

Supporting Information

Lee et al. 10.1073/pnas.0801338105

Supporting Text

For the sake of simplicity, all gene identifiers for this paper use 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 protoxin 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, pD-Cerm (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 GGATCC ATG TTA AAA TTT ACT TCA AAT ATT TTA G; SagA_rev, ACA GCG GCC GCT TAT TTA CCT GGC GTA TAA CTT CC; SagB_fwd, CGG ATC CAT GTC ATT TTT TAC AAA GG; SagB_rev, AAA AGC GGC CGC CTA TTG AGA CTC CTT AGT TCC; SagC_fwd, AAA AGG ATC CAT GAA ATA TCA ACT TAA TAG TAA TG; SagC_rev, AAA AGC GGC CGC TCG ATT ACT CGT CAA GGA G; SagD_fwd, AAA AGG ATC CAT GTT ATA CTA TTA TCC TTC TTT TAA TC; SagD_rev, AAA AGC GGC CGC GAA TTC TAA 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 thrombin 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 DH5 α . Plasmid DNA was isolated by miniprep kit (Invitrogen) and sequenced at Eton Bioscience using the T7 forward, T7 reverse, and MBP forward (5'-ATG AAG CCC TGA AAG ACG-3') primers to verify the presence of the gene and to ensure no mutations were introduced during the cloning.

The gene encoding ClosA, locus tag CLL0566 (2), from *Clostridium botulinum* F str. Langeland was chemically synthesized (Genscript). The construct was provided as a GST fusion in the pGS21a vector. ClosA was then subcloned into the pET28-MBP vector and sequenced, as described above, using the following primers: ClosA_fwd, AAG GAT CCA TGC TGA AAT TTA ACG AAC ATG TGC TGA CC; ClosA_rev, AAG CGG CCG CTT AGT TGC CGC CCT GAC CCG C.

Genes encoding the synthetases from *Pyrococcus furiosus* (PagB locus tag is PF0001, PagC is PF0003, and PagD is PF0002) amplified from genomic DNA (American Type Culture Collec-

tion) and inserted into an MBP-modified pET15 vector (3, 4). This vector is similar to pET28-MBP but lacks the hexahistidine tags and TEV protease site. The primers used were as follows: PagB_fwd, AAG GAT CCA TGG TTC ATT ATT TTG TGA AGG TAA C; PagB_rev, AAG CGG CCG CTC ACA CTA ATT TAC CCA TTC CTC C; PagC_fwd, AAA AGC TTG GAT CCA TGG CAA ATT CAA AAA GGA AGA ATG CAC G; PagC_rev, AAG CGG CCG CTC ATT TAA ACC ACC TCC CCT CAC AGA CAT TAC; PagD_fwd, AAG GAT CCA TGA GGA TAA ATT GTC TTA GCA ATA TTG AAG; PagD_rev, AAG CGG CCG CTT ATG GAT ATG GAT GAG GAA TAC ACC TTA GC.

Expression and Purification of Recombinant Proteins. Constructs containing confirmed sequence were transformed into BL21(DE3)RIL cells under selection with 50 μ g/ml kanamycin (or 100 μ g/ml ampicillin) and 35 μ g/ml chloramphenicol. A starter culture (10 ml) was grown overnight and used to inoculate 2 \times 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 \times g yielded supernatant that was immediately gravity-loaded onto a column packed with 8 ml of amylose resin (NEB) and pre-equilibrated 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), 2.5% 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 (6 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-

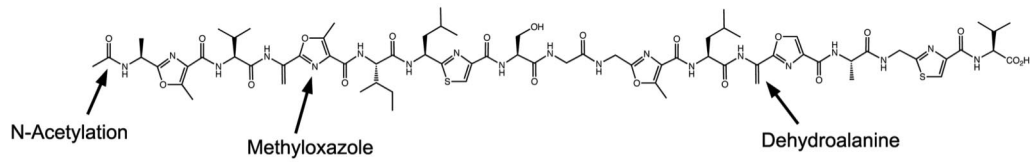
mented with 5% FBS, penicillin (100 units/ml), and streptomycin (100 μ g/ml).

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 \pm standard deviation.

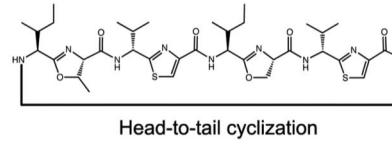
Cell Microscopy. HEK293a cells were grown to 50% confluence on BIOCOAT fibronectin-coated coverslips (BD Biosciences) and treated with synthetase reactions for 4–6 h. The coverslips were washed three times with PBS, fixed in paraformaldehyde (4% in PBS, wt/vol) for 20 min at 22°C, and permeabilized with 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 $\times 40$ objective on a Zeiss Axiovert System. The images were subsequently exported to Adobe Photoshop 7.0 for image preparation.

1. Datta V, et al. (2005) Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. *Mol Microbiol* 56:681–695.
2. Sebahia M, et al. (2007) Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res* 17:1082–1092.
3. Maeder DL, et al. (1999) Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* 152:1299–1305.
4. Robb FT, et al. (2001) Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology. *Methods Enzymol* 330:134–157.

Goadsporin (*S. sp. TP-A0584*)



Patellamide A (*P. didemni*)



Microcin B17 (*E. coli*)

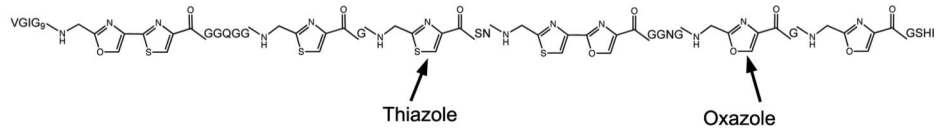


Fig. S1. Molecular structures of compounds synthesized by a *sag*-like gene cluster. Posttranslational modifications are indicated. Dehydroalanines and thiazoles are derived from cysteine, oxazole from serine, and methyloxazole from threonine. Patellamide A contains two reduced heterocycles: a methyloxazoline and oxazoline. These are derived from threonine and serine, respectively, and are not labeled for clarity.

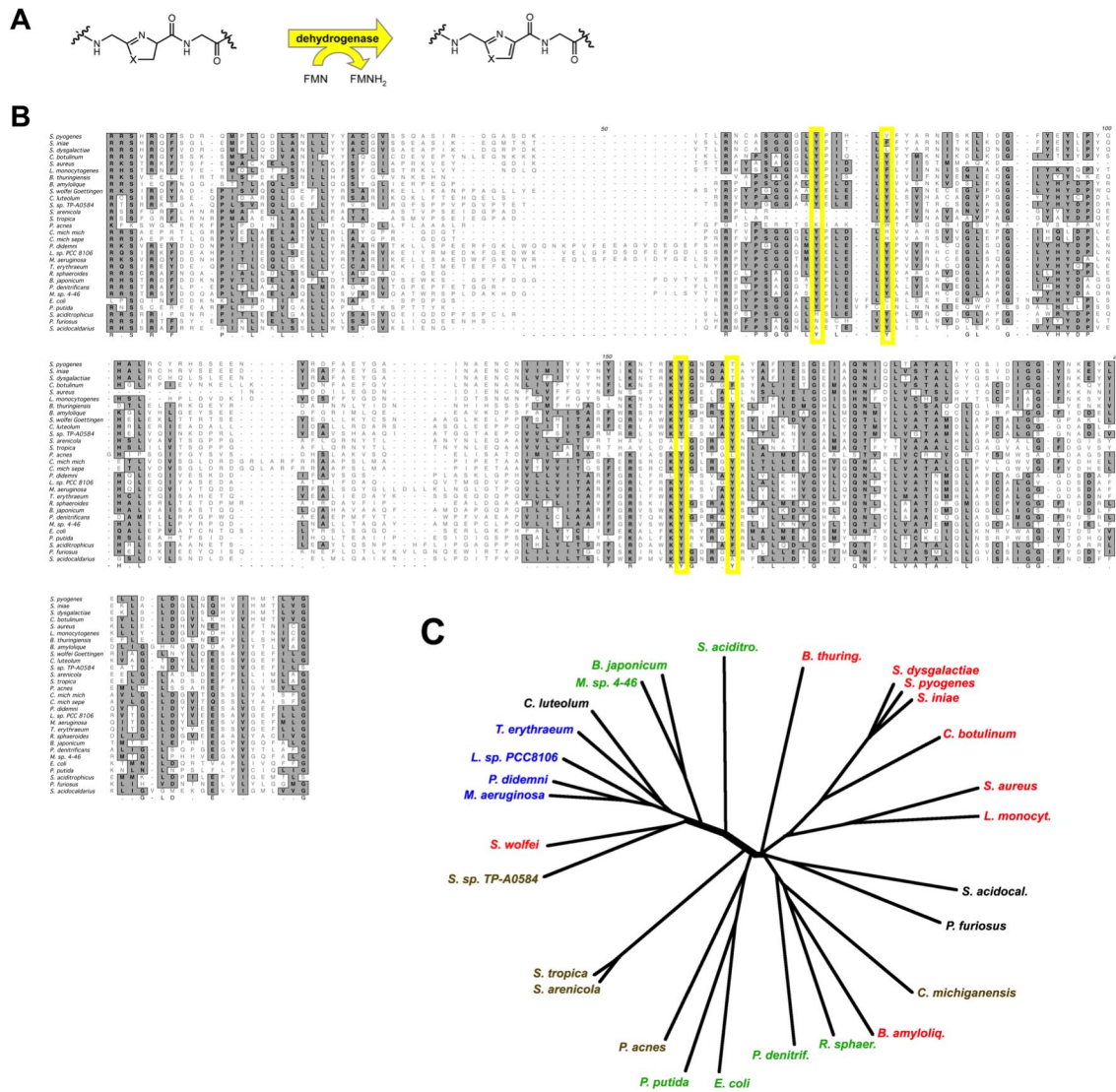


Fig. S2. Multiple sequence alignment and phylogenetic analysis of the dehydrogenase. (A) The chemical transformation carried out by SagB orthologs. Oxazoline (X = O) and thiazoline (X = S) heterocycles are oxidized by two electrons to oxazole and thiazole, respectively. During this oxidative aromatization reaction, oxidized flavin mononucleotide (FMN) is reduced by two electrons to FMNH₂. 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.

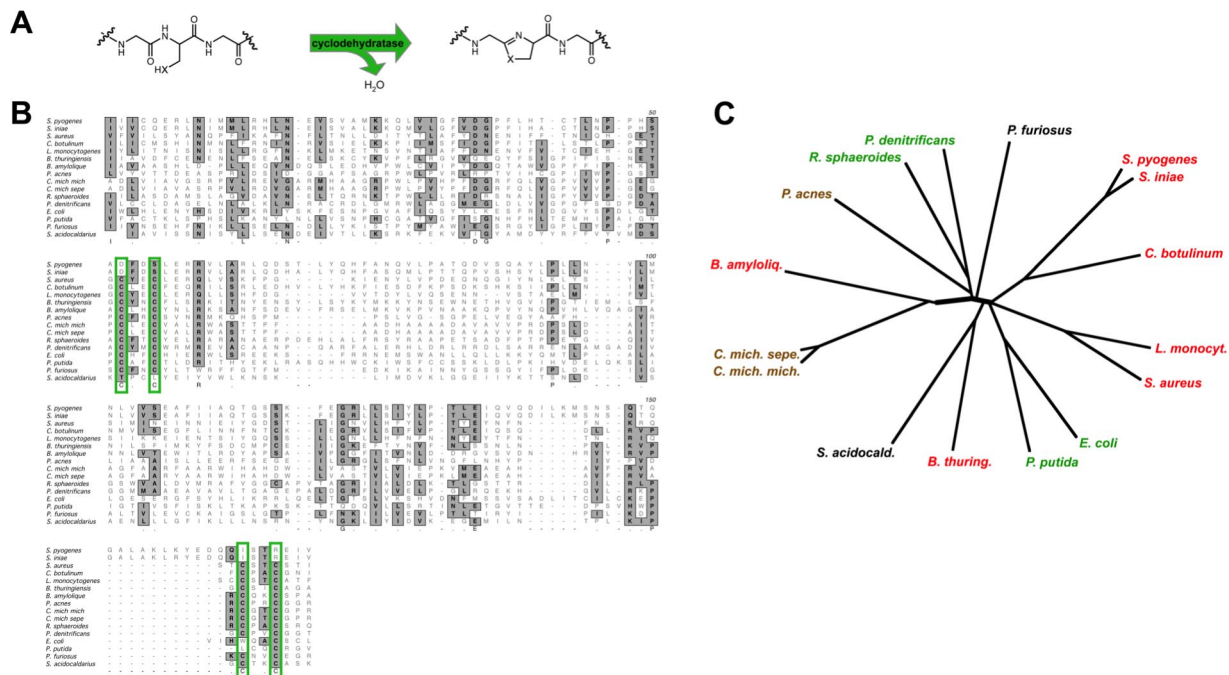


Fig. S3. Multiple sequence alignment and phylogenetic analysis of the cyclodehydratase. (A) The chemical transformation carried out by SagC orthologs. The cyclodehydratase catalyzes the conversion of serine to oxazoline ($X = O$), cysteine to thiazoline ($X = S$), and threonine to methyloxazoline with 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 bacterioidetes phyla have a tendency to fuse SagCD and are shown in a separate figure.

A

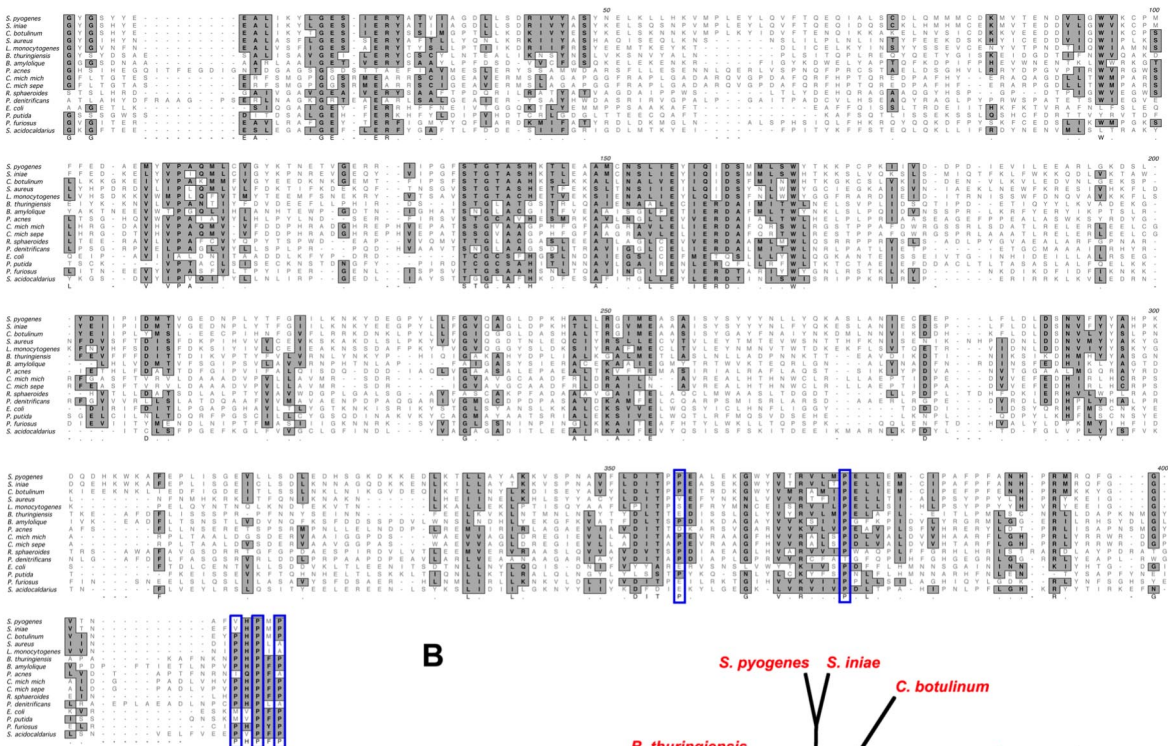
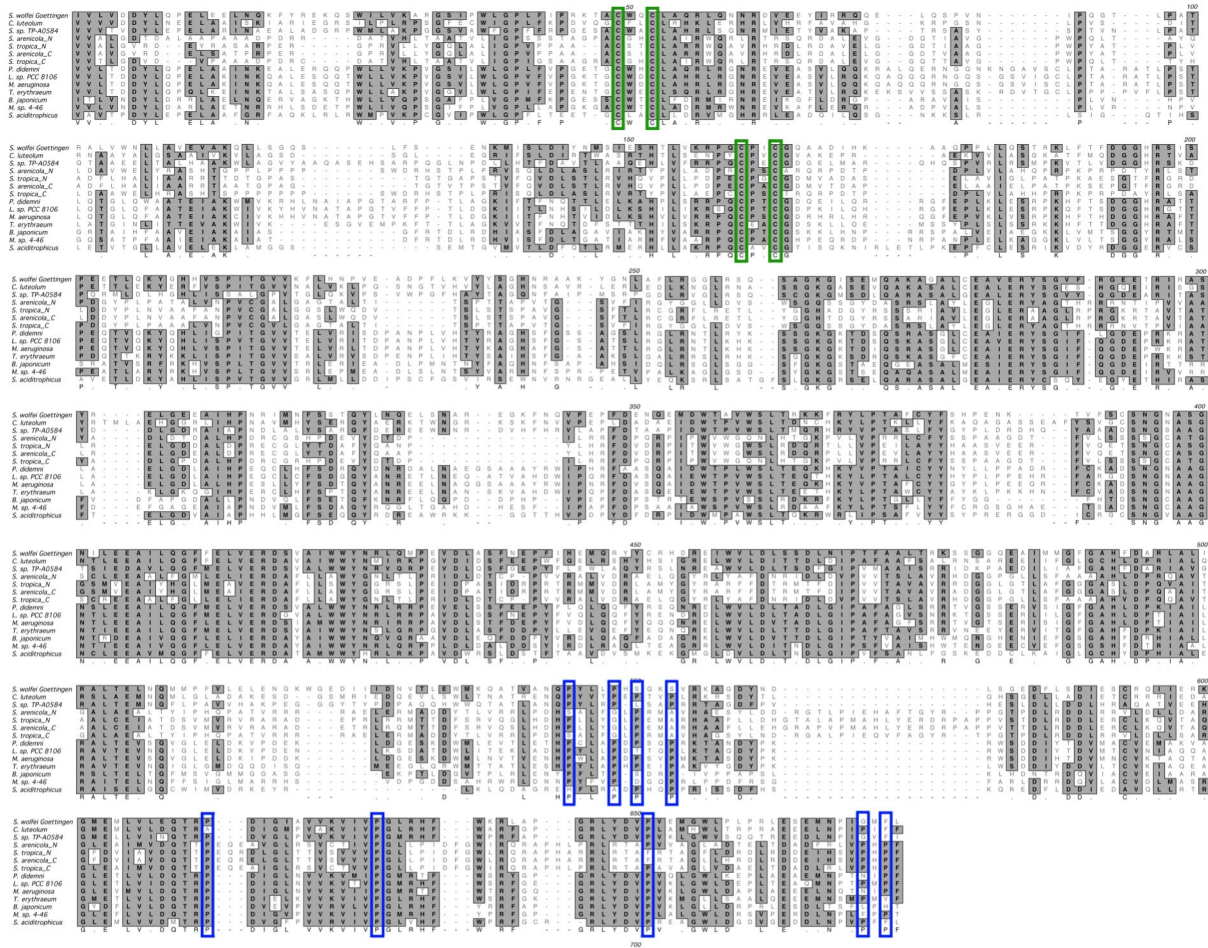


Fig. S4. Multiple sequence alignment and phylogenetic analysis for the docking protein. (A) ClustalW alignment of 16 SagD orthologs (the others are encoded as a fusion protein to SagC). Because of divergence at the N terminus, the ≈400 C-terminal residues are shown in this alignment. This region contains 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 bacteroidetes phyla have a tendency to fuse SagCD and are shown in a separate figure.

A



B

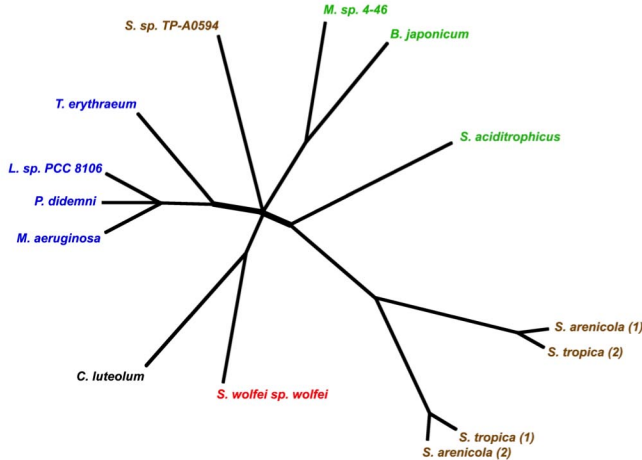


Fig. S5. Multiple sequence alignment and phylogenetic analysis for the fused cyclodehydratase/docking protein. (A) ClustalW alignment of 14 fused SagCD orthologs (the *Salinospora* clusters contain two copies of this protein). Because of divergence at the N terminus, the \approx 700 C-terminal residues 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, bacteroidetes.