Discovery of a widely distributed toxin biosynthetic gene cluster

Shaun W. Lee*†‡, Douglas A. Mitchell*†‡, Andrew L. Markley‡, Mary E. Hensler§¶, David Gonzalez‡, Aaron Wohlrab‡, Pieter C. Dorresteijn*†‡, Victor Nizet§¶, and Jack E. Dixon*†‡**

*Department of Pharmacology, †Department of Cellular and Molecular Medicine, ‡Department of Chemistry and Biochemistry, §Skaggs School of Pharmacy and Pharmaceutical Sciences, and ¶Department of Pediatrics, Division of Pharmacology and Drug Discovery, University of California at San Diego, La Jolla, CA 92093; and Howard Hughes Medical Institute, Chevy Chase, MD 20815


Bacteriocins represent a large family of ribosomally produced peptide antibiotics. Here we describe the discovery of a widely conserved biosynthetic gene cluster for the synthesis of thiazole and oxazole heterocycles on ribosomally produced peptides. These clusters encode a toxin precursor and all necessary proteins for toxin maturation and export. Using the toxin precursor peptide and heterocycle-forming synthetase proteins from the human pathogen Streptococcus pyogenes, we demonstrate the in vitro reconstitution of streptolysin S activity. We provide evidence that the synthetase enzymes, as predicted from our bioinformatics analysis, introduce heterocycles onto precursor peptides, thereby providing molecular insight into the chemical structure of streptolysin S. Furthermore, our studies reveal that the synthetase exhibits relaxed substrate specificity and modifies toxin precursors from both related and distant species. Given our findings, it is likely that the discovery of similar peptidic toxins will rapidly expand to existing and emerging genomes.

antibiotics | bacteriocin | bioinformatics | hemolytic | streptolysin

S.

treptolysin S (SLS), from the human pathogen Streptococcus pyogenes, is a ribosomally synthesized, secreted toxin responsible for the classical β-hemolytic phenotype of bacterial colonies grown on blood agar media (1, 2). S. pyogenes is associated with a wide spectrum of diseases ranging from simple pharyngitis to life-threatening necrotizing fasciitis. The expression of SLS promotes virulence in animal models of invasive infection (1, 3).

The molecular structure of SLS has remained elusive for more than a century; however, SLS is known to be an oxygen-stable, nonimmunogenic cytolysin (4). The gene locus responsible for SLS biosynthesis (streptolysin S-associated genes, sag) was recently identified through transposon and targeted mutagenesis and Fig. S1). Unmodified McbA exhibits no effect on DNA gyrase (11).

Results and Discussion

Heterocycle Formation by SagBCD. Given similarities in organization of the mcg and sag gene clusters and the observed homology of SagBCD to McbBCD, we hypothesized that the SagBCD synthetase complex would serve to posttranslationally modify a toxin precursor in the same manner as McbBCD. The sagBCD genes were cloned individually and purified as fusions to maltose-binding protein (MBP). Because numerous attempts to observe SagA by mass spectrometry failed, we used McbA, which is amenable to mass spectrometry, to detect and confirm heterocycle formation by SagBCD (12). After removal of the MBP tag, recombinant McbA has a calculated molecular mass of 6,293 Da due to addition of Gly-Ser-His to the N terminus. For each heterocycle formed, a loss of 18 Da (oxazoline/thiazoline) or 20 Da (oxazole/thiazole) is expected from the parent peptide (Fig. 1B). Treatment of McbA with recombinant SagBCD resulted in the formation of four new masses differing from the precursor peptide by multiples of 20 Da (Fig. 1 C and D). These masses are within error for linear mode MALDI mass spectrometry and correspond to heterocycle formation at four residues of McbA. The fourth heterocycle was not visible when the synthetase concentration was reduced (data not shown), and only unmodified McbA was observed when SagBCD was omitted from the reaction. These results provide experimental evidence that SagBCD functions in a manner analogous to McbBCD.

SagBCD Converts SagA into a Cytolysin. Because recombinant SagBCD was active in vitro, we next sought to confirm that SagA could be transformed into an active cytolytin in a SagBCD-dependent manner. SagA was produced as an MBP fusion protein and then subjected to modification by the addition of SagBCD in an in vitro synthetase reaction. After this reaction, samples were tested for lytic activity against sheep erythrocytes. In this assay, SLS extracted from S. pyogenes cultures caused rapid hemolysis (data not shown). The addition of SagA alone failed to induce lysis, but robust hemolysis was observed after treatment with SagBCD (Fig. 2A). All three synthetase proteins were required for the lytic phenotype, indicating a cooperative role in toxin maturation.

To determine whether SagBCD-treated SagA exhibited a broader cytolytic phenotype characteristic of the native SLS


The authors declare no conflict of interest.

See Commentary on page 5655.

**To whom correspondence should be addressed at: Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815-6789. E-mail: jedixon@ush.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0801338105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA.
indicated above the respective mass (singly charged species). Linear mode MALDI-TOF mass spectrum of McbA treated with SagBCD. The number of heterocycles on the protoxin peptide is the mass of the peptide. (obsd) molecular masses for MBP-McbA after reaction with SagBCD and thrombinolysis. Formation of a single thiazole/oxazole leads to the loss of 20 Da from a single thiazole/oxazole leads to the loss of 20 Da from the mass of the peptide. (D) Linear mode MALDI-TOF mass spectrum of McbA treated with SagBCD. The number of heterocycles on the protoxin peptide is indicated above the respective mass (singly charged species).

**Fig. 1.** Conservation of toxin biosynthesis operons in *S. pyogenes* and *E. coli*. (A) Genetic organization of the streptolysin S-associated gene cluster (sagA–I) from *S. pyogenes* and the *E. coli* microcin B17 gene cluster (mcbA–G). (B) Through the action of a trimeric synthetase complex, oxazole and thiazole heterocycles are incorporated into a peptidic protoxin scaffold (black) and are active in vitro. Chemical transformations carried out by SagC/McbB (green, cyclodehydratase) and SagB/McbC (yellow, dehydrogenase) orthologs are shown. Molecular mass change for each reaction is shown in daltons. SagD/McbD (blue) serves as an enzymatic scaffold and facilitates substrate binding. (C) Heterocycle formation on *E. coli* McbA by *S. pyogenes* SagBCD. Shown are calculated (calcd) and observed (obsd) molecular masses for MBP-McbA after reaction with SagBCD and thrombinolysis. Formation of a single thiazole/oxazole leads to the loss of 20 Da from the mass of the peptide. (D) Linear mode MALDI-TOF mass spectrum of McbA treated with SagBCD. The number of heterocycles on the protoxin peptide is indicated above the respective mass (singly charged species).

toxin, reaction mixtures were applied to HEK293a cells. Cells treated with samples containing both SagA and all of the synthetase components progressively lost their focal contacts and eventually detached from the tissue culture wells (Fig. 2B). Actin staining revealed massive cytoskeletal collapse consistent with severe membrane damage. HEK cells incubated with solutions missing any component of the synthetase complex, or the protoxin precursor, were indistinguishable from untreated cells (Fig. 2B). Lactate dehydrogenase (LDH) release into the media provided quantitative confirmation that cytotoxicity required both the SagA substrate and the SagBCD complex (Fig. 2C). SagA contains seven cysteines that serve as potential sites for thiazole formation by SagBCD. Mutation of all seven cysteines to serine (SagA-panC/S) completely abolished cytolytic activity. Furthermore, we demonstrate that SagBCD functions in vitro reconstitution of SLS activity. Taken together, these experiments describe the first in vitro reconstitution of SLS activity. Furthermore, we demonstrate that SagBCD functions similarly to McbBCD by installing heterocycles on a peptidic toxin precursor.

The **sagBCD** Gene Clusters Are Widely Distributed Among Prokaryotes. The similarities between the SLS and McbB17 biosynthetic operons were intriguing, given that (i) SLS is produced by a Gram-positive organism whereas McbB17 is from a Gram-negative organism and (ii) microcins have heretofore been classified as unique to Gram-negative bacteria (13). This suggested that other prokaryotes could use related machinery to introduce Ser/Thr/Cys-derived heterocycles into a wide variety of ribosomally produced peptides. We therefore initiated a comprehensive survey of the public genomic databases to identify related biosynthetic gene clusters (14). This search revealed that similar gene clusters are widespread among prokaryotes, most of which are not associated with a function (Fig. 3A). The genes encoding the SagB-like dehydrogenase (Fig. S2), SagC-like cyclodehydratase (Fig. S3), and SagD-like protein (Fig. S4) were present as adjacent ORFs in a diverse group of plant pathogenic organisms spanning six phyla. The SagC and SagD orthologs were also sometimes found fused as a single ORF (Fig. S5).

Of particular interest is the fact that gene clusters that highly resemble *sagA–I* are present in major mammalian pathogens such as *Clostridium botulinum*, *Listeria monocytogenes*, and *Staphylococcus aureus* RF122 (Fig. 3A). Similar gene clusters are also found in distantly related prokaryotes, such as cyanobacteria (15, 16) and archaea (e.g., *Pyrococcus furiosus* DSM 3638). Other family members are found throughout proteobacteria. For instance, a gene cluster identical in arrangement to *E. coli* mcba–G is present in the plant symbiont *Pseudomonas putida* KT2440 (13). *P. putida* colonizes the nutrient-rich rhizosphere of plants and induces plant growth while secreting antibiotics to limit the growth of competing soil bacteria (17, 18). Furthermore, two other plant symbionts, *Bacillus amyloliquefaciens* FZB42 (Ban) and *Bradyrhizobium japonicum* USDA110, encode analogous synthetase proteins. *Bacillus thuringiensis* harbors a similar operon; although not a plant symbiont, *B. thuringiensis* is of agricultural interest owing to its ability to secrete toxins that are effective against a variety of pests, such as moths, butterflies, flies, mosquitoes, and beetles (19). The conservation of this synthetase gene cluster across prokaryotes—in particular, among both pathogenic and symbiotic bacteria—suggests that these gene products play an important role in the establishment of a survival niche for these organisms.

**Presence of Protoxins Across Prokaryotic Lineages.** McbA and SagA encode protoxins of 69 and 53 residues, respectively (Fig. 3B). These are extremely short ORFs, which many gene-identification algorithms do not always recognize. For this reason, manual ORF searches were performed in the intergenic regions for each biosynthetic cluster identified. Short ORFs encoding proteins of 50–70 residues that are rich in Cys/Ser/Thr were present in all organisms containing sagBCD-like genes as adjacent ORFs (Fig. 3B). Although the functions of these potential protoxins are not known, based on similarity to McbA and SagA, we speculate that
in complex natural products (21). Inclusion of these ancillary enzymes is expected to dramatically increase the chemical diversity of the toxins, perhaps for niche purposes. The cyanobacteria *Prochloron didemni* (15) and *Trichodesmium erythraeum* IMS101 (16) have been previously identified as sources of cyclic peptides with Ser/Thr/Cys-derived heterocycles (Fig. S1). The *sag*-like biosynthetic clusters of *T. erythraeum* and *P. didemmi* harbor an additional gene encoding a putative serine protease (Fig. 3A). Although the exact function of this protein is unclear, it is proposed to catalyze a head-to-tail macrocyclization reaction by using the N terminus of a peptide, instead of water, to hydrolyze the acyl-enzyme intermediate. After this reaction, the volume occupied by the toxin is much smaller and may eliminate the need for an elaborate export system. The lack of ABC transporters in these gene clusters supports this notion (Fig. 3A). Based on a similar arrangement of genes, we hypothesize that Rhodobacter sphaeroides, Synechocystis sp. PCC 8106, and Microcystis aeruginosa will also form macrocyclized peptides.

**The Heterocycle Synthetase Complex Displays Functional Promiscuity.** A gene cluster that is highly similar to the *sag* cluster was found in the pathogenic bacterium *C. botulinum* (Fig. 3A). The organization and amino acid sequence of proteins present in the *C. botulinum* cluster *closA–I* displays the highest similarity to *saga–I*. The protoxin, which we have designated *ClosA*, shares many characteristics with *SagA* (39% identical) including the presence of numerous adjacent cysteines, other conserved heterocyclizable residues, and a putative protease site for removal of a leader peptide (Fig. 3B). To explore the possibility that *closA* encodes an SLS-like cytolsin, recombinant MBP-ClosA was prepared and subjected to *in vitro* synthetase assays with *SagBCD*. Hemolysis was observed only in samples containing both *ClosA* and the *SagBCD* synthetase complex (Fig. 4A). Microscopy revealed that cytotoxicity in HEK cells treated with *ClosA* plus *SagBCD* was identical to cells treated with *SagA* plus *SagBCD* (Figs. 2B and 4A). These data demonstrate that *SagBCD* accepts substrates beyond its cognate peptide, *SagA*.

Given the ability of *SagBCD* to accept *McbA* and *ClosA* as alternative substrates, we assessed whether other synthetases were permissive in their substrate selectivity. To this end, the heterocycle-forming synthetases from *E. coli* (McbBCD) and the thermophilic archaeon *P. furiosus* (PagBCD, identity to corresponding *Sag*: 17–24%) were prepared as MBP fusions. *SagA* and *ClosA* were then tested for erythrocyte lytic activity after reaction with McbBCD and PagBCD. The PagBCD complex efficiently converted both *SagA* and *ClosA* into a hemolysin (Fig. 4B). In contrast, the McbBCD complex converted only *ClosA* into a potent lytic species. It is intriguing that McbBCD accepted and converted an alternate substrate into a cytolytic factor given that its biological function is to produce a DNA gyrase inhibitor (22). Therefore, we conclude that the mechanism of toxin action is governed by the protoxin amino acid sequence and that the synthetase proteins are functionally redundant. The extent of substrate tolerance and the kinetics and selectivity of ring formation require further investigation.

In sum, our bioinformatic survey has uncovered the presence of *sag*-like gene clusters in a myriad of prokaryotes, leading us to define a new class of bacteriocin. This class is characterized by a biosynthetic gene cluster that encodes a small protoxin and three adjacent synthetase proteins that endow an organism with the ability to cyclize Ser/Thr/Cys residues from a ribosomally synthesized protoxin scaffold. The finding that the heterocycle synthetase genes are exclusively found as adjacent ORFs will facilitate the identification of orthologous biosynthetic clusters in emerging genomes. Using the protoxin and synthetase proteins from the human pathogen *S. pyogenes*, we demonstrate *in vitro* reconstitution of SLS activity. Furthermore, we show that the synthetase enzymes are functionally redundant and catalyze...
heterocycle formation on alternative substrates despite their significantly distinct evolutionary lineages. Our data suggest that all of the newly identified gene clusters are responsible for the production of bacteriocins containing at least one Ser/Thr/Cys-derived heterocycle. Many of the gene clusters encode ancillary modifying enzymes that will install further modifications onto the bacteriocin, thus increasing the chemical diversity of this family. Further insights into this bacteriocin family will lead to the identification of novel targets for antibiotic and vaccine development.

Fig. 3. The biosynthetic operon for producing thiazole/oxazole-containing toxins is widely distributed. (A) Gene clusters from organisms containing SLS- and MccB17-like bacteriocins. Members are sorted by prokaryotic phylum. Relative gene length and directionality are shown (scale for actinobacteria and cyanobacteria is reduced by 50%). Each gene cluster contains a protoxin (A, black), dehydrogenase (B, yellow), cyclodehydratase (C, green), and a docking scaffold (D, blue). These genes produce both single-domain and fusion proteins. Numerous ancillary enzymes are included and increase the chemical diversity of the toxin family. (B) Select protoxin amino acid sequences. Predicted leader peptide cleavage sites are denoted with an asterisk. Hyphens indicate a known cleavage site. Potential sites of heterocyclization are indicated in blue, and known sites of heterocycle formation are indicated in red. Green text signifies conversion to dehydroalanine. Toxins similar to SagA (top) are predicted cytolsins.
base (National Center for Biotechnology Information) were initiated using all proteins confirmed to have a sag-like gene locus (with bcd genes located directly adjacent to or within a few ORFs of each other). A lower limit of 30% amino acid similarity was chosen as a threshold to assign a candidate protein as being homologous (BLOSUM30 matrix). This second search identified gene clusters from three other organisms that produce a heterocycle-containing bacteriocin with a known structure (Fig. 51), Streptomyces sp. TP-A0584 (goadsporin) (26), P. didemni (patellamide A/C) (15), and T. erythraeum IMS101 (trichamide) (16) and from additional organisms.

**Mass Spectrometry. Sample preparation.** Recombinant MBP-McbA (110 μl at 40 μM) was proteolytically digested with 50 NIH units of thrombin (Sigma–Aldrich) for 4 h at room temperature in thrombin cleavage buffer: 50 mM Tris (pH 7.8) and 10 mM CaCl₂. During the course of the reaction, cut McbA precipitates (pl ~ 7) as a white solid. Precipitated McbA was subsequently harvested by centrifugation and washed two to three times with 100 μl of deionized water to remove the majority of buffer, salts, detergent, cut MBP, and uncut MBP-McbA. This precipitate was then resolubilized in 60% MeCN/2% formic acid (40 μl). Samples of MBP-McbA that were treated with BCD synthetase complex (Sag and Pag) were first reacted, as described in In Vitro Synthetase Assay, before thrombinolysis.

The MALDI target (stainless steel, Applied Biosystems) was prepared by allowing a saturated solution of sinapic acid (Sigma–Aldrich) in 50% MeCN/0.1% TFA to fully dry on the target (1.5 μl per sample, in duplicate). The precipitated McbA sample was then diluted (1:1) in the saturated matrix solution before spotting on top of the dried spot (1 μl in duplicate).

**Instrument settings.** MALDI-TOF mass spectra were acquired on a Voyager DE-STR instrument (Applied Biosystems) in linear positive mode. Specific settings were typically as follows: 2000–2100 laser power, 93% grid, 0.12% guide wire, 400 nsec delay, mass window 2000–18,000 Da.

In **Vitro Synthetase Assay.** Synthetase reactions using MBP-tagged SagA and SagBCD and other fusion proteins were performed in a manner described earlier (27, 28). Reaction mixtures consisted of MBP-SagA (10 μM) and MBP-SagBCD (2 μM) in each in a total volume of 100 μl of synthetase buffer (50 mM Tris–HCl, pH 7.5/125 mM NaCl/20 mM MgCl₂/2 mM ATP/10 mM DTT). Reactions were allowed to proceed at 37°C for 16 h unless otherwise stated. These reactions were used for hemolytic assays and LDT release assays to quantify membrane damage. Omission of either ATP or DTT resulted in loss of hemolytic activity (data not shown).

**Hemolytic Assay.** Fresh defibrinated whole sheep blood was obtained from Hemostat Laboratories. Whole blood was washed three times in PBS and diluted to a final concentration of 1:25 vol/vol in PBS. Prepared whole blood (100 μl) was then placed into individual wells of a flat-bottom 96-well microtiter plate (Costar, Corning). Reactions from in vitro synthetase assays were then added directly to the wells, and the mixture was reacted for 16 h at 37°C unless otherwise noted. After incubation, plates were centrifuged at 300 × g for 10 min and an aliquot of supernatant (100 μl) was placed in a separate microtiter plate for measuring hemoglobin absorbance at 450 nm on a Victor3 microplate reader (Perkin–Elmer).

**ACKNOWLEDGMENTS.** We thank Joyce Limm for technical assistance and Christopher Walsh (Harvard Medical School, Boston) for the mbcA–D clones. We also thank Elizabeth Komives and Michael VanNieuwenhze for their helpful suggestions. Seema Mattoo critically reviewed the manuscript, for which we are grateful. This research was supported by grants from the National Institutes of Health, the Walter Cancer Institute, the Ellison Foundation, and the Howard Hughes Medical Institute (to J.E.D.). D.A.M. is supported by a Hartwell Foundation postdoctoral fellowship. S.W.L. is supported by a University of California at San Diego Heme and Blood Proteins postdoctoral training grant.


Supporting Information

Lee et al. 10.1073/pnas.0801338105

Supporting Text

For the sake of simplicity, all gene identifiers for this paper were used to indicate the following designations unless otherwise stated: A, protoxin; B, dehydrogenase; C, cyclodehydratase; D, docking protein. This family of bacteriocin is defined as a ribosomally produced toxin with conserved machinery (the BCD enzymes) to introduce thiazole, thiazoline, (methyl)oxazole, and (methyl)oxazoline heterocycles, which are derived from cysteine, threonine, and serine residues, into the peptidic backbone provided by a structural protein.

Materials and Methods

Phylogenetic trees were constructed using the ClustalW server located at http://align.genome.jp. Slow/accurate pairwise alignments were used. All other parameters were the default options. The trees were visualized by selecting “unrooted N-J tree.”

Cloning. The genes encoding SagA (locus tag, SPy_0738), SagB (Spy_0739), SagC (Spy_0740), and SagD (Spy_0741) were previously amplified from Streptococcus pyogenes genomic DNA and cloned into the group A streptococcal expression vector, pDerm (1). For recombinant expression of these proteins in E. coli, the SagA-D inserts were amplified by PCR using the following primers (IDT), which contain 5′ BamHI and 3′ NotI restriction endonuclease sites (left to right, 5′ to 3′): SagA_fwd, GAG GATCC ATG TTA AAA TTT ACT TCA AAT ATT TTA G; SagA Rev, ACA GCG GCC GGC TAT TTA CCT GGC GTA TAA CCT CC; SagB_fwd, CGG ATC CAT GTC ATT TTT TAC AAA GG; SagB_rev, AAA AGC GGC CGC CTA TGG AGA CTC TCT AGT TCC; SagC_fwd, AAA AAG ATC CAT GAA ATA TCA ACT TAA TAG TAA TG; SagC_rev, AAA AGC GGC CGC TCG ACT ATT CGT CAA GGA G; SagD_fwd, AAA AGG ATC CAT GTT ATA CTA TTA TCC TTC TTT TAC CC; SagD_rev, AAA AGC GCC CGC GAA TTA GGC ATT GG.

The amplified inserts were purified by using a 1% agarose gel, excised, and gel-extracted (Invitrogen). Purified inserts and a modified pET28 vector (EMD Chemicals) containing an N-terminal maltose-binding protein (MBP) fusion with thiombolin and TEV protease sites were double digested (BamHI/NotI; NEB) as per the manufacturer’s instructions. Digested vector was treated with calf intestine alkaline phosphatase (CIP; Promega) for 30 min at 37°C before purifying digested inserts and vector on a 1% agarose gel. Digested plasmid and insert were gel-extracted and ligated by using T4 DNA ligase (NEB) before transformation into chemically competent DH5α. Plasmid DNA was isolated by miniprep kit (Invitrogen) and sequenced at Eton Bioscience using the T7 forward, T7 reverse, and MBP forward primers. Plasmid DNA was then subcloned into the pET15 vector (3, 4).

Expression and Purification of Recombinant Proteins. Constructs containing confirmed sequence were transformed into BL21(DE3)RIL cells under selection with 50 μg/ml kanamycin (or 100 μg/ml ampicillin) and 35 μg/ml chloramphenicol. A starter culture (10 ml) was grown overnight and used to inoculate 2× 1 liter of LB medium for each protein. These cells were grown at 37°C to an OD600 of 0.7. Cultures expressing the protoxins were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3–4 h at 37°C. Cultures expressing the BCD synthetase enzymes were lowered to 22°C before inducing expression with 0.4 mM IPTG for 16 h. The cells were harvested by centrifugation, and the pellets were stored at −80°C until purification.

Cell pellets were resuspended in MBP lysis buffer [50 mM Tris, pH 7.5/0.5 M NaCl/2.5% (vol/vol) glycerol/0.1% (vol/vol) Triton X-100] along with 5 mg/ml lysozyme and EDTA-free Complete Protease Inhibitor Mixture Tablets (Roche) for 45 min at 4°C. The resuspended cells were lysed by three rounds of sonication at 4°C. Centrifugation for 40 min at 40,000 × g yielded supernatant that was immediately gravity-loaded onto a column packed with 8 ml of amylose resin (NEB) and prequilibrated with MBP lysis buffer. The loaded column was washed with 10–15 column volumes of ice-cold wash buffer A: 50 mM Tris (pH 7.5), 0.4 M NaCl, 0.5 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), 2% glycerol, and 0.1% Triton X-100. A second wash step included 1 column volume of ice-cold wash buffer B (lacks Triton X-100) before elution with 5 column volumes of 50 mM Tris (pH 7.5), 0.15 M NaCl, 0.5 mM TCEP, 10 mM maltose, and 2.5% glycerol. Elution fractions containing a band of the correct molecular weight, as determined by Coomassie-stained SDS/PAGE, were pooled, buffer-exchanged, and concentrated by using 50-kDa molecular mass cutoff concentrators (Millipore). The final storage buffer was 50 mM 4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid (Hepes, pH 7.5), 50 mM NaCl, 0.25 mM TCEP, and 2.5% glycerol. All proteins containing the N-terminal MBP fusion tag over-expressed to sufficient levels to yield >95% pure protein after single column purification. The fractions containing homogeneous proteins were then aliquoted and stored at −80°C until needed. Concentrations were assessed by using the Bradford method (BSA standard) and denatured absorbance at 280 nm (M guanidine HCl). Typical yields: protoxin, 50 mg/liter; dehydrogenase 30 mg/liter; cyclodehydratase, 10 mg/liter; docking protein, 15 mg/liter.

Cell Culture and Treatments. Human embryonic kidney cells (HEK293) were maintained in a 5% CO2, water-saturated atmosphere at 37°C in DMEM with glutamine (Gibco) supple-
mented with 5% FBS, penicillin (100 units/ml), and streptomycin (100 μg/ml).

**Cytotoxicity Assay.** Lactate dehydrogenase (LDH) levels were determined by using a colorimetric assay (Roche). HEK293a cells were grown to 70–80% confluence in 24-well microtiter plates (Falcon, BD Biosciences) as described above. Cells were washed three times in assay medium consisting of phenol-red free DMEM (Mediatech), and treated with in vitro synthetase reactions for 10 h unless otherwise stated. Cells were treated with Triton X-100 (Sigma–Aldrich) (1%, vol/vol, in medium) for 15 min as a positive control. Culture supernatants were prepared as per the manufacturer’s instructions, and the absorbance at 490 nm was recorded (n = 3). Background was subtracted from control reactions; all readings were averaged and plotted as the mean ± standard deviation.

**Cell Microscopy.** HEK293a cells were grown to 50% confluence on BIOCOAT fibronectin-coated coverslips (BD Biosciences) and treated with synthetase reactions for 4–6 h. The coverslips were washed three times with PBS, fixed in paraformaldehyde (4% in PBS, wt/vol) for 20 min at 22°C, and permeabilized with isotonic PBS containing normal goat serum (3% vol/vol; Gibco, Invitrogen) and Triton X-100 (0.2% vol/vol) for 1 h. The samples were subsequently treated with rhodamine phalloidin (1:1,000 in PBS; Molecular Probes) for 10 min to visualize actin filaments and anti-vinculin antibody (1:3,000) for cytoplasmic staining. Samples were washed extensively with PBS before mounting in Gelvatol. Images were acquired by using a ×40 objective on a Zeiss Axiovert System. The images were subsequently exported to Adobe Photoshop 7.0 for image preparation.

Fig. S1. Molecular structures of compounds synthesized by a sag-like gene cluster. Posttranslational modifications are indicated. Dehydroalanines and thiazoles are derived from cysteine, oxazole from serine, and methyloxazole from threonine. Patellamide A contains two reduced heterocycles: a methyloxazoline and oxazoline. These are derived from threonine and serine, respectively, and are not labeled for clarity.
Fig. S2. Multiple sequence alignment and phylogenetic analysis of the dehydrogenase. (A) The chemical transformation carried out by SagB orthologs. Oxazoline (X = O) and thiazoline (X = S) heterocycles are oxidized by two electrons to oxazole and thiazole, respectively. During this oxidative aromatization reaction, oxidized flavin mononucleotide (FMN) is reduced by two electrons to FMNH$_2$. Molecular oxygen restores the flavin to the fully oxidized state in vitro; in vivo, the oxidant remains unknown. (B) ClustalW alignment of ~30 SagB orthologs (includes S. dysgalactiae). Because of divergence at the N and C termini of the proteins, only ~200 of the more highly conserved residues are shown in this alignment for each protein. This region contains several positions that contain nearly invariant tyrosines (yellow boxes). FMN binding sites are diverse, but usually the cofactor is π-stacked between Y or W. There are also invariant arginine residues, which may play a role in coordinating the phosphate group of FMN. Note that the S. aureus protein is annotated as multiple ORFs (SAB1378c, SAB1377c, and SAB1376c). (C) Phylogenetic analysis of ClustalW aligned SagB homologs reveals evolutionary clustering. The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; blue, cyanobacteria; brown, actinobacteria; black, other.
Fig. S3. Multiple sequence alignment and phylogenetic analysis of the cyclodehydratase. (A) The chemical transformation carried out by SagC orthologs. The cyclodehydratase catalyzes the conversion of serine to oxazoline (X – O), cysteine to thiazoline (X – S), and threonine to methyloxazoline with loss of water from the parent peptide. (B) ClustalW alignment of 16 SagC orthologs (the others are encoded as a fusion protein to SagD). Because of divergence at the N and C termini of the proteins, only 160 of the more highly conserved residues are shown in this alignment for each protein. This region contains several positions that contain two nearly invariant CXXC motifs (green boxes) that likely serve to tetrahedrally coordinate a structural Zn$^{2+}$ . The Streptococcal orthologs are the exception—they contain two nearby DXXE sequences (D, aspartic acid; E, glutamic acid), a known metal-binding motif. One of the DXXE sequences can be seen on the given alignment. (C) Phylogenetic analysis of ClustalW aligned SagC proteins reveals evolutionary clustering. The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; black, other. The cyanobacteria, actinobacteria, and bacteriodetes phyla have a tendency to fuse SagCD and are shown in a separate figure.
Fig. S4. Multiple sequence alignment and phylogenetic analysis for the docking protein. (A) ClustalW alignment of 16 SagD orthologs (the others are encoded as a fusion protein to SagC). Because of divergence at the N terminus, the ~400 C-terminal residues are shown in this alignment. This region contains several invariant positions and a proline (P)-rich C terminus. (B) Phylogenetic analysis of ClustalW aligned SagD proteins reveals evolutionary clustering and a conserved proline-rich C terminus (blue boxes). The labels are sorted by phylum and are colored as follows: red, firmicutes; green, proteobacteria; black, other. The cyanobacteria, actinobacteria, and bacteriodetes phyla have a tendency to fuse SagCD and are shown in a separate figure.
Fig. S5. Multiple sequence alignment and phylogenetic analysis for the fused cyclodehydratase/docking protein. (A) ClustalW alignment of 14 fused SagCD orthologs (the Salinospora clusters contain two copies of this protein). Because of divergence at the N terminus, the ~700 C-terminal residues are shown in this alignment. The N-terminal part of this region contains invariant CXXC motifs that comprise a zinc-tetrathiolate (cyclodehydratase domain, green boxes). The C-terminal part of this alignment reveals a proline-rich region (docking domain, blue boxes). (B) Phylogenetic analysis of ClustalW aligned SagD proteins reveals evolutionary clustering. For S. arenicola and S. tropica, 1 designates the SagCD ortholog at the beginning of the operon and 2 represents the ortholog at the end of the operon. The labels are sorted by phylum and are colored as follows: brown, actinobacteria; green, proteobacteria; blue, cyanobacteria; red, firmicutes; black, bacteriodetes.