}$, $sgl^{PRR1}$, $sgl^{AP205}$, $sgl^{MS2}$, and $sgl^{Hgal1}$. Our study recapitulates results for how known Sgls target key host functions, uncovers a novel Sgl-host protein interaction, and, in many cases, reveal repair pathways that may compensate for Sgl toxicity. Additionally, we found Sgl-specific suppressor phenotype for distinct inner membrane proteases, suggestive of a constraining role in evolution of phage infectivity pathways. Finally, we propose a systems-level hypothesis for the mechanism of $Sgl^{MS2}$ based on enrichment of lytic transglycosylases.

One of the major unexpected findings of this study was that the $Sgl^{PP7}$ also targets MurJ, an essential lipid II flippase in gram negative bacteria. Interestingly, the $Sgl^{PP7}$ has little to no resemblance to the other MurJ-targeting Sgl, $Sgl^{M}$. Furthermore, these two Sgls could be overcome by the same single missense change (Q244P) in MurJ, suggesting that these two disparate Sgls have not only convergently evolved to target the same protein but also the same molecular interface on MurJ. The resistance allele Q244P is located on TMD7, one of the four TMDs lining the central hydrophilic cavity of MurJ. Interestingly, TMD7 undergoes large conformational changes between periplasmic-open and cytoplasmic-open states of MurJ and Gln244 is positioned at the bend in the helix. Locking MurJ in either of the conformational states leads to accumulation of Lipid II in the inner leaflet of the inner membrane and ultimately results in cell lysis. Previously, cysteine accessibility studies (SCAM) have shown that $Sgl^{M}$ locks MurJ in periplasmic-open conformation and blocks the transfer of Lipid II across the membrane. Given the putative interaction interface of $Sgl^{PP7}$ at the highly dynamic TMD7 of MurJ, potentially $Sgl^{PP7}$ locks MurJ in the opposite conformation to $Sgl^{M}$ i.e, cytoplasmic-open conformation. Future SCAM analysis and structural studies of $Sgl^{PP7}$-MurJ complex should shed light on both the conformation state of MurJ-$Sgl^{PP7}$ complex and its interaction interface.

Recently, the number of available Sgls has expanded by 35 and they share very little similarity to the previously known Sgls [18]. The high sequence diversity of Sgls naturally implies diversity in molecular targets to affect host cell lysis. However, the possible sequence space of Sgls in the hyper-expanded ssRNA phage genomes far outweighs the available molecular targets involved in cell envelope homeostasis. Hence, convergent evolution of Sgls to target the limited number of host targets is an inevitable consequence and one should expect more cases of convergent evolution to the known

targets such as MurA, MraY, and MurJ. The fact that both $Sgl^{M}$ and $Sgl^{PP7}$ target MurJ suggests that there is more than one way to exploit the same “weak spot” in the bacterial cell wall machinery. Furthermore, a target uncovered in one species of bacteria could also serve as one in another more distant species. Thus, by studying convergently evolved Sgls one could gain insights into built-in universal molecular “weak spots” across various species.

Here we limited this study to the discovery of suppressors for unique Sgl lysis proteins from *Microviridae* and *Fiersviridae*. We anticipate this unbiased genetic screening approach to be generalizable and extendable to discover suppressors of many other toxic genes found in nature including Sgls from the recent hyperexpansion of ssRNA phage genomes. Further, this approach could be very useful in the study and annotation of dsDNA and ssDNA phage genomes and host encoded small toxic genes [38]. We demonstrate here that by repurposing Dub-seq technology for carrying out unbiased and high-throughput suppressor screens will greatly expedite hypothesis generation and target identification of Sgl lysis proteins, providing a new avenue for antibiotic and phage-derived biotechnological discovery.

## Author contributions

B.A.A. conceived the project.
B.A.A., H.C., and V.K.M built and characterized the Dub-seq libraries used in this study.
B.A.A., K.C., H.C., and J.K., designed constructs, performed experiments, processed, and analyzed data.
A.M.D. provided critical reagents, support for sequencing efforts, and advice.
V.K.M and A.P.A arranged funding and supervised the project.
B.A.A., K.C., R.F.Y., V.K.M., and A.P.A. wrote the paper.

## Acknowledgments

The authors gratefully thank Simon Roux (Joint Genome Institute) for helpful discussions at various stages of this project. This project was funded by the Microbiology Program of the Innovative Genomics Institute, Berkeley. The initial Dub-seq library characterization for this project was funded by ENIGMA, a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract DE-AC02- 05CH11231. Sequencing was performed at: Vincent J. Coates Genomics Sequencing Laboratory (University of California at Berkeley), supported by NIH S10 Instrumentation Grants S10RR029668, S10RR027303, and OD018174.

## Methods

### *Bacterial strains and growth conditions*

In general, all *E.coli* strains were grown in Lysogeny Broth (LB-Lennox) broth (Sigma) with antibiotics at 37 ˚C, 180 rpm unless stated otherwise. When appropriate, 50 µg/mL kanamycin sulfate and/or 34 µg/mL chloramphenicol (denoted with +C or +K, respectively) were supplemented to media. All bacterial strains and libraries were stored at -80ºC for long term storage in 25% sterile glycerol (Sigma). All library assays were performed in NEB10beta strain backgrounds (New England Biolabs). For a complete list of strains and plasmids, please refer to Extended Data 6.

### *Construction of sgl expression strains*

Template sequences for *sgl*$^{\alpha 3}$, *sgl*$^{AP205}$, *sgl*$^{Hgal1}$, *sgl*$^{M}$, *sgl*$^{MS2}$, *sgl*$^{PRR1}$, *sgl*$^{PP7}$, and *sgl*$^{St1}$ were identified from the NCBI-deposited genomes: NC_001330, NC_002700, NC_019922, NC_019707, NC_001417, NC_008294, NC_001628, and NC_012868, respectively. As a toxic gene control, we used protein PC02664 detected from a phage genome infecting *E.coli* [39,40]. Each gene was codon-optimized for *E.coli*, had BsaI sites removed, and synthesized *de Novo* (IDT, TWIST Bioscience). *Sgl* genes were cloned into pBA368, a golden gate *gfp*-dropout vector derived from pBbA2K-rfp. DNA assembly was performed via golden gate assembly using BsaI (New England Biolabs), pBA368, and one of the synthesized *sgl*s. Reactions were cleaned up using Zymo Clean and Concentrate (Zymo), transformed into NEB10beta competent cells (New England Biolabs), and plated on LB+K. GFP- colonies were picked, grown up, stored at -80˚C, and verified for intact *sgl*.

For all strains, lytic activity was measured via plate reader assay before constructing suppressor libraries. Strains were inoculated into LB+K media overnight. Cells were diluted 50X into LB+K media with varying levels of anhydrotetracycline (aTc) ranging from 0-200ng/mL in a flat bottom 96-well plate (Corning 3904). Sgl-mediated lysis progressed in Tecan Infinite F200 readers with orbital shaking and OD600 readings every 15 min for 3-5 hours at 37˚C. Strains with functional Sgl phenotypes typically had visible lysis after ~90 minutes.

Plasmid pBAD24-*sgl*$^{PP7}$-*lacZ*$_{\alpha}$ was constructed in multiple steps. First, the *sgl*$^{PP7}$ (NC_001628.1) was codon-optimized for *E. coli* expression (IDT) and a synthetic DNA construct was obtained (GenScript). The synthetic *sgl*$^{PP7}$ DNA was amplified using primers KC94 and KC116 and the resulting PCR product was gel purified (Qiagen), digested with restriction enzymes EcoRI and XhoI (New England Biolabs), and sub-cloned into plasmid pKC3, replacing *sgl*$^{M}$ in pKC3.

Plasmid pBAD24-*gfp*-*sgl*$^{PP7}$-*lacZ*$_{\alpha}$ was constructed via the Overlap Extension PCR method using primers KC36 and KC127 to amplify a *gfp* megaprimer in the first PCR [41]. The megaprimer was then used to insert *gfp* into pBAD24-*sgl*$^{PP7}$-*lacZ*$_{\alpha}$ plasmid during the second PCR. The product of the second PCR reaction was treated with DpnI and then transformed into competent XL1Blue cells. The constructs were verified by sequencing (Eton Biosciences) with primers KC30 and KC31.

***Construction of Dub-seq suppressor libraries***

Dub-seq suppressor libraries were constructed by transforming the plasmid Dub-seq library, pFAB5516 [22], directly into *sgl* expression strains (above) via electroporation. Competent cells were created from an overnight culture diluted 70X into 25mL LB+K and shook at 37˚C, 180rpm for ~3 hours until OD600 0.5-0.7. The resulting mid-log cultures were chilled at 4˚C. Cultures were centrifuged (Beckman-Coulter Allegra 25R) for 5 minutes at 8000xg and subjected to three washes: (i) once with 25 ml chilled water (ii) and twice with 15 ml chilled 10% glycerol. The cell pellets after the final glycerol wash were resuspended in 10% glycerol, yielding □250 ul of *sgl* competent cells.

For each *sgl* library, 5 parallel transformations were performed to minimize inefficiency bias from any individual transformation. Each transformation consisted of 40µL of competent cells and 10ng of the pFAB5516 plasmid library transferred to a chilled cuvette (1mm gap, VWR). Cuvettes were electroporated using a BTX™-Harvard Apparatus ECM™ 630 Exponential Decay Wave Electroporator with the following parameters: voltage (1800 V), resistance (200 Ω), and capacitance (25 µF). Following each transformation, cells were recovered in 1mL LB+K media at 37˚C for 1 hour. For each transformation, 980µL of each recovery was plated and spread out onto LB+K+C Agar in a 245mm x 245mm bioassay dish (Nunc). The remaining 20µL of cells were serially diluted and plated onto a standard LB+K+C Agar plate to estimate the number of transformants per electroporation. All transformations were incubated at 37˚C overnight.

After overnight incubation at 37˚C, we first quantified the transformations to ensure we had at least 250,000 total estimated colonies (i.e., ≥5X pFAB5516 library coverage). We then picked 10 colonies from each of the transformations and PCR, followed by Sanger sequencing to ensure that the *sgl* was free of mutations. If any *sgl* mutations were detected in this subset, we repeated library construction. The transformant colonies were scraped and resuspended in 25mL LB+K+C media and processed as described

above to make multiple 1mL -80˚C freezer stocks [22]. Because pFAB5516 was characterized earlier [22], there was no need to perform library mapping PCRs at this step. An overview of library composition is summarized in Supplementary Table 1 and a gene-level description is shown in Extended Data 2.

### *Liquid culture fitness experiments*

Competitive fitness experiments were performed in liquid culture with two replicate experiments were performed per *sgl* suppressor experiment. Briefly, a 1 mL aliquot of suppressor Dub-seq library was gently thawed and used to inoculate a 25 mL of LB+K+C media. The library culture was grown to an OD600 of ~1.0 at 37˚C. From this culture two 1 mL pellets were collected, comprising the 'Time-0' or reference samples in BarSeq analysis. The remaining cells were diluted to a starting OD600 of 0.02 in LB with kanamycin and chloramphenicol. 690 µL of cells were mixed with 10µL of diluted aTc (Sigma) and transferred to a 48-well microplate (700 µL per well) (Greiner Bio-One #677102) covered with breathable film (Breathe-Easy). For all experiments, unless otherwise noted, aTc concentrations were used at 15.625 ng/mL. Sgl lysis progressed in Tecan Infinite F200 readers with orbital shaking and OD600 readings every 15 min for 8-12 hours at 37˚C. At the end of the experiment, each well was collected as a pellet individually. All pellets were stored at -80˚C until prepared for BarSeq. A summary of all library experiments is described in Extended Data 1.

### *BarSeq of Dub-seq pooled fitness assay samples*

Plasmid DNA was isolated from stored pellets of enriched and 'Time 0' Dub-seq samples using the QIAprep Spin Miniprep Kit (Qiagen). We performed 98°C BarSeq PCR protocol as described previously [24]. BarSeq PCR in a 50 µL total volume consisted of 20 µmol of each primer and 150 to 200 ng of plasmid DNA. For the HiSeq4000 runs, we used an equimolar mixture of four common P1 oligos for BarSeq, with variable lengths of random bases at the start of the sequencing reactions (2–5 nucleotides). Equal volumes (5 µL) of the individual BarSeq PCRs were pooled, and 50 µL of the pooled PCR product was purified with the DNA Clean and Concentrator kit (Zymo Research). The final BarSeq library was eluted in 40 µL water. The BarSeq samples were sequenced on Illumina HiSeq4000 with 50 SE runs. Typically, 96 BarSeq samples were sequenced per lane of HiSeq.

#### *Data processing and analysis of BarSeq reads*

Fitness data for Dub-seq suppressor libraries were analyzed as previously described with a few modifications as described, using *barseq* script from the *Dub-seq* python library with default settings [22]. From a reference list of barcodes mapped to the genomic

regions (BPSeq and BAGseq), and the barcode counts in each sample (BarSeq), we estimated fitness values for each genomic fragment using *gscore* script from the Dub-seq python library. At this step, instead of pooling all Time0 samples together, the time-zero samples within each suppressor library were pooled, since the composition and abundance of library members between libraries was distinct. For instance, $\alpha 3$ experiments had time-zero samples different from those of the *levM* experiments. The *gscore* script identifies a subset of barcodes mapped to the genomic regions that are well represented in the time-zero samples for a given experiment set. A barcode was required to have at least 10 reads in at least one time-zero (sample before the experiment) sample to be considered a valid barcode for a given experiment set. The *gscore* script was used to calculate a fitness score (normalized ratio of counts between the treatment sample and sum of counts across all time-zero samples) for the strains with valid barcodes. Frm the fitness scores calculated for all Dub-seq fragments, a fitness score for each individual gene *gscore* that is covered by at least one fragment was calculated using nonnegative least squares regression [22]. The nonnegative regression determines if the high fitness of the fragments covering the gene is due to a particular gene or its nearby gene, and avoids overfitting. Raw data for reads, fscores, and gscores across all experiments are provided in Extended Data 3, Extended Data 4, and Extended Data 5, respectively.

We applied a few additional filters to ensure that the fragments covering the gene had a genuine benefit. Briefly, we identified a subset of the effects to be reliable if the fitness effect was large relative to the variation between start samples ($|score| \geq 2$) for both mean and gene fitness scores [22], the gscores and fscores appeared to be reproducible across replicate experiments, and the number of reads for those fragments was consistently sufficient for the gene score to have little noise. Due to the strong selection pressure and subsequent fitness distribution skew resulting from Sgl activity, all candidate genes passing these filters were then subjected to manual scrutiny. For each gene, all barcodes were analyzed by fscore and reads. Several genes covered by few fragments (ie $\leq 3$), had inconsistent fscores, with orders of magnitude different read depth. This bias yielded inflated gscores and were discarded from further analysis. However, genes covered by individual fragments were kept in such cases. Some genes coincided with fit genes and had few instances where they didn't co-occur with a different gene, yielding inflated gscores. For instance, in the $sgl^{\alpha 3}$ experiments, *ompC* was fit for 8 of 9 fragments, but always co-occurred with *micF*, which was fit for 14 of 15

fragments. In such cases for instance, the gene was noted, but ultimately discarded from further analysis.

### *ASKA-based validations - building strains*

To validate select lysis suppressor phenotypes from suppressor screens we performed plate reader assays using additional plasmids derived from the overexpression ASKA library [37]. ASKA plasmids were miniprepped from the ASKA collection using a QIAprep Miniprep kit (Qiagen), transformed into the corresponding *sgl* expression strain, and plated on LB+K+C Agar. Transformants were verified by Sanger sequencing.

Plate reader assays for validations were performed as follows. Strains were inoculated into LB+K+C overnight. Cells were diluted 50X into LB+K+C media and allowed to grow at 37˚C, 180rpm to OD600 = 0.5. Cells were then transferred to a Corning 3904 96-well plate (Corning) and induced with varying levels of atc ranging from 0-250ng/mL for *sgl* expression and varying levels of IPTG ranging from 0-200µM for ASKA gene expression. Sgl lysis progressed in Tecan Infinite F200 readers with orbital shaking and OD600 readings every 10 min for 3-5 hours at 37˚C. Strains with unsuppressed lysis phenotypes typically had visible lysis after ~90 minutes.

### *Heterologous murJ suppression*

Strain TB28 was co-transformed with plasmids expressing lysis proteins ($Sgl^{M}$, $Sgl^{PP7}$, and GFP-$Sgl^{PP7}$) and compatible plasmids expressing MurJ orthologs ($MurJ_{TA}$ [42] and AmJ [43]) and selected on LB-Amp-Spec-IPTG (100µM) agar plates. The transformants were grown overnight at 37□°C with the same selective media and on the following day 1:200 dilutions of the overnights were added to 25 mL LB with appropriate antibiotics and IPTG (100µM) in a 250 mL flask. The cultures were induced with 0.4% w/v L-arabinose (Sigma-Aldrich) at OD550 ~0.2. The growth was plotted using Kaleidagraph 4.03 (Synergy Software).

### *GFP-$Sgl^{PP7}$-resistant mutant selection and screening*

Cultures of XL1-Blue pBAD24-*gfp-*$sgl^{PP7}$*-*$lacZ_{\alpha}$ were grown overnight at 37□°C with aeration. To perform the Sgl screen/selection, 100□µl of overnight culture was mixed with 400□µl LB and plated on LB-Ara-Amp-IPTG-X-gal agar plates (100 mm). The colonies that turned blue after overnight incubation at 37□°C were picked and purified on the same selection media. The $Sgl^{PP7}$-resistant colonies were grown overnight and both genomic (Qiagen QIAamp DNA micro kit) and plasmid (Qiagen mini prep kit) DNA were extracted. To rule out possible mutations in the lysis gene the plasmid was sequenced with primers KC30 and KC31. The *murJ* locus in the gDNA of the $Sgl^{PP7}$-resistant mutants was amplified by PCR using Phusion high-fidelity DNA polymerase (New England Biolabs) with the primers KC230 and KC234. The amplified PCR product

was gel-purified and sequenced with the primers KC230, KC231, KC232, KC233, and KC234.

## References

1 Cahill J, Young R. Phage Lysis: Multiple Genes for Multiple Barriers. *Adv Virus Res* 2019; **103**: 33–70.

2 Chamakura K, Young R. Phage single-gene lysis: Finding the weak spot in the bacterial cell wall. *J Biol Chem* 2019; **294**: 3350–3358.

3 Chamakura KR, Young R. Single-gene lysis in the metagenomic era. *Curr Opin Microbiol* 2020; **56**: 109–117.

4 Bernhardt TG, Wang I-N, Struck DK, Young R. Breaking free: 'Protein antibiotics' and phage lysis. *Res Microbiol* 2002; **153**: 493–501.

5 Typas A, Banzhaf M, Gross CA, Vollmer W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol* 2011; **10**: 123–136.

6 Chamakura KR, Sham L-T, Davis RM, Min L, Cho H, Ruiz N *et al.* A viral protein antibiotic inhibits lipid II flippase activity. *Nature Microbiology* 2017; **2**: 1–7.

7 Bernhardt TG, Roof WD, Young R. Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. *Proc Natl Acad Sci U S A* 2000; **97**: 4297–4302.

8 Zheng Y, Struck DK, Bernhardt TG, Young R. Genetic analysis of MraY inhibition by the phiX174 protein E. *Genetics* 2008; **180**: 1459–1466.

9 Bernhardt TG, Wang IN, Struck DK, Young R. A protein antibiotic in the phage Qbeta virion: diversity in lysis targets. *Science* 2001; **292**: 2326–2329.

10 Holtje JV, van Duin J. MS2 phage induced lysis of E. coli depends upon the activity of the bacterial autolysins. *Microbial cell wall synthesis and autolysis Elsevier/North-Holland Publishing Co , New York* 1984; : 195–199.

11 Chamakura KR, Tran JS, Young R. MS2 Lysis of Escherichia coli Depends on Host Chaperone DnaJ. *J Bacteriol* 2017; **199**. doi:10.1128/JB.00058-17.

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

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

14 Chamakura KR, Edwards GB, Young R. Mutational analysis of the MS2 lysis protein L. *Microbiology* 2017; **163**: 961–969.

15 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977; **74**: 5463–5467.

16 Marvin DA, Hohn B. Filamentous bacterial viruses. *Bacteriol Rev* 1969; **33**: 172–209.

17 Pourcel C, Midoux C, Vergnaud G, Latino L. A carrier state is established in Pseudomonas aeruginosa by phage LeviOr01, a newly isolated ssRNA levivirus. *J Gen Virol* 2017; **98**:

2181–2189.

18 Chamakura KR, Tran JS, O'Leary C, Lisciandro HG, Antillon SF, Garza KD *et al.* Rapid de novo evolution of lysis genes in single-stranded RNA phages. *Nat Commun* 2020; **11**: 6009.

19 Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X *et al.* Redefining the invertebrate RNA virosphere. *Nature* 2016; **540**: 539–543.

20 Callanan J, Stockdale SR, Shkoporov A, Draper LA, Ross RP, Hill C. Expansion of known ssRNA phage genomes: From tens to over a thousand. *Sci Adv* 2020; **6**: eaay5981.

21 Starr EP, Nuccio EE, Pett-Ridge J, Banfield JF, Firestone MK. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc Natl Acad Sci U S A* 2019; **116**: 25900–25908.

22 Mutalik VK, Novichkov PS, Price MN, Owens TK, Callaghan M, Carim S *et al.* Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nat Commun* 2019; **10**: 308.

23 Mutalik VK, Adler BA, Rishi HS, Piya D, Zhong C, Koskella B *et al.* High-throughput mapping of the phage resistance landscape in E. coli. *PLoS Biol* 2020; **18**: e3000877.

24 Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *Am Soc Microbiol* 2015; **6**: e00306–15.

25 Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 2018; **557**: 503–509.

26 Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792–1797.

27 Johnson JW, Fisher JF, Mobashery S. Bacterial cell-wall recycling. *Ann N Y Acad Sci* 2013; **1277**: 54–75.

28 Dik DA, Marous DR, Fisher JF, Mobashery S. Lytic transglycosylases: concinnity in concision of the bacterial cell wall. *Crit Rev Biochem Mol Biol* 2017; **52**: 503–542.

29 van Teeffelen S, Wang S, Furchtgott L, Huang KC, Wingreen NS, Shaevitz JW *et al.* The bacterial actin MreB rotates, and rotation depends on cell-wall assembly. *Proc Natl Acad Sci U S A* 2011; **108**: 15822–15827.

30 Lee TK, Tropini C, Hsin J, Desmarais SM, Ursell TS, Gong E *et al.* A dynamically assembled cell wall synthesis machinery buffers cell growth. *Proc Natl Acad Sci U S A* 2014; **111**: 4554–4559.

31 Cho H, Uehara T, Bernhardt TG. Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell* 2014; **159**: 1300–1311.

32 Lai GC, Cho H, Bernhardt TG. The mecillinam resistome reveals a role for peptidoglycan endopeptidases in stimulating cell wall synthesis in Escherichia coli. *PLoS Genet* 2017; **13**: e1006934.

33 Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R *et al.* Phenotypic landscape of a bacterial cell. *Cell* 2011; **144**: 143–156.

34 Saito A, Hizukuri Y, Matsuo E-I, Chiba S, Mori H, Nishimura O *et al.* Post-liberation cleavage of signal peptides is catalyzed by the site-2 protease (S2P) in bacteria. *Proc Natl Acad Sci U S A* 2011; **108**: 13740–13745.

35 Willett JLE, Gucinski GC, Fatherree JP, Low DA, Hayes CS. Contact-dependent growth inhibition toxins exploit multiple independent cell-entry pathways. *Proc Natl Acad Sci U S A* 2015; **112**: 11341–11346.

36 Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA *et al.* Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* 2004; **119**: 9–18.

37 Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H *et al.* Complete set of ORF clones of Escherichia coli ASKA library (A Complete S et of E. coli K-12 ORF A rchive): Unique Resources for Biological Research. *DNA Res* 2005; **12**: 291–299.

38 Wan X, Hendrix H, Skurnik M, Lavigne R. Phage-based target discovery and its exploitation towards novel antibacterial molecules. *Curr Opin Biotechnol* 2020; **68**: 1–7.

39 Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M *et al.* The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res* 2021; **49**: D751–D763.

40 Roux S, Páez-Espino D, Chen I-MA, Palaniappan K, Ratner A, Chu K *et al.* IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res* 2021; **49**: D764–D775.

41 Horton RM, Cai Z, Ho SM, Pease LR. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. BioTechniques 8(5):528-535 (November 1990). *Biotechniques* 2013; **54**: 129–133.

42 Kuk ACY, Mashalidis EH, Lee S-Y. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. *Nat Struct Mol Biol* 2017; **24**: 171–176.

43 Meeske AJ, Sham L-T, Kimsey H, Koo B-M, Gross CA, Bernhardt TG *et al.* MurJ and a novel lipid II flippase are required for cell wall biogenesis in Bacillus subtilis. *Proc Natl Acad Sci U S A* 2015; **112**: 6437–6442.