Origin of a core bacterial gene via co-option and detoxification of a phage lysin

Amelia M Randich, David T Kysela, Cécile Morlot, Yves Brun

▶ **To cite this version:**

Amelia M Randich, David T Kysela, Cécile Morlot, Yves Brun. Origin of a core bacterial gene via co-option and detoxification of a phage lysin. Current Biology - CB, 2019, 29 (10), pp.1634-1646.e6. 10.1016/j.cub.2019.04.032 . hal-02296249

HAL Id: hal-02296249

https://hal.univ-grenoble-alpes.fr/hal-02296249

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0960982219304300
Manuscript_dff92341560dff641d49b934d0f970a2

Title

Origin of a Core Bacterial Gene via Co-option and Detoxification of a Phage Lysin

Authors

Amelia M. Randich[1], David T. Kysela[1, 2], Cécile Morlot[3], Yves V. Brun[1, 2]

Correspondence: Y.V.B. yves.brun@umontreal.ca; C.M. cecile.morlot@ibs.fr
Lead Contact: Y.V.B. yves.brun@umontreal.ca

[1] Department of Biology, Indiana University, 1001 E. 3rd Street, Bloomington, IN 47405 USA
[2] Faculté de médecine, Département de microbiologie, infectiologie et immunologie, Université de Montréal, Pavillon Roger-Gaudry, 2900, boulevard Édouard-Montpetit, bureau P-607, C.P. 6128, Succursale Centre-ville, Montréal, QC H3C 3J7, Canada
[3] Univ. Grenoble Alpes, CNRS, CEA, IBS, 38000 Grenoble, France

Summary

Temperate phages constitute a potentially beneficial genetic reservoir for bacterial innovation despite being selfish entities encoding an infection cycle inherently at odds with bacterial fitness. These phages integrate their genomes into the bacterial host during infection, donating new, but deleterious, genetic material: the phage genome encodes toxic genes, such as lysins, that kill the bacterium during the phage infection cycle. Remarkably, some bacteria have exploited the destructive properties of phage genes for their own benefit by co-opting them as toxins for functions related to bacterial warfare, virulence, and secretion. However, do toxic phage genes ever become raw material for functional innovation? Here we report on a toxic phage gene whose product has lost its toxicity and has become a domain of a core cellular factor, SpmX, throughout the bacterial order Caulobacterales. Using a combination of phylogenetics, bioinformatics, structural biology, cell biology, and biochemistry, we have investigated the origin and function of SpmX and determined that its occurrence is the result of the detoxification of a phage peptidoglycan hydrolase gene. We show that the retained, attenuated activity of the phage-derived domain plays an important role in proper cell morphology and developmental regulation in representatives of this large bacterial clade. To our knowledge, this is the first observation of a phage gene domestication event in which a toxic phage gene has been co-opted for core cellular function at the root of a large bacterial clade.

Keywords

Lysozyme, GH24, Prophage domestication, Bacterial evolution, Alphaproteobacteria, *Caulobacter*, *Asticcacaulis*

Introduction

Understanding how new genes arise is key to studying the forces that drive diversity and evolution. Although horizontal gene transfer (HGT) is widely regarded as an important mechanism for exchanging *existing* genes among bacteria, mobile genetic elements can transfer exogenous genetic material that gives rise to *novel* genes. These new genes provide the basis for evolving new traits and propelling evolutionary transitions [1,2]. Temperate bacteriophages

© 2019 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

mediate genetic transfer by integrating their genomes into bacterial hosts [3–6]. These integrated gene tracts, called prophages, remain dormant until induced by various signals to produce phage particles and proteins that lyse the cell. In many cases, prophages contain genes that benefit the host, promoting prophage retention in many bacterial lineages, even after mutations have inactivated the prophage [7–9]. Accumulation of host-specific beneficial mutations in prophages has been referred to as "domestication." Many domesticated segments of inactivated prophages unexpectedly contain lytic and virion genes, which would intuitively be useless or even detrimental to the bacterial host [7]. Bacteria can use these genes as weapons against competing bacteria and eukaryotic hosts [10–14]. In contrast, we have identified an instance in which a toxic phage gene has not been repurposed as a weapon, but has evolved into a domain in a new core bacterial gene, *spmX*. Here, we report that SpmX resulted from an ancient domestication event at the root of the alphaproteobacterial order Caulobacterales, in which co-option and detoxification of a toxic phage gene gave rise to a novel bacterial gene with roles in developmental regulation and morphogenesis.

SpmX was first identified as a developmental regulator in the model organism *Caulobacter crescentus* [15]. Like most members of Caulobacterales, stalked *C. crescentus* cells divide asymmetrically to produce a stalked "mother" cell and a motile, flagellated "daughter" or "swarmer" cell. The *Caulobacter* developmental cycle depends on strict coordination of cell growth, chromosome replication and segregation, and division by various regulatory proteins that differ in localization and timing [16]. This network depends on regulatory phospho-signaling factors localized and regulated by polar scaffolds. SpmX is one protein scaffold that localizes at the stalked pole during the swarmer-to-stalked cell transition and recruits and potentially activates the histidine kinase DivJ [15]. Intriguingly, SpmX is required for stalk synthesis initiation and elongation in the closely related *Asticcacaulis* species *A. excentricus* and *A. biprosthecum* [17]. Therefore, this gene appears to have evolved multiple roles for defining cell morphology within this family of dimorphic, stalked bacteria.

Perplexingly, SpmX contains an N-terminal phage muramidase domain generally toxic to bacteria. Phages use these enzymatic domains to cleave the bacterial cell wall and lyse bacteria to release infectious phage particles. As a part of SpmX, this domain is critical for SpmX's role in both developmental regulation and stalk biogenesis: the muramidase domain is necessary for proper SpmX localization in both *C. crescentus* [15,18] and the *Asticcacaulis* genus [17]. Various studies have shown that SpmX localizes with the polar scaffold PopZ in *C. crescentus* [19,20] entirely through the muramidase domain [18]. The inability to measure enzymatic activity from purified *C. crescentus* SpmX muramidase domain has led to the conclusion that the domain lost its enzymatic activity and was repurposed for protein interactions and oligomeric assembly [18]. However, given the remarkable sequence similarity of the SpmX muramidase domain to functional phage lysozymes, including the canonical catalytic glutamate, total loss of enzymatic activity seems unlikely. Why would this domain be so highly conserved if its new function were merely for non-essential protein-protein interactions?

To better characterize the SpmX muramidase domain and the constraints underlying its conservation, we performed an in-depth bioinformatics study of more than 60 available SpmX genes together with structural determination, biochemical analysis, and comparative cell biology between *Caulobacter* and *Asticcacaulis.* We show that *spmX* arose prior to the diversification of Caulobacterales, a large order of stalked bacteria. We establish that the SpmX muramidase domain is a close relative of GH24 autolysin/endolysins that have been laterally exchanged via prophages. We find that the SpmX muramidase domain exhibits attenuated ancestral phage

activity, consistent with its remodeled active cleft. Finally, we demonstrate that this enzymatic activity is necessary for SpmX function in three representative species. We conclude that, close to the time of the genesis of the full-length *spmX* gene, the co-opted muramidase domain accumulated mutations that attenuated its hydrolytic activity on peptidoglycan and detoxified it for bacterial use. To our knowledge, this is the first case of phage gene domestication in which a toxic phage gene has been incorporated into a new core bacterial gene shared by a large bacterial order.

Results

The SpmX muramidase domain was co-opted from prophage in an early Caulobacterales ancestor. We first determined the prevalence of SpmX and its homologues in the bacterial domain. Simple pBLAST analysis revealed that SpmX, as defined by its three-part architecture with an N-terminal muramidase domain, a charged and proline-rich intermediate domain, and two C-terminal transmembrane (TM) segments (**Figure 1A**), is taxonomically constrained to Caulobacterales and one member of its sister taxa, Parvularculales. It is conserved as a single-copy gene in all sequenced members (**Table S1**). In all 69 identified *spmX* orthologues, the muramidase domains exhibit high amino acid sequence conservation (**Figure S1**), the intermediate domains high variability in length and sequence conservation, and the TM segments moderate sequence conservation among genera (**Figure 1A**). Apart from these orthologues, BLAST searches using SpmX only returned hits for the muramidase domain. These hits came from Gram-negative bacterial genomes that span the entire bacterial domain and from viral genomes. Most of these bacterial genes are likely to be in prophage regions, as evidenced by their position in tracts of prophage genes. We did not detect sequences homologous to SpmX TMs in our search, although we occasionally detected homologous phage muramidase domains fused to other, non-homologous TM segments.

Consistent with finding close SpmX muramidase relatives in prophages, NCBI's Conserved Domain Database (CDD) tool [21,22] clustered SpmX muramidase with glycoside hydrolase 24 (GH24) lysozymes in the autolysin/endolysin class. The sequence cluster diagram in **Figure S2** illustrates the inferred, ancient evolutionary relationships between lysozyme families based on sequence and structural alignments. These relationships allow us to determine a root for the GH24v lysozymes, with SpmX emerging relatively recently within this ancient clade of phage lysozymes. Autolysin/endolysins are closely related to classical phage T4 lysozyme-like (T4L-like) peptidoglycan hydrolases, which cleave peptidoglycan and lyse cells during the lytic cycle. These lysozymes are distinct from lytic transglycosylases (**Figure S2**, GH24λ), which include known housekeeping bacterial hydrolases with roles in cell growth and division. Lytic transglycosylases are also assigned to the GH24 group but share no sequence similarity with T4L-like muramidases [23–25]. Thus, although core bacterial genomes encode peptidoglycan hydrolases, the SpmX muramidase domain is most closely related to peptidoglycan hydrolases encoded by prophages and phage genomes.

Unlike its close relatives that have been transferred horizontally through the bacterial domain via prophage, the SpmX muramidase domain coding region has been inherited vertically as part of the *spmX* gene in Caulobacterales. The SpmX gene tree mirrors the phylogeny of Caulobacterales from concatenated gene alignments (**Figure 1B**). None of the *spmX* genes appear in tracts of prophage genes. The genomic context of *spmX* appears to be well maintained in members of Caulobacterales, with the gene occurring between a putative Mg^{2+} transporter and

a putative isovaleryl-CoA dehydrogenase in most species. Together, these findings suggest that SpmX muramidase domain is derived from an autolysin/endolysin no longer within a prophage island but instead under direct cellular control. It likely fused with the intermediate and TM domains in a common ancestor of Parvularculales and Caulobacterales. The vertical transmission of *spmX* and strong sequence conservation of the muramidase domain suggests an important cellular function for the gene among Caulobacterales members.

The SpmX muramidase domain retains the canonical GH24 motif but contains mutations in the catalytic cleft known to inactivate phage lysozymes. To determine if critical enzymatic residues in SpmX muramidase were conserved, we compared SpmX amino acid sequences to other GH24v lysozymes. By definition, lysozymes catalyze the hydrolysis of β1,4-linked glycosidic bonds in peptidoglycan and chitin [25]. This superfamily includes at least seven distinct groups (five are represented in **Figure S2**) that are unrelated by sequence similarity but share a common fold in which the catalytic Glu and the beta-hairpin motif in the N-terminal lobe pack against the C-terminal lobe to form the catalytic cleft (**Figure 2A)** [26]. This beta-hairpin, or GH motif, contains family-specific residues critical for enzyme activity in all lysozyme superfamily members [26].

We compared the SpmX GH motif to those of lysozymes from the T4L and endolysin/autolysin classes, which should share the same family-specific residues. **Figure 2B** shows the amino acid conservation in the GH motif of T4L-like, autolysin/endolysin, closely related non-SpmX muramidase, and SpmX muramidase protein sequences. Because the autolysin/endolysin class and the closest non-SpmX relatives are likely to be active phage enzymes, highly conserved residues shared by these groups with T4L delineate positions that are evolutionarily constrained for phage lysozyme activity and stability in this clade. For example, D10 is not conserved outside of T4L-like enzymes because the autolysin/endolysin class does not have a salt bridge between D10 and the C-terminal lobe [27]. On the other hand, all of the putative phage sequences (**Figure 2B(i-iii)**) conserve the T4 lysozyme "catalytic triad": the catalytic residue E11 and active site residues D20 and T26. While the exact roles of D20 and T26 are not clear, they are critical for effective catalysis [26–29]. Position D20 is very sensitive to mutation, with only substitutions D20C/A retaining the hydrolytic activity of T4L or P22 phage lysozymes [30]; these substitutions are tellingly well represented amongst the putative phage sequences. Remarkably, SpmX muramidase domains demonstrate strong conservation of residues required for the GH motif, but low conservation of residues associated with catalysis, with the exception of the main catalytic residue, E11 (**Figure 2B(iv)**). The majority of SpmX genes contain the mutation D20L/R, both of which reduced T4L activity to less than 3% of WT in previous studies [27] and which are distinctly unrepresented in the other phage muramidases. Moreover, the T26 position no longer appears to be under selective constraint in SpmX. The conservation of the GH motif coupled with the apparent inactivation of the catalytic triad across all SpmX genes suggests that the catalytic cleft has been remodeled structurally and that the muramidase domain may therefore not retain the same level of activity or function as phage GH24v lysozymes.

The SpmX muramidase domain has a wider, more dynamic catalytic cleft than related phage lysins. Obtaining the structure of the SpmX muramidase domain (residues 1-150) from *Asticcacaulis excentricus* (SpmX-Mur-*Ae*) (**Table S2**) allowed us to directly visualize the effect of the D20L and T26X mutations on the catalytic cleft. Overall, SpmX-Mur-*Ae* exhibits the

characteristic T4 lysozyme structure: the predicted catalytic glutamate occurs at the C-terminal end of the first alpha-helix, within the catalytic cleft formed between the N- and C-terminal lobes (**Figure 3A**). P22 lysozyme (the model for molecular replacement) and the active conformation of the distantly related SAR endolysin protein R21 (PDB 3HDE) from bacteriophage P21 (**Figure S2**) are overlaid in the structural alignment in **Figure 3A** to emphasize the manner in which the SpmX muramidase domain deviates from these phage lysozymes: besides the extended beta-hairpin in the C-terminal lobe, the canonical GH beta-hairpin in the N-terminal lobe of SpmX-Mur-*Ae* splays away from the catalytic cleft relative to those of the phage lysozymes. This GH beta-hairpin region exhibited the most conformational differences among the three molecules of SpmX-Mur-*Ae* in the asymmetric unit. The overlay of the three SpmX-Mur-*Ae* chains in **Figure 3B** illustrates how the orientation of the GH beta-hairpin is tilted by about 16° between chains A and B, suggesting a heightened flexibility in this region compared to other T4L-like lysozymes, which may reduce the ability of the enzyme to coordinate peptidoglycan hydrolysis in the catalytic cleft.

GH motif sequence alignments (**Figure 2B)** show that SpmX muramidase domains have lost a highly conserved tyrosine residue at position 18. Although T4L enzymatic activity is not sensitive to mutation at this position [27], it is invariant across all the phage lysozyme classes we analyzed. Visualization of Y18 in the P22 lysozyme structure (**Figure 3C**) shows that it interacts with R14 at the base of the beta-hairpin, possibly a critical interaction for coordinating the beta-hairpin with the catalytic glutamate. In SpmX-Mur-*Ae*, Y18S still appears to make hydrogen-bonding contact with R14; however, most SpmX muramidase domains have non-polar residues at position 18 (**Figure 2B(iv)),** which may reduce coordination. It has been previously shown that the Y18 position is a hot-spot for compensatory mutations that restore activity to inactive catalytic mutants [31], and it is intriguing to imagine that mutations at this position in SpmX muramidase are associated with the ability of its remodeled, more flexible catalytic cleft to still bind and/or cleave peptidoglycan.

Figure 3 shows the catalytic clefts of both P22 lysozyme (**D**) and SpmX-Mur-*Ae* (**E**). In P22 lysozyme, the E11-carbonyl, D20-carboxyl, and T26-hydroxyl groups point into the aqueous catalytic cleft. In SpmX-Mur-*Ae,* the cleft is slightly reorganized, with the T26M S-methyl thioether still within 20 Å of E11 and potentially capable of interacting with peptidoglycan. In about two thirds of the SpmX genes, position 26 is either a valine or an isoleucine, which do not have any polar moieties to contribute to the cleft (**Figure 2B(iv)**). With this structural data, we can infer that the SpmX muramidase domain has a remodeled catalytic cleft with a correctly positioned catalytic glutamate. However, the increased flexibility between the GH motif and the glutamate, as well as the loss of key coordinating residues might reduce, if not eliminate, SpmX hydrolytic activity, and would explain why previous groups could not detect hydrolytic activity from purified SpmX muramidase [18].

SpmX retains reduced hydrolytic activity on peptidoglycan. Given SpmX's reported inactivity [18] and the structural data suggesting the catalytic cleft is capable of interacting with peptidoglycan, we hypothesized that the domain retains ancestral function in binding peptidoglycan. To test this, various constructs from *C. crescentus, A. excentricus,* and *A. biprosthecum* were purified and incubated with sacculi from all three species. Both muramidase and entire soluble domains including the intermediate domain bound sacculi from all three species (**Figure S3**). Since the purified protein was capable of binding its putative substrate, we also tested its ability to hydrolyze peptidoglycan. We used remazol brilliant blue (RBB) assays to

compare the activity of SpmX muramidase from *C. crescentus* (SpmX-Mur-*Cc*) to P22 lysozyme (P22Lyso) and its D20L mutant (P22Lyso-D20L) (**Figure 4A**), and found that both SpmX-Mur-*Cc* and P22Lyso-D20L exhibit similarly attenuated hydrolytic activity in comparison to P22Lyso. Both reached maximal levels of RBB release near enzyme concentrations of 15 μM while P22Lyso reached the same levels near 5 μM. Mutants in which the catalytic glutamate was replaced with alanine (SpmX-Mur-*Cc*-E11A and P22Lyso-E11A) did not exhibit activity (**Figure S4A**). These data indicate that the "inactivating" substitution D20L attenuates enzymatic activity whereas mutating the catalytic glutamate abolishes it altogether.

Because the D20L mutation reduced P22Lyso's activity close to that of SpmX muramidase, it was possible that this mutation was responsible for SpmX's attenuated activity. However, restoring the ancestral D20 (SpmX-Mur-*Cc*-L20D) did not increase SpmX activity *in vitro* (**Figure S4B**). We suspect that the additional accumulation of mutations in SpmX muramidase, such as the drift observed at Y18 and T26 in the cleft, has made it impossible to restore ancestral phage lysozyme activity with a single mutation. Because the D20L mutation is ancestral in the SpmX phylogeny (**Figure 1B**) and capable of attenuating P22 lysozyme activity to SpmX-like levels, we infer that this mutation likely occurred first. The increased flexibility of the GH motif observed in the SpmX-Mur-*Ae* structure is therefore the consequence of many mutations that accumulated either neutrally after the D20L substitution attenuated the activity, or selectively to shape the new function of the domain as part of SpmX.

The enzymatic activity of the P22Lyso-D20L was puzzling in light of early work that reported that D20 mutations inhibited T4 lysozyme in phage plaque assays [27]. One possible explanation is that the D20L mutation reduces lysozyme activity to the point that it is not suitable for cell lysis at *in vivo* expression levels, and that T4 lysozyme with the mutation was unable to complete infection and form plaques. To explore this possibility, we designed an experimental system to test the activity of P22Lyso and SpmX-Mur-*Cc* mutants in the *E. coli* periplasm using fusions to the N-terminal PelB leader sequence (pET22b). Lemo21(DE3) cells expressing P22Lyso lysed without induction (**Figure 4B),** indicating that marginal P22Lyso levels can drive cell lysis. In contrast, cells expressing P22Lyso-D20L lysed only after induction **(Figure 4C)**, confirming that much higher enzyme concentrations were needed. Thus the D20L mutation may represent a critical detoxification step that reduced the ability of the domain to lyse the cell and made it available for co-option.

Although purified P22Lyso-D20L and SpmX-Mur-*Cc* had similar activation curves *in vitro,* Lemo21(DE3) strains expressing SpmX-Mur-*Cc* never lysed (**Figure 4C**). This was despite equivalent periplasmic expression levels to P22Lyso-D20L (**Figure S4D)**. Different growth conditions and media increased the amount of SpmX-Mur-*Cc* in the periplasm but did not affect cell viability (**Figure S4C**). Moreover, SpmX-Mur-*Cc* was active on sacculi isolated from Lemo21(DE3) (**Figure S4E**), eliminating the possibility that it could not cleave *E. coli* peptidoglycan. It is possible that SpmX-Mur-*Cc* cannot fold correctly in the *E. coli* periplasm, or that its activity is further attenuated in the periplasmic environment. However, the periplasmic expression tests in *E. coli* confirm that the D20L mutation attenuates P22Lyso hydrolytic activity and thereby increases the amount of protein required to induce lysis. This tuning of enzymatic activity might have served as a critical detoxifying step in the co-option of the muramidase domain from phage. Because SpmX has retained the ancestral catalytic glutamate and its modified catalytic cleft is capable of hydrolytic activity, we conclude that this attenuated activity is under purifying selection in SpmX and must be important for SpmX function.

Inactivating the muramidase domain interferes with SpmX localization *in vivo.* To determine the role of the preserved, albeit attenuated, activity of the muramidase domain in SpmX function, we inactivated it by mutating the conserved catalytic glutamate to alanine (E11A, E19A in SpmX numbering) at the chromosomal locus in various species and observed the effects *in vivo* (**Figure 5**). We determined the effects of the E11A mutation on cellular morphology, as the *C. crescentus*, *A. excentricus*, and *A. biprosthecum* Δ*spmX* strains all have morphological phenotypes (**Figure 5ABCii**): In *C. crescentus*, Δ*spmX* cells have a characteristic elongated morphology resulting from failed division cycles and often grow stalks prematurely from daughter cells that fail to divide completely (**Figure 5Aii**) [15]. In *Asticcacaulis*, Δ*spmX* cells lack stalks without other apparent developmental phenotypes (**Figure 5BCii**) [17]. If enzymatic activity is critical for overall SpmX function, we expected that eliminating catalytic activity with the E11A mutation would phenocopy Δ*spmX*. However, we observed intermediate phenotypes for this mutation. In *C. crescentus*, the E11A mutant population contained both WT-like cells and cells exhibiting the division defect, but with less severity than in Δ*spmX* (**Figure 5Aiii**). In both *Asticcacaulis* species, the E11A mutants still grew stalks (**Figure 5BCiii**). Nevertheless, the *A. biprosthecum* E11A mutant exhibited a significant loss of bilateral stalks (3.5 fold reduction) and an increase in the frequency of cells with a single stalk (**Figure S5D**). These results suggest that eliminating catalytic activity does not fully inhibit SpmX function.

WT and mutant SpmX GFP fusions allowed us to monitor changes in SpmX cellular localization. As shown previously [15,17], WT SpmX localized at the future position of the stalk, at the pole as in *C. crescentus*, or at sub-polar or bilateral positions in *Asticcacalis*, and was retained at this position during stalk elongation (**Figure 5ABCiii**). Both *C. crescentus* and *A. biprosthecum spmX* E11A mutants exhibited an increase in delocalized fluorescence throughout the cell body compared to WT (**Figure 5ABiv**). Quantification of the fluorescence data indicated that while the overall mean cell fluorescence was the same as WT, the SpmX foci were significantly less intense in the mutants (**Figure S5AB**). We also observed a 3X increase in the stalk fluorescence in *A. biprosthecum* expressing SpmX E11A compared to WT (**Figure S5B**). Although no difference in focal fluorescence intensity was observed in *A. excentricus spmX* E11A mutant cells, more cells had a second SpmX focus at stalk tips than WT cells (**Figure 5iv, S5C**), indicating altered localization. Western blots of cells expressing WT SpmX-eGFP and SpmX mutants confirmed that the delocalized fluorescence was not due to clipping of the GFP tag, but to delocalized SpmX protein (**Figure S5E**). Together these data show that the E11A mutation disrupts SpmX localization in all three species and may underlie the morphological defects observed in *C. crescentus* and *A. biprosthecum*.

Because the E11A intermediate phenotype suggested that the mutation might be disrupting SpmX localization by interfering with peptidoglycan interactions, we also mutated a position associated with peptidoglycan binding, but not catalysis, in T4 lysozyme. N/Q105 has been shown to coordinate peptidoglycan in the active cleft [32] and the mutation Q105R abolished activity in T4 phage plaque assays [27]. The mutation N105R (N91R in SpmX numbering) in *C. crescentus* and *A. biprosthecum* resulted in similar delocalization and intermediate morphological phenotypes as E11A (**Figure 5BCiiv**). We also investigated the effects of restoring the phage active site D20 (L28D in SpmX numbering) to the catalytic cleft. However, this had no evident effect on SpmX localization or cell morphology (**Figure 6i**), suggesting that the D20L substitution in SpmX, while ancestral, is not strictly necessary for SpmX function. This finding is in line with the observation that SpmX-Mur-*Cc*-L20D activity was not significantly different from WT in our *in vitro* RBB assays (**Figure S4B**). Therefore, the

D20L substitution was likely a key first step in SpmX detoxification but no longer appears to be under fitness constraints.

Overall, these data show that inactivating enzymatic activity or reducing the peptidoglycan-binding capability of the muramidase domain affects SpmX localization and function. Although it is not clear whether disrupting SpmX localization with the E11A mutation stems from eliminating SpmX's hydrolytic activity or decreasing SpmX's binding affinity for peptidoglycan, the similar phenotype from mutating a predicted peptidoglycan-interacting residue (N105R) underscores the importance of SpmX-peptidoglycan interactions. That the catalytic mutant has an intermediate morphological phenotype in *C. crescentus* and one *Asticcacaulis* species indicates that the muramidase domain may coordinate SpmX functions similarly in the two genera and that this function likely relies on its interactions with peptidoglycan.

Replacing the muramidase domain, or removing it, depletes native SpmX protein levels *in vivo*. Because SpmX localization depended on the ability of the muramidase domain to interact with peptidoglycan, we were interested in whether swapping alternative muramidase domains into SpmX would support WT function. We first made chimeras wherein P22 lysozyme replaced the domain with the hypothesis that (1) P22 lysozyme would be too active and therefore toxic to the cells and that (2) P22 lysozyme E11A might be able to support some level of SpmX localization. While SpmX and the SpmX-E11A mutant exhibited the previously determined morphological and delocalization phenotypes (**Figure 6Aiii-iv**), chimeras with P22 lysozyme were surprisingly viable but phenocopied the parent Δ*spmX* strain and lacked fluorescence. We were unable to detect any GFP-fusion products in this chimera by Western blot (**Figure 6B**), but confirmed by sequencing that P22Lyso-SpmX had been correctly inserted at the *spmX* locus, suggesting that the chimeras were likely expressed but quickly degraded in *C. crescentus*. Therefore the phenotype of this chimera is due to the loss of SpmX and not the addition of the P22 lysozyme domain. Inactivating P22 lysozyme (E11A) did not change the outcome, suggesting that the toxicity of the phage muramidase was not driving SpmX degradation.

To determine whether the loss of SpmX protein levels was particular to using P22 lysozyme, we verified the phenotype when SpmX lacked the muramidase domain entirely. Deletion of the muramidase domain from the *spmX* locus in all three species also resulted in strains with the Δ*spmX* phenotype that failed to produce detectable amounts of Δmur-SpmX-sfGFP by Western blot (**Figure S5E**). These results suggest that the SpmX muramidase domain is necessary to produce and/or maintain WT levels of SpmX in all three species, and that P22 lysozyme, despite high sequence similarity (51%) and structural homology (RMSD 1.7 Å), is not sufficient to replace it. P22Lyso and SpmX-Mur-*Cc* are nonetheless fairly distantly related, so we tested the ability of other SpmX muramidase domains to replace that of *C. crescentus*. Previously, *C. crescentus* and *Asticcacaulis* muramidase domains were shown to be interchangeable [17], so we extended the sequence distance to SpmX muramidases from the next closest relative *Brevundimonas subvibrioides*, which has D20R in the catalytic cleft, and the most distant relative *Parvularcula bermudensis,* which shares the D20L mutation. We found that the muramidase domain from *B. subvibrioides* supported the WT phenotype in *C. crescentus* (**Figure 6v**), but that the SpmX muramidase domain from *P. bermudensis* did not. We were surprised to see no evidence of delocalization in the *B. subvibrioides* SpmX chimera because the L20R point mutant of SpmX in *C. crescentus* showed some delocalization (**Figure 6vi**). This result suggests that the L20R mutation in the brevundimonads must coexist with other

compensatory mutations. The SpmX muramidase domain from *P. bermudensis*, like P22 lysozyme, must be too distant from *C. crescentus* to support WT expression levels. In combination with data from the P22Lyso chimeras, these data indicate that a T4L GH fold alone is not sufficient for SpmX function, and that the SpmX muramidase domain must contain other mutations necessary for stable protein levels in Caulobacterales. It could suggest that this domain has additional constraints on it unrelated to potential peptidoglycan interactions, such as binding interfaces specific to its function as a recruiting factor and protein scaffold.

Discussion

Bacteriophages shape bacterial evolution in various ways: they increase bacterial diversity by selectively preying on species [2,33], drive horizontal gene transfer [8,34], and serve as reservoirs of raw material for genetic innovation [35,36]. Phages are heralded as a major source of genetic material for novel gene emergence in bacteria [2,35,36], but, as we discuss later in this section, very few examples of novel gene emergence from prophage exist in the literature. We have investigated the origin and function of a taxonomically restricted gene from Caulobacterales, *spmX*, and determined that its occurrence is the result of the fusion and domestication of a phage peptidoglycan hydrolase gene. Although SpmX functions as a scaffold in developmental regulation and morphology, its muramidase domain retains high sequence similarity to phage lysozymes, which are toxic to bacteria. The active cleft contains mutations that have attenuated the toxic activity of the domain, presumably making it available for genetic innovation and bacterial use. We show here that the domain remains enzymatically active on peptidoglycan and that eliminating this activity alters the function of the full-length protein *in vivo*. Thus, the SpmX gene represents a core gene innovation specific to the Caulobacterales order that originally arose from a prophage gene with antibacterial activity.

Previously, it was suggested that the SpmX muramidase domain functions only in protein-protein and self-oligomerizing interactions in SpmX's role as a developmental regulator and scaffold in *C. crescentus* [18]. This conclusion was based on the lack of detectable activity from the purified domain and the inability of the catalytic E11R (E19R in SpmX numbering) mutant to self-oligomerize. It is highly likely that the E11R mutation greatly destabilizes the muramidase domain structure. We found that even the E11A mutant eluted in multiple fractions during purification, indicating decreased conformational stability. Moreover, the E11R protein product was no longer detectable in the cells expressing the gene [18]. Thus, the effect of the E11R mutation is similar to using a distantly related muramidase domain (like P22Lyso) or deleting portions of the muramidase domain entirely [15]. These data indicate that the muramidase domain plays an unanticipated role in maintaining stable SpmX protein levels across all tested species: without an appropriate muramidase domain, SpmX is misfolded, misprocessed, and/or quickly degraded.

Inactivating the SpmX muramidase domain resulted in developmental defects in *C. crescentus* and significant decrease in bilateral stalks in *A. biprosthecum*. Curiously, inactivating the enzymatic domain did not yield a null phenotype or complete delocalization. It is possible that enough peptidoglycan-interactions are maintained in the mutants for the domain to function as a peptidoglycan-binding domain. It is also possible that SpmX recruits proteins with redundant enzymatic activity that cannot be recruited in the Δ*spmX* mutant, as it is already known that SpmX interacts with targeting factors via its C-terminal domains in *Asticcacaulis* [17] and possibly via its transmembrane segments with DivJ [15]. Finally, it is hard to

distinguish whether there is a direct relationship between catalytic activity and peptidoglycan binding, or if cleaving peptidoglycan could indirectly localize SpmX. The multiple domains and pleiotropic effects of SpmX make it difficult to assess the effects of an individual domain on its *in vivo* function. However, our data support a model in which the muramidase domain of SpmX is still active, and this activity is used to localize SpmX. We conclude that the muramidase domain functions in localizing SpmX via its interactions with peptidoglycan rather than self-oligomerization as previously hypothesized. This proper localization is necessary for its roles in development and morphology.

SpmX emerges in the genomic record at the root of Caulobacterales with the attenuating D20L mutation (**Figure 1B**). The D20L mutation is therefore ancestral and potentially the initial step in the co-option of the domain. D20L conservation throughout most of Caulobacterales suggests evolutionary constraint on this position despite no observable phenotype from the SpmX-Mur-*Cc* L20D reversion mutation *in vivo* or *in vitro*. After the D20L substitution detoxified the muramidase domain, the domain likely accumulated both neutral and occasional adaptive mutations in the context of its new function. The active cleft contains several modifications, including the loss of selection on the third catalytic triad position, T26, and the invariant residue Y18. This pair is interesting in that Y18 was identified as a hotspot for spontaneous second site revertants of T26 mutants in T4 lysozyme [31]. It is possible that the changes we see at these two positions are compensatory mutations retaining attenuated activity, although there is no clear history of covariation. Accumulation of these types of mutations likely underlies the inability to restore phage lysozyme-like activity by reversing the D20L substitution. The ancestral D20L mutation has diverged in two groups: *Oceanicaulis* and *Maricaulis* (D20R/G), and *Brevundimonas* (D20R) (**Figure 1B, S1**). Interestingly, D20R/G is covariant with residue N105S/D (**Figure S1**), a peptidoglycan-interacting residue in T4L [32]. The covariance of peptidoglycan-interacting residues in these diverging genera further underscores the importance of this domain in peptidoglycan interactions, rather than just protein-protein interactions.

spmX arose recently enough to see the hallmarks of novel gene emergence and adaptation in a constrained bacterial clade. The gene either arose from a fusion event in the bacterial genome, or the original phage gene contained the transmembrane segments. Detoxification of the muramidase appears concomitant with the origin of the full SpmX gene comprising three fused domains. Maintenance of the muramidase domain since the emergence of SpmX and its activity in current living Caulobacterales members suggest that its attenuated activity was selected for in the ancestral protein and still involved in its modern functions. In contrast, SpmX's downstream intermediate domain is highly variable throughout Caulobacterales (**Figure 1A**). This domain appears to experience comparatively minimal sequence constraint and has undergone multiple independent events of elaboration and reduction in this clade. This region of charged residues and prolines drives SpmX self-oligomerization *in vitro* [18], and may also facilitate other protein interactions. For example, the intermediate domain appears to be responsible for targeting SpmX to sub-polar and bilateral positions in *Asticcacaulis* [17].

In several reported cases bacteria have domesticated phage genes for genetic manipulation and transfer, bacterial warfare, virulence, and secretion. However, these events are distinct from that which created the novel bacterial gene *spmX*. Phage genes for DNA replication and recombination have replaced bacterial functional homologues within bacterial genomes several times [37–40], however these genes retain their original function and carry out the same tasks. Gene transfer agents (GTAs) pose an interesting case where virion proteins from cryptic

prophage package random DNA from the bacterial genome to presumably share with other bacteria [41]. Although a specific GTA has been stably maintained across several alphaproteobacterial orders, this domesticated island of phage genes still shuttles DNA around, as it once did in ancestral infectious cycles [42,43]. Phage tails have been weaponized many times, resulting in type VI secretion systems [44,45], tailocins and phage tail-like bacteriocins [12–14], phage tail-like systems with insecticidal properties [11,46,47], and phage tail-like arrays [48]. All of these represent a "guns for hire" acquisition scheme in which phage genes are co-opted for their ancestral toxicity and function [2]. Many of these genes reside in genomic islands and confer environmental, niche-specific advantages that directly exploit their ancestral activity for the benefit of the host. Similarly, in two other known cases of phage lysozyme domestication in bacteria, muramidase domains have been fused to colicins [49] or are predicted to be secreted with type III secretion systems [50], presumably for use in bacterial warfare or infection. In one strange case, a phage lysozyme gene has been co-opted in bivalve genomes, which apparently still use the gene for its antibacterial properties [51].

The domestication of the muramidase domain in SpmX is distinct from the above cases of "guns for hire" because the phage gene has been incorporated into a novel bacterial gene with new function in basic cellular processes in a large bacterial order. The SpmX muramidase domain, although active, no longer lyses bacterial cells; instead it plays a role in localizing SpmX for its function in developmental regulation and morphogenesis. The co-option of phage genes for core cellular function is likely a common event in nature, but identifying such genes may require a careful search. Based on our findings, we suggest a future strategy for their detection: searching for phage gene homologues with long histories of vertical inheritance and signs of innovation in bacterial genomes.

Acknowledgements

We thank members of the Brun, Vernet, and Dessen laboratories for support, advice and encouragement. Many thanks to Breah LaSarre, Farrah Bashey-Visser, Jay Lennon, Daniel Schwartz, and Ernesto Vargas for critical manuscript reading and editing. Enthusiastic thanks for critical preprint review by the journal clubs of the (Pamela) Brown Lab (University of Missouri) and the Süel Lab (UCSD). We thank David Flot (ESRF, beamline ID30a1) for support in data collection. The NIH supported this work with grants 2R01GM051986 and R35GM122556 (to Y.V.B.), and NRSA F32GM112362 (to A.M.R.). A Fulbright US Research Scholar Award supported the work of Y.V.B. at the Institut de Biologie Structurale in Grenoble. Y.V.B is supported by a Canada 150 Research Chair in Bacterial Cell Biology. This work used the platforms of the Grenoble Instruct-ERIC Center (ISBG : UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Author Contributions

Conceptualization, A.M.R. and Y.V.B.; Methodology, A.M.R., D.T.K., C.M., and Y.V.B.; Investigation, A.M.R. with the following exceptions - D.T.K. performed the sequence conservation analysis (Figure 1A) and the phylogenetic analysis (Figure 1B), Y.V.B. grew crystals, and C.M. collected X-ray diffraction data and solved the crystal structure; Resources, C.M. and Y.V.B.; Writing – Original Draft, A.M.R.; Writing – Review and Editing, A.M.R., D.T.K., C.M., and Y.V.B.; Visualization, A.M.R.; Supervision, C.M. and Y.V.B.; Funding Acquisition, A.M.R., C.M., and Y.V.B.

Declaration of Interests

The authors declare no competing interests.

Main Figure Titles and Legends

Figure 1. The SpmX is vertically inherited in Caulobacterales. (A) Schematic of SpmX architecture, including the conserved muramidase domain (see **Figure S1** for alignments), the variable intermediate domain, and two C-terminal transmembrane (TM) segments. Bar indicates amino acid sequence conservation among *spmX* alleles (see **Table S1** for a list of *spmX* genes used in this study). **(B)** Phylogenetic trees of representative species from Caulobacterales and other Alphaproteobacteria for concatenated housekeeping gene alignments (left) and for SpmX (right), with branch colors indicating the amino acid identity at position 20 of SpmX (D20L in yellow, D20R in red, and D20G in green). See **Table S6** for genome IDs. The concatenated housekeeping tree is fully supported with posterior probability of 1.0 for all clades. Asterisks indicate clades in the SpmX tree with posterior probabilities > 0.95. See **Figure S2** for the relationship of the SpmX muramidase domain within the lysozyme superfamily.

Figure 2. The SpmX muramidase domain retains the canonical GH motif but contains inactivating mutations in the catalytic cleft. (A) P22 lysozyme (PDB 2ANX) as a model lysozyme colored with rainbow gradient from blue N-terminus to red C-terminus. The catalytic glutamate appears in fuchsia and the GH beta-hairpin in light blue. **(B)** HMM logos of GH lysozymes made using WebLogo 3 [52]. Logos were constructed from protein sequences of **(i)** T4 lysozyme-like genes (n = 94), **(ii)** representative autolysins/endolysins from the Conserved Domain Database including P22 lysozyme (n = 20) but excluding SpmX genes, **(iii)** closest BLAST hits from non-SpmX muramidases (n = 60), and **(iv)** SpmX muramidases (n = 66), and organized in a cladogram to resemble the sequence cluster tree diagram in **Figure S2**. Amino acids are color-coded according to chemical properties, with uncharged polar residues in green, neutral residues in purple, basic residues in blue, acidic residues in red, and hydrophobic residues in black. The height of each letter is proportional to the relative frequency of a given identity and the height of the stack indicates the sequence conservation at that position. T4L numbering is used for ease of comparison. Asterisks mark positions critical for enzymatic activity and open circles mark positions associated with GH motif stability [26,27]. Refer to **Figure S1** for alignments of SpmX muramidases, which are listed in **Table S1.** See also **Table S5** for non-SpmX GH24 gene IDs.

Figure 3. The structure of SpmX muramidase domain has a wider, more dynamic catalytic cleft than related phage lysins. (A) Structural alignment of P22 lysozyme (PDB 2ANX, the model used for molecular replacement) in purple, R21 endolysin from P21 (PDB 2HDE, a distantly related GH24 T4L lysozyme) in navy blue, and SpmX-Mur-*Ae* in gold (PDB 6H9D). The catalytic glutamate is shown in red. Root mean square deviation (rmsd) 1.7 Å and 40% identity over 141 aligned Cα atoms, Dali Z-score 21.5 between P22 lysozyme and SpmX-Mur-*Ae*. See **Table S2** for data collection and refinement statistics. **(B)** Structural alignment of the three SpmX-Mur-*Ae* molecules, chains A (green), B (light blue), and C (dark blue), from the asymmetric unit. The surface of chain B is shown in partially transparent light blue. The double-headed arrow indicates the tilt of about 16° between the GH beta-hairpins of chains B and A. **(C)**

Overlays of ribbon diagrams and surfaces of P22 lysozyme (2ANX, left) and SpmX-Mur-*Ae* (6H9D, right) illustrating the conformation of the critical residues E11 (red), D20 (dark blue), R14 (yellow), and Y18 (orange). T4L numbering is used for ease of comparison. These structures have been rotated 180° around the y-axis from their representation in (A, B, D, E)**. (D)** Surface representation of P22 lysozyme (2ANX) with inset showing ribbon diagram and conformation of catalytic cleft with the canonical E11/D20/T26 catalytic triad. **(E)** Surface representation of SpmX-Mur-*Ae* (6H9D) with inset showing ribbon diagram and conformation of remodeled catalytic cleft with E11/D20L/T26M.

Figure 4. The D20L mutation attenuates P22 hydrolytic activity. (A) Remazol brilliant blue assays on *C. crescentus* sacculi using purified P22 lysozyme, P22 lysozyme D20L mutant, and *C. crescentus* SpmX muramidase. Active enzymes release peptidoglycan monomers covalently-bound to RBB into the supernatant that are detected by absorbance at 595 nm. Error bars are ± standard deviation for each normalized absorbance (n = 3). Lines are drawn to help guide the eye toward basic trends. Data points are from various days and sacculi preparations, but with internal normalization to Hen Egg White Lysozyme (HEWL). See **Figure S3** for peptidoglycan binding activity and **Figure S4AB** for SpmX mutant activity in RBB assays. **(B** and **C)** Growth curves of Lemo21(DE3) *E. coli* expressing P22 lysozyme (blue), P22 lysozyme D20L mutant (green), and *C. crescentus* SpmX muramidase (red). Proteins were expressed from pET22b with a N-terminal PelB signal sequence. In **(B)**, strains were grown in 5 mM rhamnose without IPTG for maximal repression of basal expression from the plasmids. In **(C)**, strains were grown without rhamnose and induced with 400 μM IPTG at the indicated time. See **Figure S4CDE** for enzymatic activity and periplasmic expression of SpmX-Mur and various mutants. **(D)** Phase/fluorescent overlays show live/dead staining of Lemo21(DE3) cells expressing P22Lyso-D20L and SpmX-Mur-*Cc* after four hours of induction. Green, membrane permeable SYTO 9 stains DNA in live cells and red, membrane impermeable propidium iodide nucleic acid dyes labels released nucleoids and DNA from lysed bacteria. The rounding of the *E. coli* in (i) is characteristic of spheroplast formation and lysis by hydrolytic activity on the cell wall. Scale bars are 5 μm.

Figure 5. Inactivating the muramidase domain partially delocalizes SpmX *in vivo*. Phase and fluorescent images of (**A**) *C. crescentus*, (**B**) A. *biprosthecum,* and **(C)** *A. excentricus.* In the top panel, phase images with derived schematics emphasizing stalks and morphologies are shown for (**i**) WT and (**ii**) Δ*spmX* cells. In **Aii**, *C. crescentus* cells exhibiting characteristic Δ*spmX* divisional defects are marked with asterisks and a cell growing stalks from both poles has its stalks marked with red arrowheads. Phase and fluorescent images of cells expressing (**iii**) SpmX-eGFP, (**iv**) SpmX-E11A-eGFP, or (**v**) SpmX-N105R-eGFP from the native chromosomal locus are shown in the lower panels. In **Aiv** and **Av**, cells with divisional defects are marked with white asterisks. In **Biii** and **Biv**, cells with one lateral or subpolar stalk are marked with white arrowheads. In **Civ**, cells with foci at the tips of stalks are marked with white arrowheads. All scale bars are 5 μm. See **Figure S5** for quantification of fluorescence and morphology data.

Figure 6. Removing or replacing the muramidase domain depletes native SpmX protein levels *in vivo*. (A) Phase and fluorescent images of strains in which the native *spmX* allele was replaced with the following gene fusions in the Δ*spmX* parent strain (ii): (i) *spmX-L20D-sfGFP* (iii) WT *spmX-sfGFP*, (iv) *spmX-E11A-sfGFP*, (v) *MurBs-*Δ*mur-SpmX-sfGFP* where *MurBs* is the muramidase domain from *Brevundimonas subvibrioides* SpmX, and (vi) *spmX-L20R-sfGFP*.

All scale bars are 5 μm. **(B)** Western blot comparing the Δ*spmX* parent strain to SpmX mutants and chimeras inserted at the *spmX* locus. In all cases, the primary antibody is directed against the C-terminal GFP fusion.

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yves Brun (ybrun@indiana.edu).

Experimental Model and Subject Details

E. coli strains were grown in LB as described in the Method Details sections concerning purification and periplasmic expression. All *C. crescentus*, *A. excentricus*, and *A. biprosthecum* strains used in this study were grown in liquid PYE medium. *C. crescentus* (CB15N/NA1000) was grown at 30°C, and *Asticcacaulis excentricus* (CB48/ATCC 15261) and *Asticcacaulis biprosthecucum* (C19/ ATCC 27554) species at 26°C. Strains were maintained on PYE plates supplemented with antibiotics as necessary (kanamycin 20 μg/mL, gentamycin 5μg/mL, and spectinomycin 100 μg/mL). For microscopy, *C. crescentus* and *A. excentricus* were inoculated from colonies, grown overnight, then diluted back 1:50 and grown for another 3-4 hours before being imaged in mid- to late-exponential phase. *A. biprosthecum* was inoculated from colonies and grown overnight to reach mid- to late-exponential phase for imaging. A detailed list of strains is included as **Table S4**.

Method Details

Bioinformatics and gene trees. Sequences of the SpmX genes in **Table S1** and members of the GH24 family were retrieved by BLAST searches on the Integrated Microbial Genomes and Microbiomes (IMG/M) database [53] and the National Center for Biotechnology Information (NCBI) "nr" database. See **Tables S1** and **S5** for lists of gene ID numbers. Multiple alignments were achieved with MUSCLE [54] and manually adjusted and visualized with Jalview [55]. Sequence conservation of SpmX residues was determined from the multiple sequence alignment of *spmX* alleles using ConSeq [56]. To improve visualization of conservation patterns, the ConSeq scores were averaged across a 20-residue sliding window. For estimating bacterial species phylogeny, assembled genome data were obtained from the genome database of the National Center for Biotechnology Information [57]. See **Table S6** for lists of genome IDs. Amino acid sequences of 37 conserved housekeeping genes were automatically identified, aligned, and concatenated using Phylosift [58]. All phylogenetic reconstruction was performed using MrBayes v3.2.6 [59] to estimate consensus phylogenies and clade posterior probability support values. Sequence substitution was modeled according to a WAG substation model with gamma-distributed rate variation between sites. Trees were visualized and formatted using iTol [60]. The sequence cluster tree was built with NCBI's Conserved Domain Database tool (CDD) [21,22]. This tool uses reverse position-specific BLAST, a method that compares query sequences to databases of position-specific score matrices and obtains *E*-values, such as in PSI-BLAST [21,22]. WebLogo3 was used to plot the amino acid distribution at each position of the GH motif [52]. To create the alignments for logo generation, 94 T4 lysozyme-like sequences, 20 endolysin/autolysins from the CDD analysis, 60 SpmX muramidase-like sequences (BLAST

hits), and 66 SpmX muramidase sequences were simultaneously aligned to T4 lysozyme. Only sequences with unambiguous alignment in the GH motif were included in this analysis.

Recombinant DNA methods. DNA amplification, Gibson cloning, and restriction digests were performed according to the manufacturer. Restriction enzymes and Gibson cloning mix were from New England Biolabs. Cloning steps were carried out in *E. coli* (alpha-select competent cells, Bioline) and plasmids were purified using Zyppy Plasmid Kits (Zymo Research Corporation). Sequencing was performed by the Indiana Molecular Biology Institute and Eurofins MWG Operon Technologies with double stranded plasmid or PCR templates, which were purified with a DNA Clean & Concentrator kits (Zymo Research Corporation). Chromosomal DNA was purified using the Bactozol Bacterial DNA Isolation Kit (Molecular Research Center). Plasmids were introduced into all *E. coli* strains using chemical transformation according to the manufacturer's protocols. Plasmids were introduced into *C. crescentus*, *A. excentricus*, and *A. biprosthecum* by electroporation based on previously published studies [61]. Briefly, for a given electroporation, 1 mL of culture in early stationary phase was pelleted at 4600 *xg* and washed twice with 1 mL of water. The pellet was resuspended in 50 µL water and placed in a 2.0 mm gap electroporation cuvette. 0.5-1 µg of DNA in 1-5 µL water was added before pulsing (2.5 kV, 25 µF, 200Ω). The cells were resuspended in 500 µL PYE, allowed to recover overnight, and plated the next day on selective plates. Allelic exchange in was achieved with pNPTS138, large genetic insertions with pMCS-2 [62], and eGFP insertional fusions with pGFPC-1 and pGFPC-2 [62].

Plasmid construction. *Expression plasmids*: *spmX* gene fragments encoding amino acids 2-150 of SpmX (SpmX-Mur) were amplified from genomic DNA and inserted into linearized expression vectors using Gibson cloning (NEB) according to manufacturers protocols. P22 lysozyme (P22Lyso) was amplified from a synthetic gene strand (Eurofins) for similar construction with Gibson cloning. For pTB147SUMO, the vector was linearized with SapI and XhoI to insert SpmX-Mur-*Ae*. For pET28a, the vector was linearized with NdeI and EcoRI to insert SpmX-Mur-*Cc*, BamHI and XhoI to insert SpmX-Mur-*Ae*, SacI to insert SpmX-Mur-*Ab,* and EcoRI to insert P22Lyso. In all pET28a plasmids, the constructs were cloned in frame with the N-terminal His-tag and a stop codon to eliminate the C-terminal His-tag. For pET22b, the vector was linearized with EcoRI and the C-terminal His-tag was preserved. Point mutants in expression vectors were obtained by using standard "quick-change" site-directed mutagenesis procedures and primers with 3' single stranded overhangs for increased efficiency.

Integrating plasmids for allelic exchange: For allelic exchange, the desired mutation was engineered into pNPTS138, bracketed by 1 kb up- and downstream of the corresponding genetic region. Integrants were isolated by antibiotic selection and secondary recombination events were selected by sucrose counter-selection using standard procedures. The resulting clones were confirmed by PCR and sequencing isolated genomic DNA.

For genomic deletions of *spmX* in *Asticcacaulis*, pNPTS138 was linearized with EcoRI and codons on either end of the gene were retained to avoid introducing frame-shifts in the surrounding area. Therefore the final gene deletion in *A. excentricus* lacks residues 5-808 and in *A. biprosthecum* lacks residues 5-815. For SpmX Δmur truncations, resides 2-150 were removed in all three species. In all cases pNPTS138 was linearized with EcoRI. Point mutations E19A and N91R were integrated into the Δ*spmX* background for ease of clone isolation and included full-length SpmX flanked by 1 kb genetic context. pNPTS138 containing mutated SpmX were

690 constructed using Gibson cloning with fragments on either side of the intended mutation and
691 overlapping primers containing the mutation amplified from genomic DNA. In all cases
692 pNPTS138 was linearized with EcoRV, except for SpmX*Ab*-E19A, where pNPTS138 was
693 linearized with EcoRI.
694 *Plasmids for insertional eGFP fusions*: The last 600 bp of *spmX* from *C. crescentus* was
695 amplified from genomic DNA and cloned into pGFPC-2 using Gibson cloning.
696 *Plasmids for integration at the* Δ*spmX locus*: These constructs were designed to allow
697 insertion of various SpmX mutants fused to C-terminal sfGFP into the Δ*spmX* locus in *C.*
698 *crescentus*. For SpmX-sfGFP and SpmX-E19A-sfGFP, fragments containing 1 kb of genomic
699 DNA upstream of *spmX* and *spmX* or *spmX-E19A* were amplified from existing pNPTS138
700 constructs and fused to a fragment containing monomeric sfGFP amplified from pSRKKm-Plac-
701 sf*gfp* [63] using Gibson cloning. In the final construct, SpmX and sfGFP are connected with the
702 linker sequence GSAGSAAGSGEF [64]. Chimeras with P22 lysozyme (P22Lyso) and its
703 catalytic mutant were made by Gibson cloning together fragments containing 1kb of upstream
704 genomic DNA, P22Lyso (with no stop codon), SpmXΔmur (residues 151-431) and sfGFP with
705 the same linker. P22Lyso and P22Lyso-E11A were amplified from pET28a plasmids containing
706 these genes. Chimeras with SpmX muramidase from *Brevundimonas subvibrioides* (residues 1-
707 140) and *Parvularcula bermudensis* (residues 1-168) were similarly made with the muramidase
708 fragments amplified from genomic DNA and synthetic gene strands (Eurofins), respectively.
709
710 **Production of SpmX-Mur-Ae for crystallography.** The muramidase domain of SpmX from *A.*
711 *excentricus* (SpmX-Mur-*Ae*, residues 2-150) was fused to a hexahistidine tag followed by the
712 SUMO cleavage site of the Ulp1 protease (His-SUMO tag) [65] and overexpressed in *E. coli*
713 BL21 (DE3) RIL cells. Cells were grown at 37 °C in 2 l of Terrific Broth (BD Biosciences)
714 supplemented with ampicillin (100 μg/mL) until the OD_{600nm} reached 0.8. Production of the
715 recombinant protein was induced by the addition of isopropyl β-D-1-thiogalactopyranoside
716 (IPTG) to 0.5 mM after the culture was cooled to 25°C. Cell growth was continued overnight at
717 25°C, and cells were harvested by centrifugation. Cell pellets were resuspended in $1/20^{th}$ volume
718 of buffer A (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 25 mM imidazole, 10% (vol/vol)
719 glycerol) containing the Complete™ cocktail of protease inhibitors (Roche). Cells were lysed by
720 six passages through a cell disruptor (Constant Systems Limited) at 20 kPsi, and cell debris were
721 pelleted by centrifugation at 40,000 × g for 30 min at 4 °C. The centrifugation supernatant was
722 loaded on a Ni-NTA agarose resin (Qiagen) equilibrated with buffer A. After extensive washing
723 with buffer A, His-SUMO-SpmX-Mur-*Ae* was eluted with a linear 0-100% gradient of buffer B
724 (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 500 mM imidazole, 10% (vol/vol) glycerol) over 10
725 column volumes. Peak fractions were pooled, mixed with a 1:100 dilution of a His-tagged Ulp1
726 (SUMO) protease preparation [66] and dialyzed overnight at 4°C in buffer C (50 mM Tris-HCl
727 (pH 8.0), 300 mM NaCl, 10% (vol/vol) glycerol). Cleavage reactions were passed through Ni-
728 NTA resin to remove free His-SUMO tag and His-Ulp1, and untagged protein was collected in
729 the flow through. Flow-through fractions were concentrated with Amicon Ultra Centrifugal filter
730 units with a molecular weight cutoff of 10 kDa (Millipore) and were injected onto an ENrich™
731 SEC650 10x300 gel-filtration column (Biorad). SpmX-Mur-*Ae* was eluted with buffer D (25 mM
732 Tris-HCl (pH 8.0), 150 mM NaCl) and again concentrated with Amicon Ultra Centrifugal filter
733 units. Protein concentration was measured using absorbance at 280 nm.
734

Protein crystallization and structure determination. High-throughput crystallization trials were performed with a Cartesian PixSys 4200 crystallization robot (Genomic Solutions, U.K.). Hanging drops containing 100 nL of protein (25 or 12.5 mg/mL) and 100 nL of reservoir solution were set up in 96-well Crystal Quick plates (Greiner) and incubated at 20°C. Initial crystal hits were refined manually by setting up hanging drops containing 1 µL of protein (25 or 12.5 mg/mL) and 1 µL of reservoir solution in 24-well plates (Molecular Dimensions) incubated at 20°C. Large needle-shaped crystals (dimensions of about 40 x 40 x 400 µm) were finally obtained for SpmX-Mur-*Ae* in 0.1 M Tris-HCl pH 8.5, 12% PEG 3350, 0.2 M $MgCl_2$, at 20°C within 24–48 h. SpmX-Mur-*Ae* crystals were cryoprotected by transfer into 0.1 M Tris-HCl pH 8.5, 13% PEG 3350, 0.2 M $MgCl_2$, 10% glycerol, and then flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on the ID30a1 (MASSIF-1) beamline [67,68].

Diffraction data were indexed and scaled using the XDS program suite [69]. SpmX-Mur-*Ae* crystals belong to the trigonal space group $P3_221$, with unit cell dimensions of 100.44 x 100.44 x 96.62 Å and three molecules per asymmetric unit. Phase determination was carried out by the molecular replacement method with PHASER [70], using as a search model the structure of the phage P22 lysozyme (PDB entry 2ANX)) to 1.9 Å resolution (*R*cryst 21.1%, *R*free 25.5%,) (**Table S2**). The molecular replacement solution model was rebuilt de novo using PHENIX [71] to prevent bias from the model.

The structure of SpmX-Mur-*Ae* was completed by cycles of manual building with COOT [72] and addition of water molecules with ARP/wARP [73]. Several cycles of manual building and refinement with REFMAC [74], as implemented in the CCP4 program suite, were performed until R_{work} and R_{free} converged [75]. Stereochemical verification was performed with PROCHECK [76]. The secondary structure assignment was verified with DSSP [77], with all residues within most favorable or allowed regions of the Ramachandran plot. Figures were generated with PyMol (http://www.pymol.org). Coordinates of the final refined model were deposited at the Protein Data Bank (PDB, http://www.rcsb.org) and were assigned PDB entry code 6H9D. The data collection and refinement statistics are summarized in **Table S2**.

Protein production for *in vitro* assays. Fresh BL21(DE3) competent cells (Novagen) were transformed with pET28a constructs containing various muramidase genes with N-terminal His-tags and grown overnight in LB with 1% glucose and 50 µg/mL kanamycin. Overnight cultures were diluted 100-fold in LB medium with 1% glucose and 50 µg/mL kanamycin. Typically 500 mL cultures of cells were grown for 1.5-2 hours to an OD600 of 0.6-0.7 and shifted to 20°C. When the OD600 reached 0.8–0.9, the cells were induced with 0.5 mM IPTG. After growing for 4 h at 20 °C, cells were harvested and resuspended in 30 mL lysis buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 20 mM imiadazole, 5 mM BME) with a EDTA-free Protease Inhibitor Mini Tablet (Pierce) and phenylmethanesulfonyl fluoride (PMSF, 1 mM). The 30 mL cell mixture was lysed on ice using a sonicating horn and spun down at 10,000g for 20-30 min. The clarified lysate was loaded onto a 5 ml HiTrap Chelating HP cartridge (GE Healthcare) charged with Ni^{2+} and pre-equilibrated with lysis buffer. After loading, the column was washed with lysis buffer followed by an elution via a 0-100% linear gradient of buffer B (25 mM HEPES pH 7.5, 100 mM NaCl, 500 mM imidazole, 2 mM BME). Muramidase-containing fractions were pooled based on SDS-PAGE analysis and concentrated to 2.5 mL. Imidizole was removed by passing the concentrated fraction over a PD10 desalting column (GE Healthcare) equilibrated with 25 mM HEPES pH 7.5, 100 mM NaCl, 2 mM BME.

Sacculi preparation, RBB labeling, and calibration. Sacculi were prepared from all species in the same manner. For a typical 2L prep, cells were grown to an OD of 0.5-1 in their respective medium (see culturing details in Experimental Model and Subject Details) and harvested by centrifugation at 6,000*g* for 20 minutes. *C. crescentus* cells usually required multiple centrifugation steps to collect all the cells. Cells were resuspended in 25 mL water (or PBS for *E. coli*) and added drop-wise into 50 mL of boiling 7.5% SDS under stirring. The mixture was boiled for 30 minutes and then allowed to cool to room temperature. Sacculi were then pelleted by ultracentrifugation at 100,000*g* for 30 minutes at room temperature. The resulting pellets were resuspended in 100 mL pure water, and washed repeatedly until SDS was no longer detected in the supernatant. The pellet was confirmed to be clear of SDS by mixing 0.2 mL of the supernatant with 1 µL 0.5% methylene blue, 0.1 mL 0.7M $NaPO_4$ pH 7.2, and 0.6 mL chloroform and checking to make sure that, after vortexing and allowing to settle, the solution had an upper blue phase and a lower clear phase [78]. At this point, the pellets were resuspended in 10 mL PBS with 20 mM $MgSO_4$, 250 U/µL Pierce Universal Nuclease (Thermo Fisher Scientific), and 10 mg/mL amylase (Sigma). The mixture was incubated at 37°C for 1-4 hours. Afterwards, 10 mg/mL trypsin and 10 mM $CaCl_2$ was added and the mixture incubated overnight at 37°C. 800 µL of 15% SDS were then added to the mixture and it was brought to a boil for about 10 minutes and allowed to cool to room temperature. The sample was then pelleted (100,000*g*, 30 min, room temperature) and resuspended in 4-5 wash steps until SDS was no longer detected. The final pellet was then resuspended in 2 mL water and added to 0.8 mL 0.2M remazol brilliant blue (Sigma), 0.4 mL 5M NaOH, and additional water to 8 mL. The mixture was incubated, shaking, overnight at 37°C. After neutralizing the solution with 0.4 mL 5M HCl, the mixture was then pelleted (21,000*g,* 20 minutes, room temperature), and resuspended in water until the supernatant became clear.

To calibrate the concentration of RBB-labeled sacculi for dye-release assays and peptidoglycan-binding assays, activity curves with Hen Egg White Lysozyme (HEWL, Sigma) were produced using different dilutions of the RBB-labeled sacculi. The RBB-labeled sacculi were used at the dilution that resulted in an A595 of 0.5 when 5 µL of the RBB-labeled sacculi were incubated with 4 uM HEWL.

Peptioglycan-binding assays. 5µL of calibrated RBB-labeled sacculi were incubated with 1 µM of purified protein (protein constructs used are shown in Figure S4) in PBS pH 7.4 to a final volume of 50 µL for 30 minutes at 37°C and then pelleted (16,000*g*, 20 minutes). Fractions were separated and the pellet resuspended in 50 µL PBS. 10 µL of each fraction was loaded onto Any kD Mini-PROTEAN TGX Precast Protein Gels (BioRad) to visualize whether the protein associated with the insoluble sacculi fraction. BSA (Sigma) was used at 1 µM as a negative control.

Remazol brilliant blue dye-release assays. Methods were adapted from [79,80]. Assays were carried out in 25-µL reactions using 25 mM HEPES pH 7.5, 100 mM NaCl and 5 µL of calibrated RBB-labeled sacculi. Enzymes were added at various concentrations (see Figs. 4 and S4) and incubated overnight at 37°C. Reactions were then centrifuged for 20 minutes at 16,000*g,* and the supernatant carefully separated from the pellet. Final values in Figure 4 and S4 are normalized against absorbances measured for reactions with HEWL that were run in tandem for every measurement to correct for differences in different sacculi preparations.

Fluorescence microscopy and image preparation. Fluorescence imaging was done using an inverted Nikon Ti-E microscope using a Plan Apo 60X 1.40 NA oil Ph3 DM objective with a GFP/Cy3 filter cube and an Andor DU885 EM CCD camera. Images were captured using NIS Elements (Nikon). Cells were mounted on 1% (w/v) agarose pads made with PYE (or PBS, in the case of *E. coli*) for imaging. In general, the fluorescent channel of each image was background subtracted and a Gaussian Blur filter was applied using Fiji [81].

Western blots. Strains were grown to saturation (overnight for *C. crescentus* and *A. excentricus*, usually 48 hours for *A. biprosthecum*). OD600 was determined and cells were collected at a normalized density of OD600 = 1/1mL. 1 mL of each normalized culture was pelleted, resuspended in 100 µL water, and prepared for analysis using standard procedures using SDS-PAGE, transfer, and western blotting. 10 µL of each sample was loaded onto Any kD Mini-PROTEAN TGX Precast Protein Gels (BioRad). The JL-8 monoclonal GFP antibody (Clontech) was used as the primary antibody and Goat Anti-mouse HRP (Pierce) was used for the secondary antibody. Transferred blots were visualized with SuperSignal West Dura Extended Duration HRP substrate (ThermoFisher Scientific) using a Bio-Rad Chemidoc.

Periplasmic expression in *E. coli*. Fresh Lemo(DE3) competent cells (NEB) were transformed with pET22b constructs containing various muramidase genes with N-terminal H-tags and plated. Lemo21(DE3) carries a rhamnose-inducible copy of LysY that inhibits T7 polymerase and allows for tunable dampening of expression of toxic products. We could not transform expression strains BL21(DE3) or Tuner(DE3) with the pET22b-P22Lyso construct, but were able to isolate a few transformants carrying this construct using Lemo21(DE3) cells under high rhamnose repression (2 mM). P22Lyso-D20L, and all the SpmX-Mur-*Cc* constructs, efficiently transformed into all expression strains tested, and could be carried by Lemo21(DE3) without rhamnose.

In the case of pET22b-P22Lyso, where cells eventually lyse from leak, cell cultures were grown directly from colonies in the presence of 5 mM rhamnose and monitored over time. Figure 4A shows the same treatment for all tested constructs. For testing induction of CCM and P22Lyso-D20L, colonies were grown overnight in LB with 100 µg/mL carbenicillin and 30 µg/mL chloramphenicol. Overnight cultures were diluted 50-fold in LB medium with 100 µg/mL carbenicillin and 30 µg/mL chloramphenicol. In experiments cases rhamnose was added at specified concentrations. Typically 4 mL cultures of cells were grown for 1-1.5 hours to an OD600 of 0.3-0.4, induced with 400 uM IPTG, and shifted to 20°C.

Growth curves and live-dead staining: Optical densities were measured over time and cells were routinely checked for lysing by microscopy using standard procedures. Briefly, 1 µL of a 1:1 mixture of solutions A and B from a LIVE/DEAD *Bac*Light Bacterial Viability Kit (ThermoFisher Scientific) was directly added to 100 µL of cells diluted 1:10 in PBS. Cells were visualized on 1% agar pads made with PBS using the methods described in microscopy.

Periplasmic expression levels: After growing for 4 h at 20 °C, OD600 was determined and cells were collected at a normalized density of OD600 = 1/1mL. One mL of the normalized sample was pelleted at 4000g for 15 min and the pellet resuspended in 250 µL 20% sucrose, 1 mM EDTA, 30 mM TRIS pH 8 at room temperature. The sample was mixed gently by rotation at room temperature for 10 minutes before being spun down at 13,000g for 10 minutes. The supernatant was carefully removed and the pellet rapidly suspended in 250 µL ice-cold pure

water. The sample was mixed gently by rotation at 4°C for 10 minutes before being spun down at 13,000g at 4°C. The supernatant (periplasmic fraction) and pellet (cell fraction) were then separated and prepared for analysis using standard procedures using SDS-PAGE, transfer, and western blotting. Blots were incubated with His-Probe Antibody (H-3) sc-8136 HRP (Santa Cruz Biotechnology) and visualized with SuperSignal West Dura Extended Duration HRP substrate (ThermoFisher Scientific) using a Bio-Rad Chemidoc.

Quantification and Statistical Analysis

HHM logo generation. WebLogo3 was used to plot the amino acid distribution at each position of the GH motif [52]. To create the alignments for logo generation, 94 T4 lysozyme-like sequences, 20 endolysin/autolysins from the CDD analysis, 60 SpmX muramidase-like sequences (BLAST hits), and 66 SpmX muramidase sequences were simultaneously aligned to T4 lysozyme. See **Table S5** for the genes used in logo construction. Because only sequences with unambiguous alignment in the GH motif were included in this analysis, many endolysin/autolysins from the CDD analysis had to be excluded from the logo. T4 lysozyme-like sequences were chosen from BLAST hits from various bacterial prophage sources.

RBB assays. RBB assays were carried out in triplicate. The averages of the replicates were divided by the average measurement for Hen Egg White Lysozyme (HEWL) activity to obtain the normalized values plotted in the activity curves. The data points plotted in Figure 4A represent measurements of P22-lyso and SpmX-Mur-*Cc* activity over several days. Lines are drawn to help guide the eye toward basic trends and do not reflect line fitting. Data points are from various days and sacculi preparations, but are normalized to parallel reactions with HEWL on the same day and with the same sacculi as each data point.

Fluorescent microscopy. Quantification of stalk morphotypes and stalks with multiple foci was done by hand using tools in Fiji. Quantification of fluorescence data was achieved using MicrobeJ [82]. Mean stalk intensity was measured in *A. biprosthecum* cells by using Fiji to draw line ROIs that did not overlap with the focus at the base of the stalk, measuring mean fluorescence along the ROI. Figures and statistics were performed using GraphPad Prism version 8.00 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com. The statistical details describing the quantification of cell morphology and fluorescent image analysis in Figure S6 can be found in the figure legend.

Data and Software Availability

The accession number for the atomic coordinates and structure factors reported in this paper are PDB: 6H9D. The data collection and refinement statistics are summarized in **Table S2**.

References

1. Hall, J.P.J., Brockhurst, M.A., and Harrison, E. (2017). Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. Phil Trans R Soc B *372*, 20160424.

2. Koonin, E.V. (2016). Viruses and mobile elements as drivers of evolutionary transitions. Phil Trans R Soc B *371*, 20150442.

3. Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A.A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat. Rev. Microbiol. *13*, 641–650.

4. Harrison, E., and Brockhurst, M.A. (2017). Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. BioEssays *39*, 1700112.

5. Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. *11*, 1511–1520.

6. Menouni, R., Hutinet, G., Petit, M.-A., and Ansaldi, M. (2015). Bacterial genome remodeling through bacteriophage recombination. FEMS Microbiol. Lett. *362*, 1–10.

7. Bobay, L.-M., Touchon, M., and Rocha, E.P.C. (2014). Pervasive domestication of defective prophages by bacteria. Proc. Natl. Acad. Sci. *111*, 12127–12132.

8. Touchon, M., Moura de Sousa, J.A., and Rocha, E.P. (2017). Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Curr. Opin. Microbiol. *38*, 66–73.

9. Wang, X., Kim, Y., Ma, Q., Hong, S.H., Pokusaeva, K., Sturino, J.M., and Wood, T.K. (2010). Cryptic prophages help bacteria cope with adverse environments. Nat. Commun. *1*, 147.

10. Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. Proc. Natl. Acad. Sci. *106*, 4154–4159.

11. Sarris, P.F., Ladoukakis, E.D., Panopoulos, N.J., and Scoulica, E.V. (2014). A Phage Tail-Derived Element with Wide Distribution among Both Prokaryotic Domains: A Comparative Genomic and Phylogenetic Study. Genome Biol. Evol. *6*, 1739–1747.

12. Ghequire, M.G.K., and De Mot, R. (2015). The Tailocin Tale: Peeling off Phage Tails. Trends Microbiol. *23*, 587–590.

13. Hockett, K.L., Renner, T., and Baltrus, D.A. (2015). Independent Co-Option of a Tailed Bacteriophage into a Killing Complex in Pseudomonas. mBio *6*, e00452-15.

14. Scholl, D. (2017). Phage Tail–Like Bacteriocins. Annu. Rev. Virol. *4*, 453–467.

15. Radhakrishnan, S.K., Thanbichler, M., and Viollier, P.H. (2008). The dynamic interplay between a cell fate determinant and a lysozyme homolog drives the asymmetric division cycle of Caulobacter crescentus. Genes Dev. *22*, 212–225.

16. Lasker, K., Mann, T.H., and Shapiro, L. (2016). An intracellular compass spatially coordinates cell cycle modules in Caulobacter crescentus. Curr. Opin. Microbiol. *33*, 131–139.

17. Jiang, C., Brown, P.J.B., Ducret, A., and Brun, Y.V. (2014). Sequential evolution of bacterial morphology by co-option of a developmental regulator. Nature *506*, 489–493.

18. Perez, A.M., Mann, T.H., Lasker, K., Ahrens, D.G., Eckart, M.R., and Shapiro, L. (2017). A Localized Complex of Two Protein Oligomers Controls the Orientation of Cell Polarity. mBio *8*, e02238-16.

19. Bowman, G.R., Comolli, L.R., Gaietta, G.M., Fero, M., Hong, S.-H., Jones, Y., Lee, J.H., Downing, K.H., Ellisman, M.H., McAdams, H.H., *et al.* (2010). Caulobacter PopZ forms a polar subdomain dictating sequential changes in pole composition and function. Mol. Microbiol. *76*, 173–189.

20. Holmes, J.A., Follett, S.E., Wang, H., Meadows, C.P., Varga, K., and Bowman, G.R. (2016). Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. Proc. Natl. Acad. Sci. *113*, 12490–12495.

21. Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., *et al.* (2007). CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. *35*, D237–D240.

22. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., *et al.* (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.

23. Blackburn, N.T., and Clarke, A.J. (2001). Identification of Four Families of Peptidoglycan Lytic Transglycosylases. J. Mol. Evol. *52*, 78–84.

24. Evrard, C., Fastrez, J., and Declercq, J.-P. (1998). Crystal structure of the lysozyme from bacteriophage lambda and its relationship with V and C-type lysozymes11Edited by K. Nagai. J. Mol. Biol. *276*, 151–164.

25. Vollmer, W., Joris, B., Charlier, P., and Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. FEMS Microbiol. Rev. *32*, 259–286.

26. Wohlkönig, A., Huet, J., Looze, Y., and Wintjens, R. (2010). Structural Relationships in the Lysozyme Superfamily: Significant Evidence for Glycoside Hydrolase Signature Motifs. PLOS ONE *5*, e15388.

27. Rennell, D., Bouvier, S.E., Hardy, L.W., and Poteete, A.R. (1991). Systematic mutation of bacteriophage T4 lysozyme. J. Mol. Biol. *222*, 67–88.

28. Anand, N.N., Stephen, E.R., and Narang, S.A. (1988). Mutation of active site residues in synthetic T4-lysozyme gene and their effect on lytic activity. Biochem. Biophys. Res. Commun. *153*, 862–868.

29. Sun, Q., Kuty, G.F., Arockiasamy, A., Xu, M., Young, R., and Sacchettini, J.C. (2009). Regulation of a muralytic enzyme by dynamic membrane topology. Nat. Struct. Mol. Biol. *16*, 1192–1194.

30. Hardy, L.W., and Poteete, A.R. (1991). Reexamination of the role of Asp20 in catalysis by bacteriophage T4 lysozyme. Biochemistry *30*, 9457–9463.

31. Poteete, A.R., Dao-Pin, S., Nicholson, H., and Mathews, B.W. (1991). Second-site revertants of an inactive T4 lysozyme mutant restore activity by restructuring the active site cleft. Biochemistry *30*, 1425–1432.

32. Anderson, W.F., Grütter, M.G., Remington, S.J., Weaver, L.H., and Matthews, B.W. (1981). Crystallographic determination of the mode of binding of oligosaccharides to T4 bacteriophage lysozyme: Implications for the mechanism of catalysis. J. Mol. Biol. *147*, 523–543.

33. Braga, L.P.P., Soucy, S.M., Amgarten, D.E., Silva, D., M, A., and Setubal, J.C. (2018). Bacterial Diversification in the Light of the Interactions with Phages: The Genetic Symbionts and Their Role in Ecological Speciation. Front. Ecol. Evol. *6*. Available at: https://www.frontiersin.org/articles/10.3389/fevo.2018.00006/full [Accessed August 30, 2018].

34. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillman, M.-L., and Brüssow, H. (2003). Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. *6*, 417–424.

35. Cortez, D., Forterre, P., and Gribaldo, S. (2009). A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biol. *10*, R65.

36. Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli., Bacterial Genomes as New Gene Homes: The Genealogy of ORFans in E. coli. Genome Res. Genome Res. *14*, *14*, 1036, 1036–1042.

37. Brézellec, P., Vallet-Gely, I., Possoz, C., Quevillon-Cheruel, S., and Ferat, J.-L. (2016). DciA is an ancestral replicative helicase operator essential for bacterial replication initiation. Nat. Commun. *7*, 13271.

38. Brézellec, P., Petit, M.-A., Pasek, S., Vallet-Gely, I., Possoz, C., and Ferat, J.-L. (2017). Domestication of Lambda Phage Genes into a Putative Third Type of Replicative Helicase Matchmaker. Genome Biol. Evol. *9*, 1561–1566.

39. Forterre, P. (1999). Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol. Microbiol. *33*, 457–465.

40. Sabehi, G., Shaulov, L., Silver, D.H., Yanai, I., Harel, A., and Lindell, D. (2012). A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. Proc. Natl. Acad. Sci. U. S. A. *109*, 2037–2042.

1023 41. Lang, A.S., Zhaxybayeva, O., and Beatty, J.T. (2012). Gene transfer agents: phage-like
1024 elements of genetic exchange. Nat. Rev. Microbiol. *10*, 472–482.

1025 42. Lang, A.S., and Beatty, J.T. (2007). Importance of widespread gene transfer agent genes in
1026 α-proteobacteria. Trends Microbiol. *15*, 54–62.

1027 43. Shakya, M., Soucy, S.M., and Zhaxybayeva, O. (2017). Insights into origin and evolution of
1028 α-proteobacterial gene transfer agents. Virus Evol. *3*. Available at: https://academic-oup-
1029 com.proxyiub.uits.iu.edu/ve/article/3/2/vex036/4705971 [Accessed August 31, 2018].

1030 44. Ho, B.T., Dong, T.G., and Mekalanos, J.J. (2014). A View to a Kill: The Bacterial Type VI
1031 Secretion System. Cell Host Microbe *15*, 9–21.

1032 45. Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014). Type VI secretion system effectors:
1033 poisons with a purpose. Nat. Rev. Microbiol. *12*, 137–148.

1034 46. Hurst, M.R.H., Glare, T.R., and Jackson, T.A. (2004). Cloning Serratia entomophila
1035 Antifeeding Genes—a Putative Defective Prophage Active against the Grass Grub Costelytra
1036 zealandica. J. Bacteriol. *186*, 5116–5128.

1037 47. Yang, G., Dowling, A.J., Gerike, U., ffrench-Constant, R.H., and Waterfield, N.R. (2006).
1038 Photorhabdus Virulence Cassettes Confer Injectable Insecticidal Activity against the Wax
1039 Moth. J. Bacteriol. *188*, 2254–2261.

1040 48. Shikuma, N.J., Pilhofer, M., Weiss, G.L., Hadfield, M.G., Jensen, G.J., and Newman, D.K.
1041 (2014). Marine Tubeworm Metamorphosis Induced by Arrays of Bacterial Phage Tail–Like
1042 Structures. Science *343*, 529–533.

1043 49. Patzer, S.I., Albrecht, R., Braun, V., and Zeth, K. (2012). Structural and Mechanistic Studies
1044 of Pesticin, a Bacterial Homolog of Phage Lysozymes. J. Biol. Chem. *287*, 23381–23396.

1045 50. Michalska, K., Brown, R.N., Li, H., Jedrzejczak, R., Niemann, G.S., Heffron, F., Cort, J.R.,
1046 Adkins, J.N., Babnigg, G., and Joachimiak, A. (2013). New sub-family of lysozyme-like
1047 proteins shows no catalytic activity: crystallographic and biochemical study of STM3605
1048 protein from Salmonella Typhimurium. J. Struct. Funct. Genomics *14*, 1–10.

1049 51. Ren, Q., Wang, C., Jin, M., Lan, J., Ye, T., Hui, K., Tan, J., Wang, Z., Wyckoff, G.J., Wang,
1050 W., *et al.* (2017). Co-option of bacteriophage lysozyme genes by bivalve genomes. Open
1051 Biol. *7*, 160285.

1052 52. Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence
1053 Logo Generator. Genome Res. *14*, 1188–1190.

1054 53. Chen, I.-M.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A.,
1055 Huang, J., Andersen, E., Huntemann, M., *et al.* (2017). IMG/M: integrated genome and
1056 metagenome comparative data analysis system. Nucleic Acids Res. *45*, D507–D516.

54. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

55. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

56. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38*, W529–W533.

57. Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. *42*, D553–D559.

58. Darling, A.E., Jospin, G., Lowe, E., Iv, F.A.M., Bik, H.M., and Eisen, J.A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ *2*, e243.

59. Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics *19*, 1572–1574.

60. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44*, W242–W245.

61. Gonin, M., Quardokus, E.M., O'Donnol, D., Maddock, J., and Brun, Y.V. (2000). Regulation of Stalk Elongation by Phosphate inCaulobacter crescentus. J. Bacteriol. *182*, 337–347.

62. Thanbichler, M., Iniesta, A.A., and Shapiro, L. (2007). A comprehensive set of plasmids for vanillate- and xylose-inducible gene expression in Caulobacter crescentus. Nucleic Acids Res. *35*, e137–e137.

63. Figueroa-Cuilan, W., Daniel, J.J., Howell, M., Sulaiman, A., and Brown, P.J.B. (2016). Mini-Tn7 Insertion in an Artificial attTn7 Site Enables Depletion of the Essential Master Regulator CtrA in the Phytopathogen Agrobacterium tumefaciens. Appl Env. Microbiol *82*, 5015–5025.

64. Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. *17*, 691–695.

65. Marblestone, J.G., Edavettal, S.C., Lim, Y., Lim, P., Zuo, X., and Butt, T.R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. Protein Sci. Publ. Protein Soc. *15*, 182–189.

66. Uehara, T., Parzych, K.R., Dinh, T., and Bernhardt, T.G. (2010). Daughter cell separation is controlled by cytokinetic ring□activated cell wall hydrolysis. EMBO J. *29*, 1412–1422.

67. Bowler, M.W., Nurizzo, D., Barrett, R., Beteva, A., Bodin, M., Caserotto, H., Delagenière, S., Dobias, F., Flot, D., Giraud, T., *et al.* (2015). MASSIF-1: a beamline dedicated to the

1092 fully automatic characterization and data collection from crystals of biological
1093 macromolecules. J. Synchrotron Radiat. *22*, 1540–1547.

1094 68. Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D., and Bowler, M.W. (2015). Fully
1095 automatic characterization and data collection from crystals of biological macromolecules.
1096 Acta Crystallogr. D Biol. Crystallogr. *71*, 1757–1767.

1097 69. Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

1098 70. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read,
1099 R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

1100 71. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H.,
1101 Hung, L.-W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure
1102 refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr.
1103 D Biol. Crystallogr. *64*, 61–69.

1104 72. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta
1105 Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

1106 73. Langer, G.G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated
1107 macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat.
1108 Protoc. *3*, 1171–1179.

1109 74. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A.,
1110 Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of
1111 macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

1112 75. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of
1113 crystal structures. Nature *355*, 472–475.

1114 76. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a
1115 program to check the stereochemical quality of protein structures. J Appl Crystallogr, 283–
1116 291.

1117 77. Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern
1118 recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577–2637.

1119 78. Hayashi, K. (1975). A rapid determination of sodium dodecyl sulfate with methylene blue.
1120 Anal. Biochem. *67*, 503–506.

1121 79. Uehara, T., Parzych, K.R., Dinh, T., and Bernhardt, T.G. (2010). Daughter cell separation is
1122 controlled by cytokinetic ring-activated cell wall hydrolysis. EMBO J. *29*, 1412–1422.

1123 80. Zhou, R., Chen, S., and Recsei, P. (1988). A dye release assay for determination of
1124 lysostaphin activity. Anal. Biochem. *171*, 141–144.

81. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

82. Ducret, A., Quardokus, E.M., and Brun, Y.V. (2016). MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat. Microbiol. *1*, 16077.

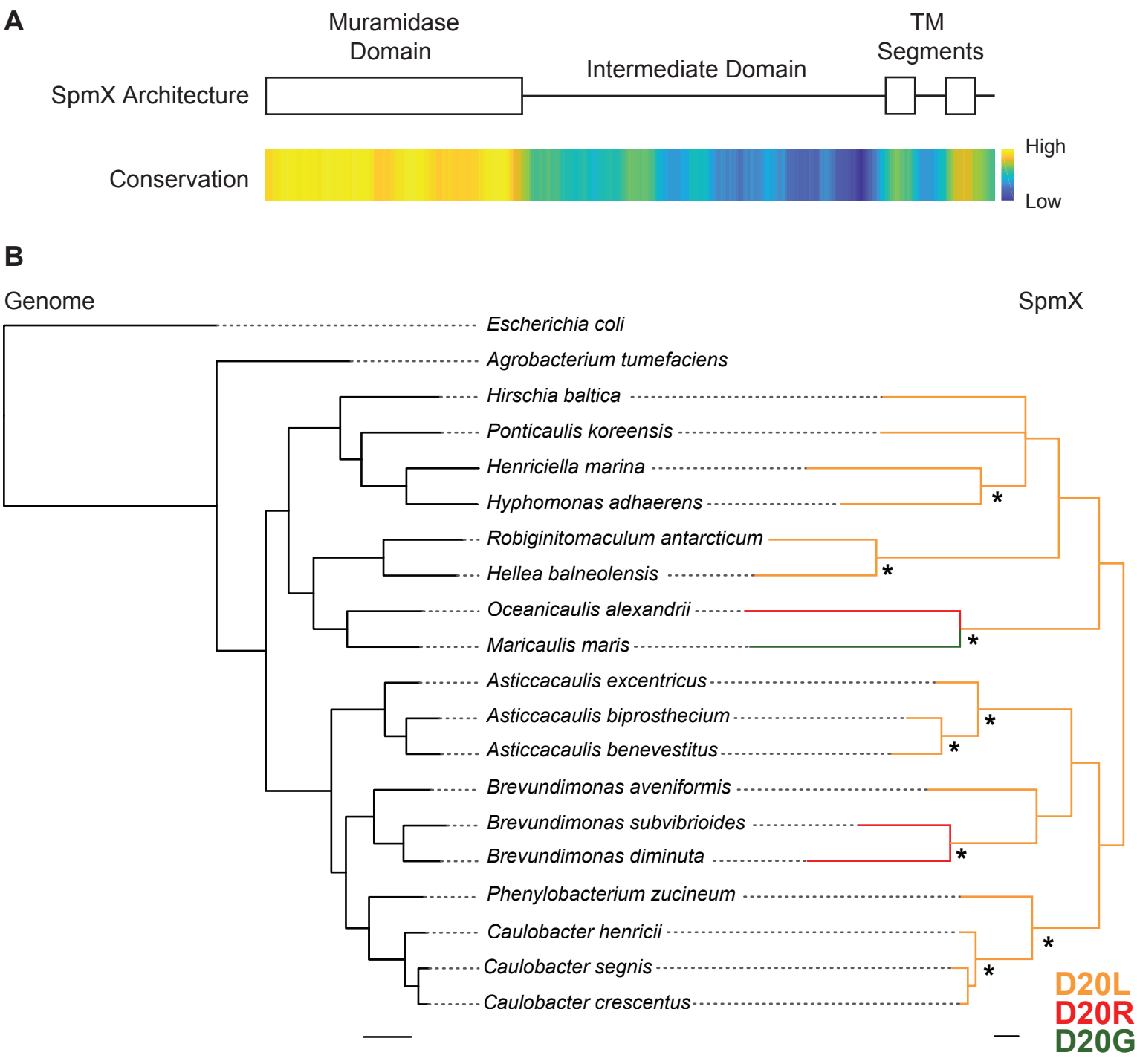
83. Evinger, M., and Agabian, N. (1977). Envelope-associated nucleoid from Caulobacter crescentus stalked and swarmer cells. J. Bacteriol. 132, 294–301.

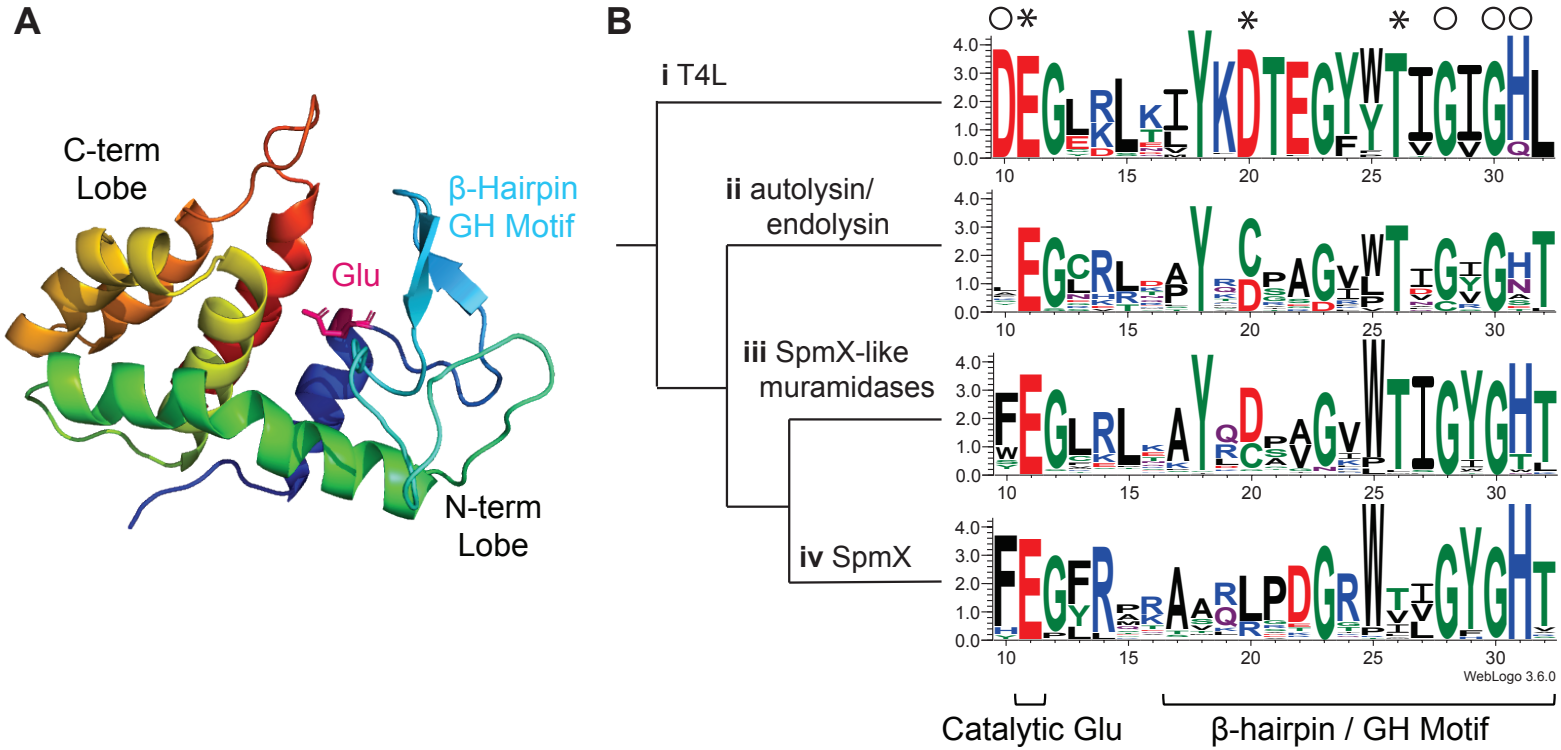
84. Poindexter, J.S. (1964). BIOLOGICAL PROPERTIES AND CLASSIFICATION OF THE CAULOBACTER GROUP1. Bacteriol. Rev. 28, 231–295.

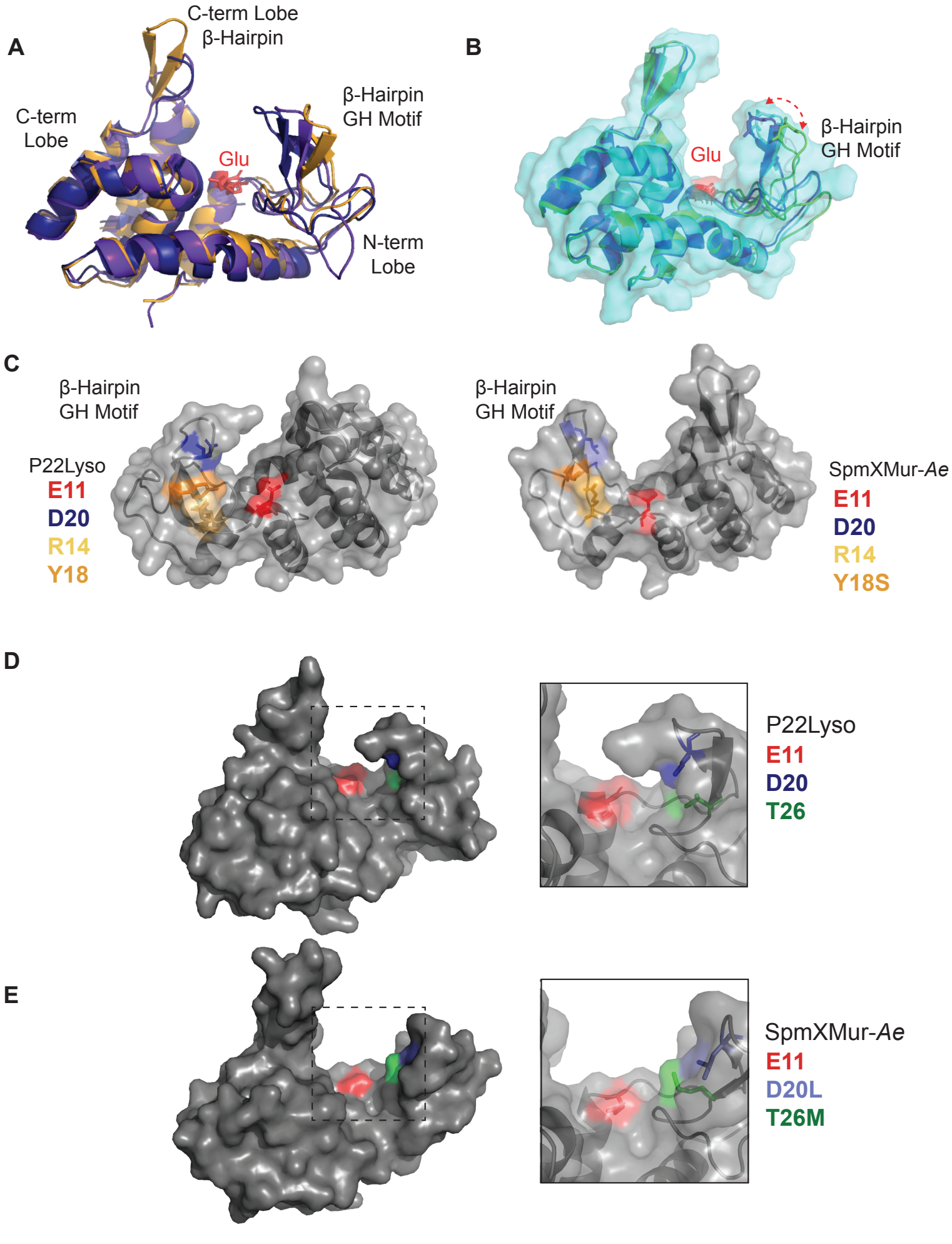
85. Pate, J.L., and Ordal, E.J. (1965). THE FINE STRUCTURE OF TWO UNUSUAL STALKED BACTERIA. J. Cell Biol. 27, 133–150.

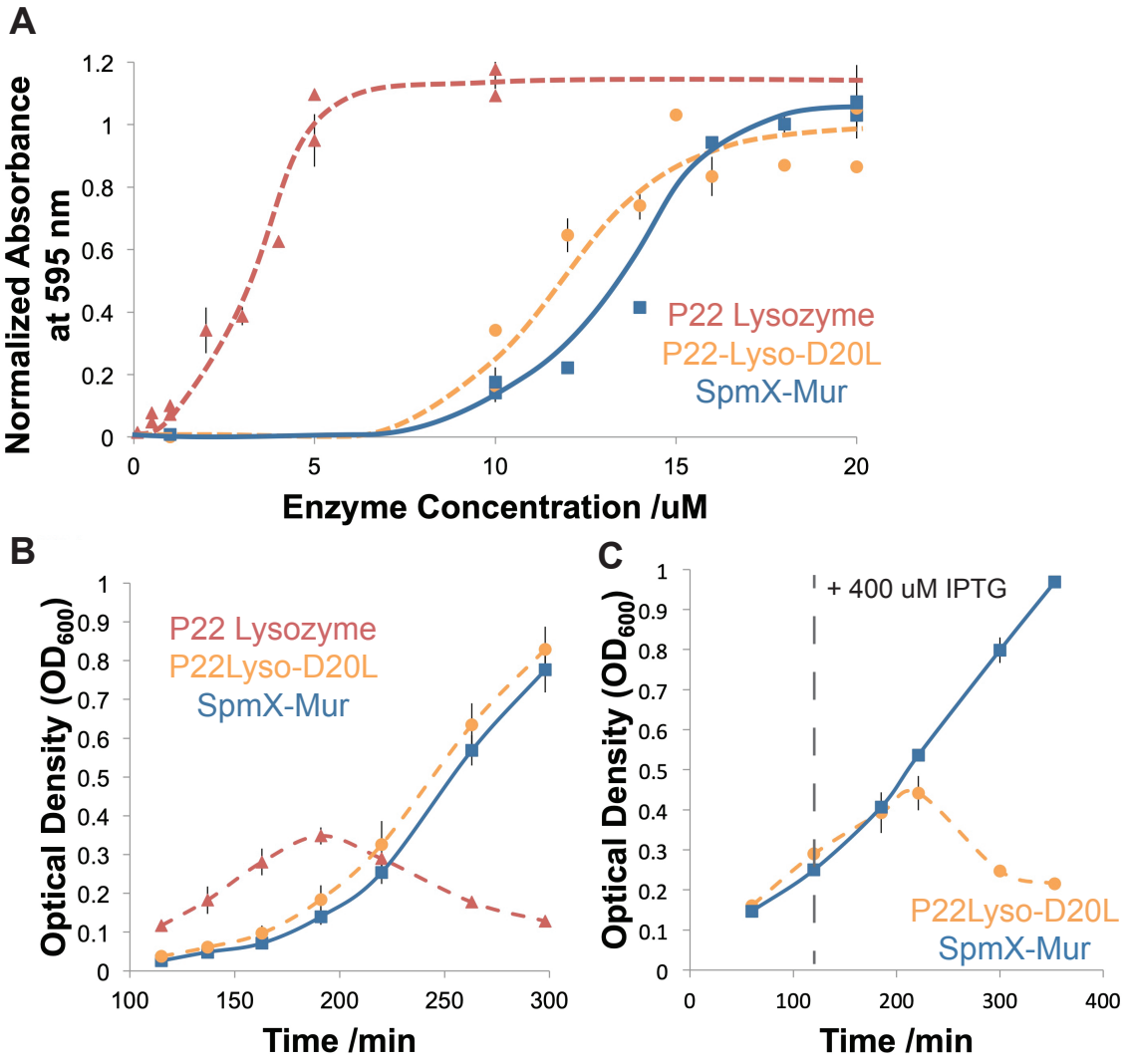
86. Mooers, B.H.M., and Matthews, B.W. (2006). Extension to 2268 atoms of direct methods in the ab initio determination of the unknown structure of bacteriophage P22 lysozyme. Acta Crystallogr. D Biol. Crystallogr. 62, 165–176.

87. Meisner, J., Llopis, P.M., Sham, L.-T., Garner, E., Bernhardt, T.G., and Rudner, D.Z. (2013). FtsEX is required for CwlO peptidoglycan hydrolase activity during cell wall elongation in Bacillus subtilis. Mol. Microbiol. 89, 1069–1083.


A
Muramidase Domain
Intermediate Domain
TM Segments
SpmX Architecture
Conservation
High
Low
B
Genome
SpmX
Escherichia coli
Agrobacterium tumefaciens
Hirschia baltica
Ponticaulis koreensis
Henriciella marina
Hyphomonas adhaerens
Robiginitomaculum antarcticum
Hellea balneolensis
Oceanicaulis alexandrii
Maricaulis maris
Asticcacaulis excentricus
Asticcacaulis biprosthecium
Asticcacaulis benevestitus
Brevundimonas aveniformis
Brevundimonas subvibrioides
Brevundimonas diminuta
Phenylobacterium zucineum
Caulobacter henricii
Caulobacter segnis
Caulobacter crescentus
D20L
D20R
D20G


A
C-term
Lobe
β-Hairpin
GH Motif
Glu
N-term
Lobe
B
i T4L
ii autolysin/
endolysin
iii SpmX-like
muramidases
iv SpmX
4.0
3.0
2.0
1.0
0.0
10
15
20
25
30
WebLogo 3.6.0
Catalytic Glu
β-hairpin / GH Motif

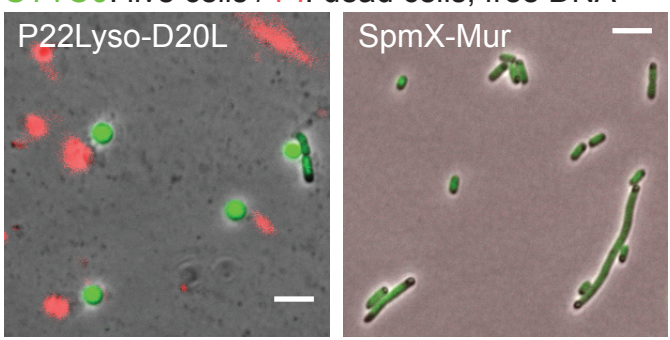

A
C-term Lobe
β-Hairpin
C-term
Lobe
β-Hairpin
GH Motif
Glu
N-term
Lobe
B
Glu
β-Hairpin
GH Motif
C
β-Hairpin
GH Motif
P22Lyso
E11
D20
R14
Y18
β-Hairpin
GH Motif
SpmXMur-Ae
E11
D20
R14
Y18S
D
P22Lyso
E11
D20
T26
E
SpmXMur-Ae
E11
D20L
T26M



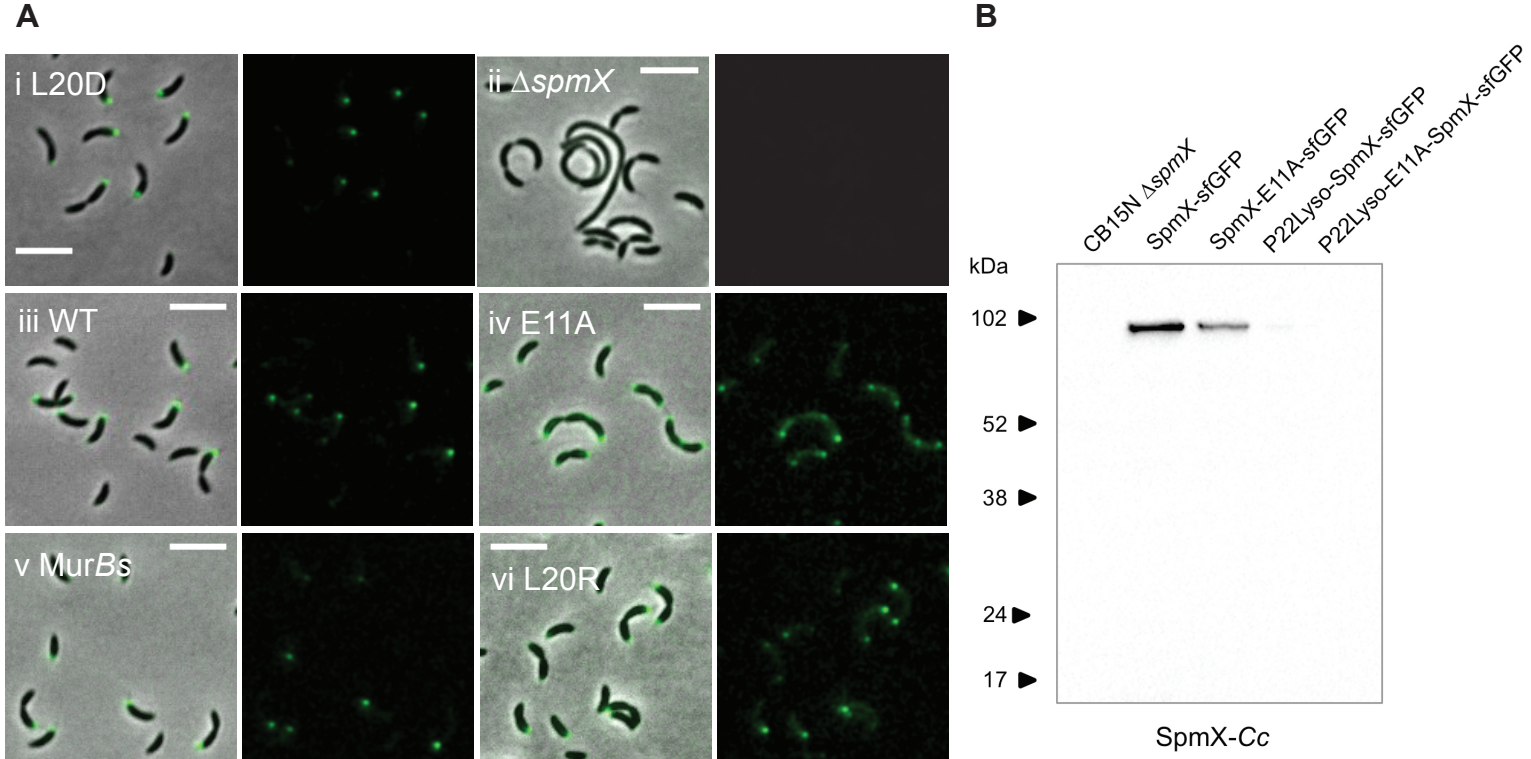
D

Live-Dead Stain:

SYTO9: live cells / PI: dead cells, free DNA



A C. crescentus
B A. biprosthecum
C A. excentricus
WT
ΔspmX
WT
E11A (Catalytic Glu)
N105R (PG-binding)
i
ii
iii
iv
v


A
i L20D
ii ΔspmX
iii WT
iv E11A
v MurBs
vi L20R
B
CB15N ΔspmX
SpmX-sfGFP
SpmX-E11A-sfGFP
P22Lyso-SpmX-sfGFP
P22Lyso-E11A-SpmX-sfGFP
kDa
102
52
38
24
17
SpmX-Cc

Injection

Phage Lytic Cycle

Cell Wall Degradation

Lysis

Lysogeny

Phage Biogenesis

Prophage Inactivation

Caulobacterales

Gene Detoxification and Co-option