HHS Public Access

Author manuscript

Cell Chem Biol. Author manuscript; available in PMC 2020 May 16.

Published in final edited form as:

Cell Chem Biol. 2019 May 16; 26(5): 724–736.e7. doi:10.1016/j.chembiol.2019.02.004.

Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry

Katherine D. Bauman[1], **Jie Li**[1], **Kazuya Murata**[1,2], **Simone M. Mantovani**[1,3], **Samira Dahesh**[4], **Victor Nizet**[4,5], **Hanna Luhavaya**[1,*], and **Bradley S. Moore**[1,5,*,±]

[1]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA, USA

[2]Current address: Faculty of Pharmacy, Kindai University, Osaka, Japan

[3]Current address: Amyris Pharmaceuticals, Emeryville, CA, USA

[4]Department of Pediatrics, University of California at San Diego, La Jolla, CA, USA

[5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA, USA.

Summary

The disconnect between the genomic prediction of secondary metabolite biosynthetic potential and the observed laboratory production profile of microorganisms is well-documented. While heterologous expression of biosynthetic gene clusters (BGCs) is often seen as a potential solution to bridge this gap, it is not immune to many challenges including impaired regulation, the inability to recruit essential building blocks, and transcriptional and/or translational silence of the biosynthetic genes. Here we report the discovery, cloning, refactoring, and heterologous expression of a cryptic hybrid phenazine-type BGC (*spz*) from the marine actinomycete *Streptomyces* sp. CNB-091. Overexpression of the engineered *spz* pathway resulted in increased production and chemical diversity of phenazine natural products belonging to the streptophenazine family, including bioactive members containing an unprecedented *N*-formylglycine attachment. An atypical discrete adenylation enzyme in the *spz* cluster is required to introduce the formylglycine moiety and represents a phylogenetically distinct class of adenylation proteins.

Bauman *et al.*, designed and utilized promoter cassettes to transcriptionally activate the cryptic streptophenazine biosynthetic pathway from marine *Streptomyces* bacteria. Pathway refactoring

[*]**co-corresponding author** Correspondence: Dr H. Luhavaya: hluhavaya@ucsd.edu, Prof B. S. Moore: bsmoore@ucsd.edu.
[±]lead contact

Author contributions

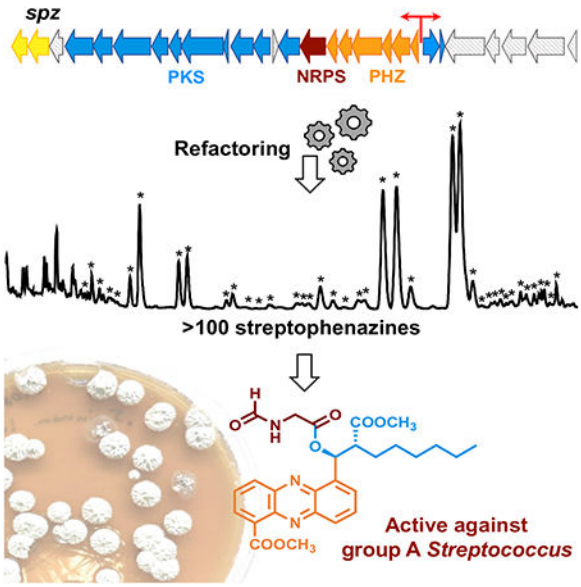
H.L. and B.S.M. designed the research. K.D.B., J.L., K.M., and S.M.M. performed the experiments and analyzed the data. V.N. and S.D. designed and performed the bioactivity testing. K.D.B., H.L., and B.S.M. wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

resulted in production of over 100 compounds, including streptophenazines with a rare *N*-formylglycine moiety that showed enhanced antibiotic activity.

Graphical Abstract



Keywords

streptophenazine; refactoring; promoter; cryptic biosynthetic gene cluster; adenylation enzyme; heterologous expression; natural product; antibiotic; marine *Streptomyces*

Introduction

Small molecules provide a fundamental means by which microorganisms interact with each other and their environment, facilitating diverse yet critical functions that can include defense, cell signaling, and/or nutrient acquisition (Demain and Fang, 2000). Nature has engineered a chemically diverse array of molecules designed to interact with a variety of biological targets and thereby mediate microbial communication. The potent biological activities that define secondary metabolites have been harnessed by humankind as well, most famously as sources of and inspiration for pharmaceuticals (Koehn and Carter, 2005; Newman and Cragg, 2016). The enzymes necessary for production of these metabolites are genetically encoded, and in bacteria, are grouped together on contiguous stretches of DNA, forming biosynthetic gene clusters (BGCs) (Osbourn, 2010). The plummeting cost of high-throughput microbial DNA sequencing and resulting flood of genomic information has revealed that up to 90% of the chemical potential of microorganisms is unexplored as many BGCs are silent, particularly under laboratory conditions, and thus their molecular products have not been identified (Bentley *et al.*, 2002; Omura *et al.*, 2001).

Controlled heterologous expression is a powerful tool to unlock the biosynthetic potential encoded by silent and cryptic BGCs. This process can be utilized to circumvent issues with

culturing and difficulties with genetic manipulation of the native producers, allowing access to the molecules produced by unculturable and symbiotic organisms (Gomez-Escribano and Bibb, 2014). However, the heterologous production of secondary metabolites can have low efficiency, and certain BGCs may remain silent even within validated heterologous expression systems (Rutledge and Challis, 2015). Additional genetic manipulations are often necessary to improve or activate production within heterologous hosts. These may include manipulating the pathway-specific regulatory genes (Olano *et al.*, 2008), or using synthetic biology tools to rewrite native genetic elements with optimized components (e.g., promoters, ribosome binding sites (RBSs), *etc.*) (Freestone *et al.*, 2017; Montiel *et al.*, 2015; Myronovskyi and Luzhetskyy, 2016).

Phenazines are a large class of heterocyclic, nitrogen containing compounds produced by both Gram-positive and Gram-negative organisms (Laursen and Nielsen, 2004; Pierson and Pierson, 2010; Price-Whelan *et al.*, 2006). Phenazine metabolites are derived from either phenazine-1-carboxylic acid (PCA) (**1**) or phenazine-1,6-dicarboxylic acid (PDC) (**2**), which are biosynthesized by a conserved set of five essential genes, *phz*BDEFG, typically encoded in a single operon (Blankenfeldt and Parsons, 2014; Mavrodi *et al.*, 2010). While all phenazines share a dibenzo annulated pyrazine core structure, they exhibit incredible diversity in modifications of that core that result in potent and often therapeutically relevant bioactivities (Figure 1A). These activities can vary widely, including broad-spectrum antibiotic (saphenic acid, **4**), antitumor (esmeraldin B, **8**), and mosquito larvicidal (saphenamycin, **5**) activities (Geiger *et al.*, 1988; Laursen and Nielsen, 2004; Rui *et al.*, 2012). Phenazines play important ecological roles as well. For instance, pyocyanin (**3**) is an important signaling factor for its microbial producer *Pseudomonas aeruginosa* and a key facilitator of microbial survival in anoxic biofilms (Dietrich *et al.*, 2006, 2008; Ramos *et al.*, 2010). Because of the diverse and important bioactivities of phenazines, their structural similarity to cofactors in primary metabolism, and their abilities to modify gene expression, it has been suggested that phenazine metabolites blur the line between primary and secondary metabolism, and refute the paradigm that so-called ‘secondary’ metabolites are not essential for survival (Price-Whelan *et al.*, 2006).

Herein, we report the identification, capture, engineering, and heterologous expression of a unique cryptic phenazine-type BGC (*spz*) from the marine actinomycete *Streptomyces* sp. CNB-091. Refactoring of the *spz* cluster resulted in tremendous structural diversity of the streptophenazine suite of natural products, PDC-derived phenazines with variable alkyl side chains isolated from *Streptomyces* bacteria (Bunbamrung *et al.*, 2014; Kunz *et al.,* 2014; Liang *et al.*, 2017; Mitova *et al.,* 2008). Our refactoring efforts resulted in the production of previously unseen *N*-formylglycine analogues with antibacterial activity.

Results

Identification of the *spz* BGC

We previously sequenced the 8.2 Mb draft genome of the marine actinomycete *Streptomyces* sp. CNB-091 (GCA_000377965.1) and identified 41 metabolically diverse BGCs, revealing the astounding biosynthetic potential of this organism (Ray *et al.*, 2016). Beyond the previously characterized salinamides, a class of anti-inflammatory and antibacterial

depsipeptides (Degen *et al.*, 2014; Moore *et al.*, 1999; Trischman *et al.*, 1994), no other natural products have been reported from this strain. To explore the metabolic capacity of this organism, we cultured it in diverse growth conditions and monitored its molecular profile by Liquid Chromatography Mass Spectrometry (LCMS). We identified a group of orange pigmented molecules that were produced sporadically and at low titers. Comprehensive analysis of UV (Figure 1B) and high resolution mass spectrometry (HRMS) revealed production of a series of streptophenazine-like compounds with accurate mass and fragmentation identical to previously characterized streptophenazines A (**16**), B (**17**), F (**18**), and G (**19**), first isolated from *Streptomyces* sp. HB202 (Mitova *et al.,* 2008) (Figure 1C).

Inspection of the *S.* sp. CNB-091 genome revealed a single set of phenazine biosynthetic genes (*spz*16–21) co-localized with genes encoding the biosynthesis of polyketide extender unit(s) and polyketide synthase (PKS) extension cycle(s) (*spz*4–12) as well as nonribosomal peptide synthetase (NRPS) functions (*spz*15), suggestive of a putative streptophenazine BGC (*spz*) (Table 1, Figure 1D) (Weber *et al.*, 2015). While the presence of the PKS genes in the *spz* cluster can be rationalized by the installation of C-6 alkyl chains in streptophenazines (Figure 1C), there have been no previously reported streptophenazine structures that incorporate amino acid(s), indicative of NRPS biochemistry. Therefore, we speculated that the *spz* BGC may encode the biosynthesis of streptophenazine-like molecules with different chemical scaffolds. We thus set out to interrogate the *spz* BGC with the intent of not only exploring its full biosynthetic potential, but also to probe issues of BGC expression that might enable the future discovery of additional metabolites from *S.* sp. CNB-091.

Transformation-associated recombination (TAR) cloning and heterologous expression of the *spz* BGC

To exploit the full biosynthetic potential of the *spz* BGC and facilitate genetic manipulations, we captured the *spz* pathway using TAR cloning methodology (Kouprina and Larionov, 2016; Tang *et al.*, 2015; Yamanaka *et al.*, 2014). Initially, a 49.5 kb DNA region covering the proposed *spz* BGC was captured from *S.* sp. CNB-091 genomic DNA (gDNA), resulting in an approximately 60.3 kb construct, pSMM. Restriction digest was used to verify correct assembly of the construct isolated from yeast (Figure S1). However, pSMM proved to be unstable upon transformation and manipulation in *Escherichia coli,* limiting its use for future genetic alterations. Inspection of peripheral genes of the DNA insert captured into pSMM revealed almost 12 kb of DNA sequence unlikely associated with streptophenazine biosynthesis. Therefore, the *spz* BGC was recaptured using a PCR-based TAR method (Shao *et al.*, 2009) targeting strictly genes *spz*1–28. Eight approximately 5 kb DNA fragments covering the *spz* pathway were amplified and purified for further TAR cloning into the pCAP03 capture vector (Tang *et al.*, 2015). The resulting construct, pKDB01, remained stable upon transformation into *E. coli* (Figure S1). We then transferred pKDB01 into *E. coli* ET12567 (MacNeil *et al.*, 1992) for further conjugation into the *Streptomyces coelicolor* M1146 heterologous host and chromosomal integration into the ϕC31 attachment site (Gomez-Escribano and Bibb, 2011; Jones *et al.*, 2013). The resulting strain was cultured under a variety of growth conditions, but we ultimately failed to detect any streptophenazines in its metabolic profile (Figure 2).

While disappointing at first, the silence of the *spz* BGC within *S. coelicolor* M1146 presented an ideal opportunity to explore issues of expression within heterologous host systems and apply genetic tools to activate a silent pathway. The *spz* BGC is an excellent model system to probe these questions for several reasons: (1) phenazine metabolites are pigmented, facilitating visual detection and aiding purification processes and (2) streptophenazine-like compounds were detected in the native producer which simplifies the search for these molecules within the heterologous system. To awaken the *spz* cluster in the *S. coelicolor* M1146 host strain, we chose to harness synthetic biology tools (Kim *et al.*, 2015) to establish genetic control over the expression and regulation of the *spz* biosynthetic genes. Ultimately, we predict these tools can be applied to other BGCs from *S.* sp. CNB-091, allowing access to the entire biosynthetic potential of this organism.

Genetic manipulation of the pathway-specific *spz* regulatory genes

Three putative pathway-specific regulatory genes are found within the *spz* BGC: *spz*28 and *spz*3, encoding TetR and LysR homologs, respectively, which are hypothesized to be repressors, and *spz*24, which encodes a LuxR homolog predicted to be an activator of phenazine biosynthesis. To examine the role of the TetR homolog, we created a construct, pKDB02, with the *spz* pathway lacking the *spz*28 gene (Figure S1). The cosmid was then integrated into the genome of *S. coelicolor* M1146 but the profile of this mutant strain did not reveal any streptophenazine metabolites (Figure 2B). Using λ-Red recombineering, we then replaced the second putative repressor gene, *spz*3, with an apramycin resistance cassette in both plasmids, pKDB01 and the TetR deletion mutant pKDB02 (Figure S2). The resulting strains *S. coelicolor* M1146–pKDB01Δ*spz*3 and the double mutant –pKDB02Δ*spz*3 had identical metabolite profiles, characterized by trace production of four streptophenazine compounds (Figure 2B). However, this production was inconsistent and observed at levels much lower than in the native organism. This observation suggested that the LysR-homolog (Spz3) may serve as a repressor of the *spz* pathway. Finally, we introduced the well characterized strong constitutive promoter *ermE*p* (Bibb *et al.*, 1985) in front of *spz*24 (Figure S2). The resulting cosmid pKDB01-*ermE*p-spz*24 was then integrated into the genome of *S. coelicolor* M1146. LCMS analysis revealed that the upregulation of the *spz* LuxR homolog did not activate streptophenazine production (Figure 2B). Ultimately, genetic manipulations of the *spz* regulatory genes failed to achieve reliable production of streptophenazines in the heterologous host.

Refactoring the *spz* pathway with strong constitutive promoters

We performed reverse transcription-PCR (RT-PCR) to test expression of every gene in the cloned *spz* pathway integrated into the *S. coelicolor* M1146 genome. Results revealed that the *spz* cluster was not transcriptionally active in a heterologous environment (Figure 3A), which strongly suggests a malfunction with the cluster's native promoters. Therefore, we decided to bypass the native regulation of the pathway and refactor the *spz* BGC by introducing characterized, constitutively-active, actinomycete-compatible promoters in four locations based on the operonic structure of the cluster (Figure 3B). We created three cassettes (two unidirectional and one bidirectional) that harbor selection markers, promoters, and RBS sequences. Importantly, all cassettes were designed to lack *oriT* that is present in

readily available λ-Red cassettes (Gust *et al.*, 2003, 2004), to be compatible with the pCAP series of vectors that already harbor an *oriT* site (Yamanaka *et al.*, 2014).

We designed the first cassette, the *sp44–p21* cassette, to include two promoters: *sp44* and *p21,* facing opposite directions. The *sp44* promoter is a synthetic derivative of the *kasO** promoter from the coelimycin BGC and was shown to be 20-fold stronger than the *ermE** promoter, traditionally used in *Streptomyces* research (Bai *et al.*, 2015). The *sp44* promoter was introduced in front of the phenazine operon to allow robust production of the PDC core for further biosynthetic derivatization. The *p21* promoter is a derivative of *ermEp1*, with 3-fold greater strength than *ermE*p* (Myronovskyi and Luzhetskyy, 2016; Siegl *et al.*, 2013). This cassette also included the apramycin resistance gene *aac(3)IV* sandwiched between two FRT sites to allow excision of the resistance marker after a successful recombination event. Eight additional nucleotides were added between the *sp44* and FRT sequences to ensure an in-frame nonpolar deletion. The second cassette, named *actIp*. was designed to govern transcription of the PKS genes (*spz*4–12). This cassette included the *actI* promoter, RBS, the *actII-ORF4* gene required for activation of the *actI* promoter from the actinorhodin BGC (Fayed *et al.*, 2015; Fernández-Moreno *et al.*, 1991; Rowe *et al.*, 1998), and the *aac(3)IV* gene to leverage apramycin resistance for selection. Finally, the third cassette, *ermE*p*, was used to introduce the *ermE** promoter in front of the cytochrome *bd* oxidase-encoding genes (*spz*1,2). To avoid cross resistance, the ampicillin resistance gene *bla* was included to allow for selection in *E. coli*. Gibson assembly was used to stitch together functional elements as discussed in the STAR methods. Cassette sequences were confirmed by restriction digest and Sanger sequencing (Table S1). These cassettes were then used to replace putative native promoter regions within the *spz* cluster in the pKDB01 cosmid in a step wise manner using λ-Red recombineering (Figure S2B). Insertion of the *sp44–p21* cassette followed by excision of the apramycin resistance marker resulted in construct pKDB03. The latter was used for insertion of the *actIp* cassette to give construct pKDB04, and subsequent insertion of the *ermE*p* cassette yielded pKDB05. Each version of the refactored cosmid was conjugated into *S. coelicolor* M1146 for comparative metabolite profiling.

We performed an analogous RT-PCR experiment, as was done prior to refactoring, on RNA isolated from *S. coelicolor* M1146-pKDB03 carrying the refactored *spz* pathway. Interestingly, insertion of the bidirectional promoter cassette alone was sufficient to activate transcription of all biosynthetic genes within the cluster (Figure 3A).

Streptophenazine production by the refactored *spz* BGC

S. sp. CNB-091 (positive control) and *S. coelicolor* M1146 with integrated pCAP03 (negative control), pKDB01 (non-refactored *spz*), and pKDB03, −04, and −05 (refactored versions of *spz*) were grown for comparative metabolomics. After two days of culturing the production media turned dark orange for the strains only carrying the refactored *spz* BGC, yielding bright orange organic extracts (Figure 2C). Analysis of LCMS data of these pigmented extracts from *S. coelicolor* M1146-pKDB03, −04, and −05 revealed intense production of at least 38 streptophenazines, as corroborated by HRMS, UV, and MS/MS fragmentation (Figures 2A, S3, and Table S2). This chemical bounty included the 16 streptophenazine compounds previously detected in *S.* sp CNB091 (Figure 2A), but with

increased production levels, as judged by LCMS analysis. Importantly, the production profile proved to be consistent and achievable under a variety of growth conditions, which speaks to the advantage of genetic control. No changes in chemical diversity or production titers from *S. coelicolor* M1146-pKDB03, – 04, and –05 strains were observed (Figure S3). Therefore, it was decided to continue experiments with *S. coelicolor* M1146-pKDB03, the strain possessing the *spz* pathway with the fewest genetic alterations.

We next used molecular networking to compare *S. coelicolor* M1146-pKDB01, M1146-pKDB03, CNB-091, and the negative control (M1146-pCAP03) to identify analogs that might have been neglected in the initial visual analysis (Figure S4) (Wang *et al.*, 2016). Immediately, the cluster of nodes with molecular ions corresponding to the streptophenazines stood out (Figure 4). Each node in the putative streptophenazine cluster was inspected for association with streptophenazines by manual analysis of the corresponding MS/MS spectra, UV, and accurate mass (Table S2). The network identified 72 nodes corresponding to streptophenazine-type compounds; 67 of the nodes were found within the refactored strain (pink and blue nodes), 34 of which were found in both the refactored strain and in *S.* sp. CNB-091 (blue nodes). Five nodes were identified only in *S.* sp. CNB-091 (orange nodes) and one node included in the cluster (grey) was found in the negative control and is likely to be an artifact of the algorithm. Using the molecular network as a guide, manual inspection of the LCMS data ultimately revealed 112 streptophenazines produced by the refactored *spz* BGC. In most cases, extracting the LCMS data for the precursor mass found in the network resulted in more peaks than nodes in the network. For example, extracting LCMS data for *m/z* 411.2 revealed eight peaks corresponding to eight different streptophenazines, rather than the three nodes seen in the network. Several pairs of nodes (e.g., *m/z* 351.1 and 337.1) appeared separately in the molecular network and did not form clusters with other nodes corresponding to streptophenazines. This finding can be explained by low abundances of these masses in the LCMS data, which resulted in MS/MS fragment spectra with additional peaks from chemical noise, causing the algorithm to create separate clusters (Caraballo-Rodríguez *et al.*, 2017). Table S2 summarizes all streptophenazine products of the *spz* BGC and contains predicted molecular formulas for compounds where appropriate.

Interestingly, a number of nodes with even mass number were present in the network (e.g., *m/z* 510.2, 524.2, and 538.2), which suggested an additional nitrogen atom in the chemical formula. These molecular ions were present only in the sample from the refactored *spz* pathway. We identified three isomeric compounds with *m/z* 510.2. The early eluting compound (31.8 min) showed a distinctive loss of 117 Da, while the latter two isomers (36.2 and 37 min) instead had identical losses of 103 Da. These fragments suggested formylalanine and formylglycine attachments, respectively, with different methylation patterns at the C-1 carboxylic group (STAR methods). Such modifications are reasonable as the *spz* pathway contains the gene *spz*15 that encodes a putative adenylation protein with predicted specificity for alanine.

Isolation and characterization of streptophenazines

Production cultures of *S. coelicolor* M1146-pKDB03 were scaled up to 6 L to enable isolation and structural elucidation of observed streptophenazines (STAR methods). Ethyl acetate extraction of cultures resulted in 8 g of crude mixture which underwent multiple rounds of flash chromatography (C18 resin and silica resin). Final purification of the targeted compounds was achieved by reversed phase preparative and/or semipreparative HPLC. Ultimately, we isolated 15 streptophenazine analogs.

For tables with NMR assignments, spectra, and key COSY and HMBC correlations for all isolated compounds, see Data S1. Analysis of ^{1}H NMR and 2D spectra revealed that isolated compounds **16, 17, 13, 14, 19,** and **18** correspond to reported streptophenazines A, B, C, D, F, and G, respectively. To confirm stereochemistry, we performed circular dichroism (CD) measurements for compound **18** and compared results to the literature data for streptophenazine G with established 1'*S*,2'*R* absolute configuration (Kunz *et al.*, 2014; Liang *et al.*, 2017; Yang *et al.*, 2012) (Figure S5A). Due to the proximity of the C-1' stereocenter to the phenazine chromophore, this chiral center will affect the absorption of circularly polarized light and generate a characteristic shift in optical rotation (Cotton effect) evident in a CD spectrum. A strong negative Cotton effect at 251 nm, which mirrored that of previously characterized streptophenazine G, confirmed the *S* conformation of the C-1' stereocenter of **18**. ^{1}H NMR shifts of compound **18** were identical to those reported in the literature for streptophenazine G, indicating that the absolute configuration of compound **18** was also 1'*S*,2'*R*. Based on the CD result and the fact that stereochemistry of these centers remains unchanged for all reported streptophenazines, we propose the same stereochemistry for all compounds isolated in this work. To further support the stereochemistry at C-1', we performed bioinformatic analysis of the ketoreductase (KR) protein Spz7, which we presume catalyzes the stereospecific reduction of the β-keto intermediate to the 1'-hydroxyl group. KR domains of modular PKS enzymes define the configuration at α- and β-carbons, and conserved amino acid motifs that control the stereochemical outcome of this reduction have been defined (Caffrey, 2003; Keatinge-Clay, 2007; Smith and Tsai, 2007). While Spz7 shows relatively low amino acid sequence similarity to modular KR domains (Figure S5B), we identified the LDD amino acid motif characteristic of type B KR domains, which would generate the 1'*S* stereocenter in streptophenazines, consistent with the established stereochemistry.

Compounds isolated in this work included **9, 10, 11**, and **12,** which were individually isolated as orange powders. HRMS analysis showed that **9** and **10** share the same molecular formula $C_{24}H_{27}N_2O_5$ (*m/z* 423.1893 $[M + H]^+$ and *m/z* 423.1925 $[M + H]^+$, respectively), while compounds **11** and **12** have molecular formula $C_{25}H_{29}N_2O_5$ (*m/z* 437.2075 $[M + H]^+$ and *m/z* 437.2087 $[M + H]^+$, respectively). These four compounds appeared to be derivatives of streptophenazines A, B, G, and F, respectively, but were missing the C-1' oxymethine low field shift ($\delta_C \sim 72$) seen in previously characterized streptophenazines. Instead, HMBC correlations revealed a carbon shift indicative of a ketone carbonyl ($\delta_C \sim 198.3$) that showed correlations to both the phenazine core and the alkyl side chain. We suggest the name oxostreptophenazine A for compound **9,** oxo-streptophenazine B for compound **10,** oxostreptophenazine G for compound **11,** and oxo-streptophenazine F for compound **12**.

Compound **15** ($C_{23}H_{27}N_2O_5$, *m/z* $[M + H]^+$ 411.1912) shares a 2D structure with streptophenazine D, but terminates in a single methyl group rather than a branched chain (Figure 1D). Paralleling the established nomenclature of streptophenazines, we propose the name streptophenazine P for this compound. Additionally, compounds **22** and **23** had NMR spectra that appeared almost identical to streptophenazines G and F, however, the proton NMR signals associated with C-3', C-2', the C-2' carbomethoxy group (2'-COO**CH$_3$**), and C-1' were slightly shifted (Data S1), which indicated that compounds **22** and **23** were diastereomers at the C-2' position of streptophenazines G and F, respectively.

Two of the three detected compounds with *m/z* 510.2 were isolated (**20** and **21**), both with MS/MS fragmentation patterns suggestive of a formylglycine moiety adjoined to the streptophenazine core structure, most plausibly via the hydroxyl group at C-1'. The molecular formula $C_{27}H_{32}N_3O_7$ was established for compounds **20** and **21** (*m/z* $[M + H]^+$ 510.2236 and 510.2234, respectively). NMR spectra of these compounds contained a signal (δ_H 8.24, δ_C 161), suspected to be the formyl group, which showed COSY correlations to two protons (δ_H 4.22, 4.18) with a carbon shift of δ_C 40.25, as well as a COSY correlation to a broad peak δ_H 6.15 with no corresponding carbon shift in the HSQC shift (amide proton) (Data S1). Two doublets (δ_H 4.22, 4.18) showed key HMBC correlations to two carbonyl groups (δ_C 161 and 168) and COSY correlations to the amide and formyl protons, establishing their position as the glycine α-protons, and thus fully confirming the presence of the *N*-formylglycine moiety. Compound **20** was found to terminate the alkyl chain in two branched methyl groups while compound **21** terminated the chain in a single methyl group (Figure 1D). We propose the names streptophenazine Q and streptophenazine R, respectively, for compounds **20** and **21**.

Anti-microbial activity of streptophenazines

A series of four streptophenazines, **9**, **13**, **16**, and **20**, were tested for antibiotic activity against two Gram-positive organisms: group A *Streptococcus* (GAS) and methicillin-resistant *Staphylococcus aureus* (MRSA) TCH1516, and two Gram-negative organisms: *Acinetobacter baumannii* 5075 and *Klebsiella pneumoniae* 1100. The four test compounds were selected because they share the same alkyl chain but differ in the C-1' substituent (keto, hydroxyl, or formylglycine) and methylation state of the C-1 carboxylic group. Results showed that compounds **9**, **13**, and **16** had minimum inhibitory concentration (MICs) higher than their tested concentrations, indicating weak or absent antimicrobial activity (Table S3). However, compound **20,** the *N*-formylglycine analogue, showed significantly more potent activity than the hydroxyl or ketone containing analogs against GAS (MIC 2.5 μg/mL). Compound **20** also inhibited the growth of MRSA and the tested Gram-negative pathogens, while no such activity was observed for the other analogs. This set of structurally similar compounds allows us to infer structure-activity relationships of streptophenazines, and this data suggests that the formylglycine moiety leads to superior bioactivity.

Spz15 – a unique standalone adenylation protein required for formylglycine modification

Arguably the most interesting aspect of the *spz* biosynthetic pathway is the production of the *N*-formylglycine-containing streptophenazines Q and R (**20, 21**). Although common in ribosomal protein biosynthesis, *N*-formylation is a rare biosynthetic tailoring modification in

NRPS-synthesized peptides, famously performed in linear gramicidin A by a specific formylation domain (Kessler *et al.*, 2004; Schoenafinger *et al.*, 2006). Bioinformatic analysis suggests that *spz*15 encodes an adenylation protein, and identification of the GXXGXPKG motif for a phosphate binding loop and AMP-binding Pfam domain (PF00501) reinforced this hypothesis (Figure S6) (Perego *et al.*, 1995). However, neither a canonical condensation domain nor a formylation domain is present in the *spz* cluster. The lack of a traditional suite of NRPS domains suggested that perhaps Spz15 is solely responsible for the addition of the *N*-formylglycine moiety. To prove involvement of the Spz15 protein in the biosynthesis of the amino acid-containing compounds, we deleted the *spz*15 gene. The resulting mutant construct (pKDB03Δ*spz*15) was confirmed by restriction digest (Figure S6) and integrated into the genome of the heterologous host. LCMS analysis of organic extracts of the mutant strain showed no signs of streptophenazines containing the *N*-formylglycine moiety. However, production of all other streptophenazines remained unchanged, ruling out polar effects on downstream genes in the *spz*15 deletion mutant (Figure S6).

Phylogenetic analysis of Spz15 revealed that this enzyme is strikingly distinct from canonical adenylation domains and instead forms a separate evolutionary clade (Figure 5). Analysis of genomes containing Spz15-like proteins revealed that these putative adenylation proteins are all found adjacent to phenazine biosynthesis genes (Figure S7). This arrangement indicates that *spz*15 may represent a genetic hook into unusual hybrid phenazine-type biosynthetic pathways. Biochemical characterization and investigation of the role of Spz15 in this unusual addition of an *N*-formylglycine moiety to streptophenazine molecules are currently ongoing.

Discussion

The discovery of the unique hybrid PHZ/PKS/NRPS BGC encoded in the genome of marine *S.* sp. CNB-091 uncovered a pathway that involves unusual biosynthetic transformations and yields a tremendous wealth of chemical products. Fundamentally, the *spz* BGC is arranged as a series of subclusters, each responsible for a distinct aspect of the biosynthesis of the final molecules, evoking the concept of a ‘cluster of clusters’ (Figure 3B). A set of contiguous genes (*spz*16–21) are responsible for the biosynthesis of the PDC core, which serves as the starter unit for a single PKS extension with an unusual long branched chain malonyl-CoA constructed by a crotonyl-CoA carboxylase (CCR)-like mechanism (Figure 6) (Wilson and Moore, 2012). To date, there is only one other example of a PKS extension of a phenazine core substrate, described in the biosynthesis of the esmeraldins and saphenamycin phenazine metabolites (Rui *et al.*, 2012).

A phylogenetically distinct adenylation enzyme (Spz15), transcribed from the same operon as the *spz*16–21 genes, is unambiguously involved in the addition of an *N*-formylglycine moiety to streptophenazines, generating unusual hybrid molecules with potent bioactivities (Figure S6). Finally, two genes at the west bound of the cluster (*spz*1,2) encode cytochrome *bd* oxidase enzymes, which may represent a resistance mechanism for the phenazine-producing strain *S.* sp. CNB-091 (Voggu *et al.*, 2006). The *spz* BGC represents unique biosynthetic machinery that does not fit neatly within any biosynthetic systems described, and therefore, opens the doors for future mechanistic biochemical studies (Figure 6).

Rewriting native genetic elements in the *spz* gene cluster with optimized promoters and RBS sequences allowed us to circumvent issues associated with the complex network of native regulation and ultimately gain mastery over this unique biosynthetic system. Interestingly, inserting just the first bidirectional promoter cassette resulted in activation of transcription of all *spz* biosynthetic genes (Figure 3). It is possible that production of the phenazine core (PDC) is required to trigger expression of the remaining biosynthetic genes ('self-induction'). Certain small molecules can bind to specific DNA regions resulting in activation of transcription of the downstream genes (Chalabaev *et al.*, 2008; Denison and Kodadek, 1998). Alternatively, the strength of the *sp44* promoter could be sufficient to activate transcription of the genes beyond the phenazine-encoding operon, and there is precedent of a single promoter driving transcription of multiple genes, spanning over 10 kb (Amagai *et al.*, 2017; Luo *et al.*, 2015).

Ultimately, refactoring the *spz* pathway produced an incredible wealth of chemical diversity (112 streptophenazines) that can serve as a library of compounds, much like a combinatorial chemist might produce. This diversity-oriented biosynthesis is generally considered an evolutionary advantage (Fischbach and Clardy, 2007), as diverse analogs may have different biological targets or act synergistically as an antibiotic cocktail (Wu and Seyedsayamdost, 2017). This may explain the lack of production of the *N*-formylglycine analogs by the wild type strain; bioactivity testing revealed that these compounds did possess distinct bioactivity, and thus they may be produced only under certain environmental and/or physiological stress. Therefore, amino-acid containing streptophenazines are excellent starting points for further chemical derivatization with the goal of ultimately achieving superior antibiotic properties.

The problem regarding inconsistent or absent production of secondary metabolites is a common occurrence, both within native organisms and heterologous hosts, and the disconnect between the genetically encoded biosynthetic potential of an organism and the actual suite of compounds produced is immense. The issue addressed here is not just successful activation of cryptic BGCs, but exploitation of the full range of capabilities housed by these pathways. Through engineering, we were able to achieve production of broad chemical diversity, including compounds not seen in the native organism that contain an unprecedented formylglycine moiety that imparts enhanced antibiotic activity. This example suggests that even transcriptionally active clusters may benefit from refactoring.

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, Dr. Bradley Moore (bsmoore@ucsd.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Streptomyces sp. CNB-091 was the original streptophenazine producer and *S. coelicolor* M1146 was used for heterologous expression of the streptophenazine biosynthetic gene cluster. *E. coli* DH10B cells were used for plasmid storage and replication, *E. coli* ET12567 and *E. coli* ET12567/pUB307 were used for triparental intergeneric conjugation. *E. coli*

BW25113 and *E. coli* DH5α/BT340 (Cherepanov and Wackernagel, 1995; Datsenko and Wanner, 2000) were used for λ-Red recombineering. Culture conditions followed literature procedures and are explained in Method Details.

METHOD DETAILS

General methods—Primers, plasmids and strains used in this study are summarized in Tables S5–S7.

Bacterial strains and growth conditions—*Streptomyces* strains were grown in TSBY liquid medium (3% tryptic soy broth, 10.3% sucrose, 0.5% yeast extract) for isolation of genomic DNA and on SFM agar plates (2% D-mannitol, 2% soya flour, 2% agar) for conjugation and strain maintenance. Liquid cultures were grown at 30 °C with shaking at 220 rpm in a rotary shaker. Solid cultures were grown at 30 °C for 10–12 days.

For streptophenazine production, seed medium (TSBY) was inoculated with *S. coelicolor* M1146 spore suspension and cultured for 36 hrs. A 5% inoculum of seed culture were used for fermentation medium (also TSBY medium). Fermentation was carried out at 30 °C and 220 rpm in a rotary incubation for 3 days.

E. coli strains were grown in Luria-Bertani (LB) broth or on LB agar at 37 °C with appropriate antibiotics for selection (apramycin 50 μg/mL, ampicillin 100 μg/mL, chloramphenicol 25 μg/mL, kanamycin 50 μg/mL, nalidixic acid 25 μg/mL). For conjugation purposes, *E. coli* was grown using 2TY medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl) with appropriate antibiotic selection.

Software and tools used—The web-based bioinformatics program antiSMASH was used to analyze the whole genome sequence of *S.* sp. CNB-091 (Weber *et al.*, 2015). The sequence of the *spz* gene cluster was further analyzed using BLAST (Altschul *et al.*, 1990). Protein sequence alignments were made using Clustal Omega software (Sievers and Higgins, 2014). Geneious software was used for all plasmid maps, promoter cassette design, and phylogenetic tree reconstruction (Kearse *et al.*, 2012). HRMS data were analyzed with MassHunter software (Agilent), low resolution MS data was analyzed with Compass DataAnlalysis software (Bruker Daltonic). NMR data was recorded with Bruker Topspin 2.1.6 software and analyzed with MestreNova software. GNPS (gnps.ucsd.edu) was used for molecular networking analysis (Wang *et al.*, 2016). The spectral networks were imported into Cytoscape 3.6.0 (Shannon *et al.*, 2003).

Preparation of genomic DNA and PCR fragments for TAR cloning—*S.* sp. CNB-091 gDNA was isolated following standard procedures from *Practical Streptomyces Genetics* (Kieser *et al.*, 2000). Approximately 200 μg of gDNA was digested with SnaBI, in an overnight reaction at 30 °C. Digested gDNA fragments were precipitated and cleaned by isopropanol precipitation. The resulting gDNA pellet was dissolved in 50 μL 10 mM Tris buffer (pH 7).

Takara PrimeSTAR HS DNA polymerase with GC buffer was used to amplify the eight fragments of the *spz* cluster for TAR cloning using gDNA as a template. PCR reactions were

carried out in a BioRad MyCycler with gradient option. Fragments were designed with average 200 bp overlap to neighboring fragments and the first and last fragment with 40 bp overlap to pCAP03 capture vector. For primers see Table S5. PCR conditions were as follows:

PCR reaction:	**Vol (μL)**
Milli-Q Water	20.5
Primer 1 (10 μM)	1
Primer 2 (10 μM)	1
Template gDNA	1
2X GC Buffer	25
dNTPs (10 mM)	1
Primestar polymerase	0.5
Total volume	50

PCR cycling condition (30 cycles)		
Initial denaturation	98 °C	1 min
Denaturation	98 °C	10 s
Annealing	60 °C	10 s
Extension	72 °C	5 min
Final extension	72 °C	10 min
	4 °C	∞

PCR fragments were cut from agarose gels and purified using the Qiagen QIAquick Gel Extraction kit according to the manufacturer's instructions.

TAR cloning of the *spz* cluster—Initially, gDNA was used as a template to directly capture the 48.5 kb region encompassing the *spz* BGC following general procedures described by (Tang *et al.*, 2015; Yamanaka *et al.*, 2014). The *spz* pathway-specific capture vector was constructed by introducing two PCR-amplified 1 kb homology arms corresponding to flanking regions of the targeted region into pCAP03. *S. cerevisiae* VL6–48N (Table S4) was grown to an OD_{600} of 0.7 to 1.0 in 50 mL of YPD medium (2% D-glucose, 1% yeast extract, 2% peptone) supplemented with adenine hemisulfate salt (100 mg/L) at 30 °C with shaking. Cells were harvested and washed with ice cold water and osmotically stabilized in 1M sorbitol at 4 °C overnight prior to spheroplast preparation. Preparation of spheroplast cells was carried out with zymolase at a final concentration of 0.1 mg/mL with 30–40 minutes of incubation. Spheroplast cells were mixed with approximately 0.5 μg of gDNA and 0.5 μg of linearized specific capture vector. Transformed spheroplasts were mixed with 8 mL of synthetic drop-out tryptophan top agar at 55 °C and overlaid on synthetic tryptophan drop-out (SD-Trp) top agar containing 5-fluoorotic acid (5-FOA) for

selection. Plates were incubated at 30 °C for 4–5 days before transformants were picked and screened by PCR for positive colonies. Plasmids were extracted and then transferred into *E. coli* DH10B cells by electroporation. Plasmids were then purified from antibiotic resistance clones, and the resulting constructions confirmed by restriction analysis.

PCR-based cloning (Shao *et al.*, 2009) followed the same procedures described above and by Yamanaka *et al.* (2014) and Tang *et al.* (2015), mixing 200 ng of each PCR fragment with the linearized capture vector pCAP03.

Heterologous expression of the *spz* cluster—pKDB01 and its derivatives were transformed into *E. coli* ET12567 by heat shock at 42 °C for 55 s, and then transferred to *S. coelicolor* M1146 by triparental intergeneric conjugation facilitated by *E. coli* ET12567/pUB307 (MacNeil *et al.*, 1992) (Table S4). Exconjugants were grown on soy flour mannitol agar containing $MgCl_2$ (10 mM). After incubation at 30 °C for 18 h, plates were overlaid with 1 mL water containing 0.5 mg nalidixic acid followed by 1 mL water containing 1 mg apramycin and then grown until appearance of exconjugants, which were then replated for a second round of selection.

RNA extraction and RT-PCR analysis—*S. coelicolor* M1146-pKDB01 and -pKDB03 were grown following the methods for production of streptophenazines. After 24 h growth in production media, 1 mL of culture was sampled, pelleted and washed with RNase free water. RNA was extracted using Ambion's PureLink RNA extraction kit following methods for bacterial cells. Briefly, cells were lysed and homogenized using a 20G needle attached to an RNase-free syringe. RNA was bound to a spin column, washed, and eluted with RNase-free water. Concentrations were analyzed using Invitrogen Qubit Fluorometric Quantitation. For primers see Table S6.

RT-PCR was performed using Qiagen's OneStep RT-PCR kit as follows:

PCR reaction:	Vol (μL)
RNase-free water	23.5
5x buffer	10
5x Q-solution	10
dNTPs (10 mM each)	2.0
Primer 1 (30 μM)	1
Primer 2 (30 μM)	1
Enzyme mix	2.0
Template RNA	0.5
Total volume	50

RT-PCR cycling condition (30 cycles)

Reverse transcription	50 °C	30 min
Initial PCR inactivation	90 °C	15 min
Denaturation	94 °C	1 min
Annealing	67 °C	1 min or 1.5 min
Extension	72 °C	1 min
Final extension	72 °C	10 min
	4 °C	∞

Promoter cassette construction—The vector pCAP03 was used as the backbone for cloning and construction of all cassettes. pCAP03 was digested with XhoI and NdeI restriction enzymes and recovered from a 0.7% agarose gel. Apramycin and ampicillin resistance genes were PCR amplified from in house plasmids (pKY01 and pETduet-1, respectively). For the bidirectional promoter cassette, the apramycin resistance gene was amplified with primers that contained flanking FRT sites added synthetically. Promoters regions were PCR amplified from in-house plasmids (not published) with primers containing appropriate homology regions for further Gibson assembly. Target fragments were recovered from a 0.7% agarose gel. Gibson assembly was used to combine the digested and purified PCR products with the vector according to manufacturer's guidelines, followed by transformation into *E. coli* DH10B competent cells using heat shock at 42 °C for 55 s. Cells were transferred onto LB plates with appropriate antibiotics for selection. After incubation overnight, transformants were picked and inoculated into 10 mL LB broth containing appropriate antibiotics and incubated at 37 °C with shaking at 220 rpm for approximately 16 h. Plasmids were isolated and identity was confirmed by restriction analysis and DNA sequencing. Finally, promoter cassettes were PCR amplified using primers with 50 bp overlap to the adjacent sequence of the targeted DNA within the *spz* cluster (Table S5).

DNA manipulation—All genetic manipulations (gene deletions and promoter replacements) in pKDB01 and pKDB02 were carried out using the λ-Red recombination-mediated PCR-targeted gene deletion/insertion method, as described by Gust *et al.* (2003). Electrocompetent *E. coli* BW25113 carrying the pIJ790 plasmid were electroporated (Eppendorf Electroporator 2510) at 2,500 V in a 2-mm cuvette with the pKDB01 or pKDB02 cosmid. Expression of the λ-Red genes was induced with arabinose (10 mM) and the cells were then transformed with the promoter cassette described above and incubated at 37 °C. Cosmid DNA was then isolated and tested by restriction digest and Sanger sequencing to confirm correct genetic alteration, prior to conjugation into *S. coelicolor* M1146. For the creation of pKDB03, the pKBD01 cosmid containing the *sp44-p21* promoter cassette was transformed into *E. coli* DH5α cells containing the temperature sensitive FLP recombination plasmid BT340. FLP production and loss of the plasmid were induced at 42 °C. Cosmid DNA was isolated and the successful excision of the resistance cassette was verified by restriction digest and Sanger sequencing prior to integration into *S. coelicolor* M1146.

For analytical restriction digests, 1.5 μL 10X CutSmart Buffer (NEB), 500 ng DNA, 1 unit of restriction enzyme, and MQ water to a total volume of 15 μL were incubated for 2 hrs at 37 °C and then heat inactivated at 65 °C for 20 minutes. For preparative restriction digests, 50 μL reactions including 5 μL 10X CutSmart Buffer (NEB), 1 μg DNA, 2 units of restriction enzyme, and MQ water were incubated for 5 hrs at 37 °C prior to heat inactivation and further use.

Streptophenazine metabolite analysis—Production cultures (liquid and solid) were extracted by an equal volume of ethyl acetate. Liquid cultures were adjusted to pH ~4 with formic acid prior to extraction. The organic phase was evaporated and samples were reconstituted in methanol and filtered through a 0.2 μm filter for subsequent LCMS analysis.

A 10 μL aliquot of methanol extract was injected onto a Phenomenex Luna C18 reversed-phase HPLC column (5 μm, 250 mm × 4.6 mm) and analyzed with a Bruker Amazon Ion Trap MS system coupled to an Agilent 1260 Infinity LC system. A solvent system of acetonitrile and water both containing 0.1% formic acid (v/v) was used. Samples were eluted over a 60 min method with a gradient from 10 to 45% acetonitrile over 15 minutes, 45 to 65% over the next 30 min, and then to 100% over 10 min. 100% acetonitrile was held for 1 min before concentration was dropped to 10%. Flow rate was 0.75 mL/min

Eluent was detected using electrospray ionization-mass spectrometry (ESI-MS) monitoring *m/z* 70–2,200 in positive mode with a speed of 32,500 *m/z* /s. HRMS was carried out on an Agilent 6530 Accurate-Mass Q-TOF in positive mode, low resolution data was acquired using an Agilent 1260 Infinity IonTrap in positive mode.

Molecular networking—Input data for molecular networking was generated through conversion of the LCMS/MS raw data to .mzXML data format using ProteoWizard MSConvert software. Data were submitted to the molecular networking workflow at the GNPS online platform (gnps.ucsd.edu).

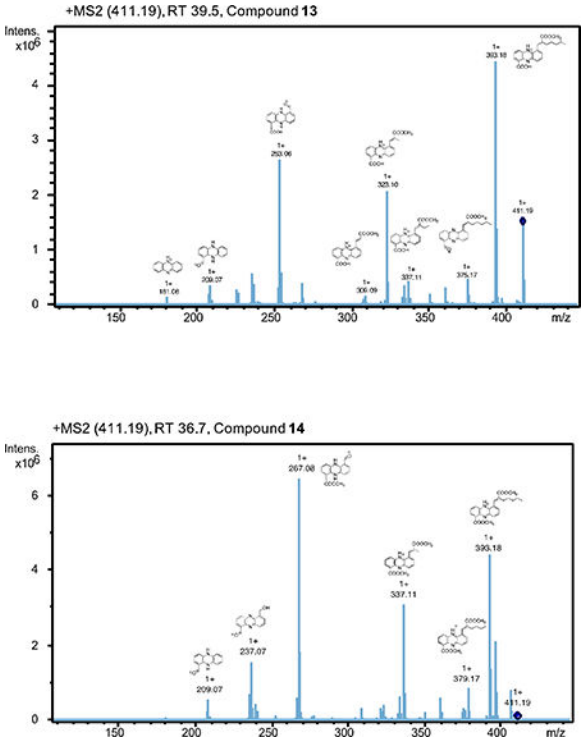
The GNPS molecular network was created using parameters as follows: A molecular network was created using the online workflow at GNPS. The data was clustered with MS-Cluster with a parent mass tolerance of 0.5 Da and a MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra. Further, consensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.5 and more than 2 matched peaks. Edges between two nodes were kept in the network if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. See table below for parameters.

Parameter	Value
Pairs_Min_Cosine	0.5

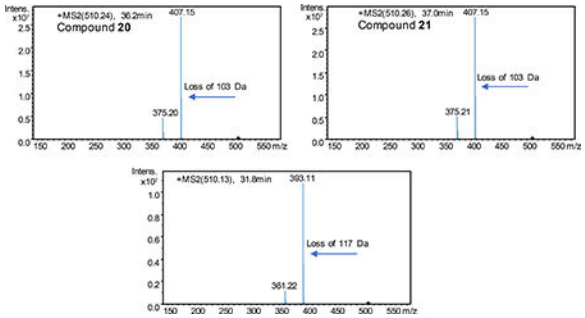
Parameter	Value
Analog_Search	0
Tolerance.PM_tolerance	0.5
Tolerance.ion_tolerance	0.5
Min_matched_peaks	2
TopK	10
Cluster_Min_Size	2
Maximum_component_size	100
Min_peak_int	0.0
Filter_Stddev_peak_int	0.0
Run_MScluster	on
Filter_precursor_window	0
Filter_library	1
Window_filter	0
Score_threshold	0.7
Min_matched_peaks_search	6
Max_shift_mass	100.0

Molecular networks can be viewed at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=24d515f76b4c4d6694a934234435b9bf (Related to Figure 2, Figure 4, Figure S4 and Table S2) and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9a2d02c583c24422bfaa16e2f8ae9ceb (Related to Figure 2 and Figure S3)

HRMS, UV, and MS/MS data for each precursors mass in the network were inspected to confirm relationship to streptophenazines. Example of characteristic MS/MS below:



MS/MS spectra of three *m/z* 510.2 compounds, including the two *N*-formylglycine-containing compounds (**20** and **21**) and the proposed formylalanine-containing compound:



Purification and isolation of streptophenazines—Approximately 6 L of 14-day-old culture of *S. coelicolor* M1146-pKDB03 were adjusted to pH ~4 and extracted with an equal volume of ethyl acetate. Organic layers were combined and solvent was evaporated, yielding 8 g of orange solid. The extract was redissolved in MeOH and separated on a bench-packed C18 column, which was washed with increasing MeOH concentrations. Fractions containing streptophenazine metabolites were combined, and the organic solvent was evaporated under vacuum followed by lyophilization. The sample was further purified using a column of silica gel (12 nm, 70 mesh) using a mixture of hexane/ethyl acetate.

Further purification was achieved by preparative HPLC using a Phenomenex Luna C18 column (5 μm, 100 mm x 2 mm), along with an Agilent Technologies system composed of a PrepStar pump, a ProStar 410 autosampler, and a ProStar UV detect (Agilent Technologies,

Inc, Santa Clara, USA). Acetonitrile and water, both with 0.1% trifluoroacetic acid (TFA) (v/v), were used as mobile phase. Samples were eluted over 60 min at a flow rate of 15 mL/min. Samples were subsequently subjected to further semi-preparative HPLC purification, using a Phenomenex Luna C18 reversed-phase HPLC column 5 μm, 250 mm × 4.6 mm) using acetonitrile and water, both containing 0.1% TFA, flow rate 2.5 mL/min. Final fractions yielded 1 mg of compounds **13, 14, 15, 22** and **23,** 3 mg compounds **9**–**12,** and over 5 mg of compounds **16**–**21.** All stages of purification were monitored by LCMS analysis.

Structural characterization of streptophenazines—All NMR data were collected using at the UCSD Skaggs School of Pharmacy and Pharmaceutical Sciences NMR Facility on a 600 MHz Bruker NMR spectrometer (Topspin 2.1.6 software, Bruker) consisting of a Magnex 54 mm bore superconducting magnet operating at 14.1 Tesla with a 1.7 mm cryoprobe. Deuterated chloroform containing 1% (v/v) TMS standard was used as a solvent. See Data S1.

Bioactivity testing of streptophenazines and MIC determination—MIC values were determined using broth microdilution in accordance with the Clinical Laboratory Standards Institute (CLSI) guidelines using cation-adjusted Mueller Hinton Broth (MHB) with minor modifications. Briefly, bacteria were grown to mid-log phase (OD_{600nm} = 0.4) at 37 ºC while shaking except GAS which was grown under static condition. Bacterial cells were then centrifuged, washed, and diluted in PBS to obtain 2×10^6 cfu/mL, and 10 μL was added to individual wells of a 96-well plate containing 170 μL MHB.

Serial dilutions of compounds **9**, **13**, **16** starting at 500 μg/mL and compound **20** at 400 μg/mL were made in a separate plate, 20 μL of each compound was then added to the test plate. The plates were sealed with parafilm and incubated at 37 ºC for 24 h. The turbidity of each plate was measured at OD_{600nm} using the EnSpire Alpha plate reader. MIC was defined as the lowest concentration of the drugs that inhibited bacterial growth.

Phylogenetic analysis of Spz15—The Geneious neighbor-joining tree building method was used for phylogenetic tree reconstruction. Proteins selected for analysis included the closest uncharacterized homologs of Spz15 found using protein BLAST analysis, characterized discrete adenylation proteins, and modular NRPS-associated adenylation domains from characterized BGCs with specificity for alanine or glycine. The sequences of A domains were trimmed according to the annotation by antiSMASH and the MIBiG repository (Medema *et al.*, 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

For bioactivity assays and MIC determination, each sample was repeated three times.

DATA AND SOFTWARE AVAILABILITY

All promoter cassette plasmids are available at Addgene: pCAP03-*sp44/p21* cassette (Addgene ID 120232), pCAP03-*actIp* cassette (Addgene ID 120233), pCAP03*ermE*p* cassette (Addgene ID 120234). The public datasets used for the GNPS molecular networking are available at: ftp://massive.ucsd.edu/MSV000083081 and ftp://

massive.ucsd.edu/MSV000083082. The molecular networks can be accessed at: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=24d515f76b4c4d6694a934234435b9bf (for Figure 3, Figure S4, and Table S2) and: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=9a2d02c583c24422bfaa16e2f8ae9ceb (for Figure S3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to P.R. Jensen and W. Fenical (Scripps Institution of Oceanography) for providing *Streptomyces* sp. CNB-091, M. Bibb (John Innes Center) for *S. coelicolor* M1146, B.M. Dungan (UC San Diego) for assistance with NMR, F.A. Tezcan and R.G. Alberstein (UC San Diego) for assistance with CD, and L. Ray, S.K. McKinnie, T. de Rond, and P.A. Jordan (Scripps Institution of Oceanography) for helpful discussions. This work was supported by grants from the National Institutes of Health through grants R01-GM085770 and R01-AI117712 to B.S.M. and NIH Marine Biotechnology Training Grant Predoctoral Fellowship (T32 GM067550) to K.D.B.

References

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol 215, 403–410. [PubMed: 2231712]

Amagai K, et al. (2017). Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host. Sci. Rep 7, 3382. [PubMed: 28611443]

Bai C, Zhang Y, Zhao X, Hu Y, Xiang S, Miao J, Lou C, and Zhang L (2015). Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. Proc. Natl. Acad. Sci. U.S.A 112, 12181–12186. [PubMed: 26374838]

Bentley SD, et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417, 141–147. [PubMed: 12000953]

Bibb MJ, Janssen GR, and Ward JM (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of *Streptomyces* erythraeus. Gene 38, 215–226. [PubMed: 2998943]

Blankenfeldt W, and Parsons JF (2014). The structural biology of phenazine biosynthesis. Curr. Opin. Struct. Biol 0, 26–33.

Bunbamrung N, Dramae A, Srichomthong K, Supothina S, and Pittayakhajonwut P (2014). Streptophenazines I–L from *Streptomyces* sp. BCC21835. Phytochem. Lett 10, 91–94.

Caffrey P (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. Chembiochem. 4, 654–657. [PubMed: 12851937]

Caraballo-Rodríguez AM, Dorrestein PC, and Pupo MT (2017). Molecular interkingdom interactions of endophytes isolated from *Lychnophora ericoides*. Sci. Rep 7, 5373–5387. [PubMed: 28710400]

Chalabaev S, Turlin E, Bay S, Ganneau C, Brito-Fravallo E, Charles J-F, Danchin A, and Biville F (2008). Cinnamic acid, an autoinducer of its own biosynthesis, is processed via Hca enzymes in *Photorhabdus luminescens*. Appl. Environ. Microbiol 74, 1717–1725. [PubMed: 18245247]

Cherepanov PP, and Wackernagel W (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibioticresistance determinant. Gene 158, 9–14. [PubMed: 7789817]

Datsenko KA, and Wanner BL (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A 97, 6640–6645. [PubMed: 10829079]

Degen D, et al. (2014). Transcription inhibition by the depsipeptide antibiotic salinamide A. ELife 3, e02451. [PubMed: 24843001]

Demain AL, and Fang A (2000). The natural functions of secondary metabolites. Adv. Biochem. Eng. Biotechnol 69, 1–39. [PubMed: 11036689]

Denison C, and Kodadek T (1998). Small-molecule-based strategies for controlling gene expression. Chem. Biol 5, R129–R145. [PubMed: 9653545]

Dietrich LEP, Price-Whelan A, Petersen A, Whiteley M, and Newman DK (2006). The phenazine pyocyanin is a terminal signaling factor in the quorum sensing network of *Pseudomonas aeruginosa*. Mol. Microbiol 61, 1308–1321. [PubMed: 16879411]

Dietrich LEP, Teal TK, Price-Whelan A, and Newman DK (2008). Redox-active antibiotics control gene expression and community behavior in divergent bacteria. Science 321, 1203–1206. [PubMed: 18755976]

Fayed B, Ashford DA, Hashem AM, Amin MA, Gazayerly ONE, Gregory MA, and Smith MCM (2015). Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in *Streptomyces* spp. and combinatorial biosynthesis. Appl. Environ. Microbiol 81, 8402–8413. [PubMed: 26431970]

Fernández-Moreno MA, Caballero J, Hopwood DA, and Malpartida F (1991). The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. Cell 66, 769–780. [PubMed: 1878971]

Fischbach MA, and Clardy J (2007). One pathway, many products. Nat. Chem. Biol 3, 353–355. [PubMed: 17576415]

Flett F, Mersinias V, and Smith CP (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. FEMS Microbiol. Lett 155, 223–229. [PubMed: 9351205]

Freestone TS, Ju K-S, Wang B, and Zhao H (2017). Discovery of a phosphonoacetic acid derived natural product by pathway refactoring. ACS Synth. Biol 6, 217–223. [PubMed: 28103011]

Geiger A, Keller-Schierlein W, Brandl M, and Zähner H (1988). Metabolites of microorganisms. 247 phenazines from *Streptomyces antibioticus*, strain Tü 2706. J. Antibiot. (Tokyo) 41, 1542–1551. [PubMed: 3058669]

Gomez-Escribano JP, and Bibb MJ (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb. Biotechnol 4, 207–215. [PubMed: 21342466]

Gomez-Escribano JP, and Bibb MJ (2014). Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. J. Ind. Microbiol. Biotechnol 41, 425–431. [PubMed: 24096958]

Gust B, Challis GL, Fowler K, Kieser T, and Chater KF (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. U.S.A 100, 1541–1546. [PubMed: 12563033]

Gust B, Chandra G, Jakimowicz D, Yuqing T, Bruton CJ, and Chater KF (2004). Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. Adv. Appl. Microbiol 54, 107–128. [PubMed: 15251278]

Jones AC, Gust B, Kulik A, Heide L, Buttner MJ, and Bibb MJ (2013). Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 Gene cluster. PLoS ONE 8, e69319. [PubMed: 23874942]

Kearse M, et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28, 1647–1649. [PubMed: 22543367]

Keatinge-Clay AT (2007). A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol 14, 898–908. [PubMed: 17719489]

Kessler N, Schuhmann H, Morneweg S, Linne U, and Marahiel MA (2004). The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. J. Biol. Chem 279, 7413–7419. [PubMed: 14670971]

Kieser T, Bibb MJ, Buttner MJ, Chater KF, and Hopwood DA (2000). Practical *Streptomyces* Genetics (John Innes Foundation).

Kim E, Moore BS, and Yoon YJ (2015). Reinvigorating natural product combinatorial biosynthesis with synthetic biology. Nat. Chem. Biol 11, 649–659. [PubMed: 26284672]

Koehn FE, and Carter GT (2005). The evolving role of natural products in drug discovery. Nat. Rev. Drug Discov 4, 206–220. [PubMed: 15729362]

Kouprina N, and Larionov V (2016). Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma 125, 621–632. [PubMed: 27116033]

Kunz AL, Labes A, Wiese J, Bruhn T, Bringmann G, and Imhoff JF (2014). Nature's lab for derivatization: new and revised structures of a variety of streptophenazines produced by a sponge-derived *Streptomyces* strain. Mar. Drugs 12, 1699–1714. [PubMed: 24670532]

Laursen JB, and Nielsen J (2004). Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. Chem. Rev 104, 1663–1686. [PubMed: 15008629]

Liang Y, Chen L, Ye X, Anjum K, Lian X-Y, and Zhang Z (2017). New streptophenazines from marine *Streptomyces* sp. 182SMLY. Nat. Prod. Res 31, 411–417. [PubMed: 27097765]

Luo Y, Zhang L, Barton KW, and Zhao H (2015). Systematic identification of a panel of strong constitutive promoters from *Streptomyces albus*. ACS Synth. Biol 4, 1001–1010. [PubMed: 25924180]

MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, and MacNeil T (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111, 61–68. [PubMed: 1547955]

Mavrodi DV, et al. (2010). Diversity and evolution of the phenazine biosynthesis pathway. Appl. Environ. Microbiol 76, 866–879. [PubMed: 20008172]

Medema MH, et al. (2015). Minimum Information about a Biosynthetic Gene cluster. Nat. Chem. Biol 11, 625–631. [PubMed: 26284661]

Mitova MI, Lang G, Wiese J, and Imhoff JF (2008). Subinhibitory concentrations of antibiotics induce phenazine production in a marine *Streptomyces* sp. J. Nat. Prod 71, 824–827. [PubMed: 18396903]

Montiel D, Kang H-S, Chang F-Y, Charlop-Powers Z, and Brady SF (2015). Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. Proc. Natl. Acad. Sci. U.S.A 112, 8953–8958. [PubMed: 26150486]

Moore BS, Trischman JA, Seng D, Kho D, Jensen PR, and Fenical W (1999). Salinamides, antiinflammatory depsipeptides from a marine streptomycete. J. Org. Chem 64, 1145–1150.

Myronovskyi M, and Luzhetskyy A (2016). Native and engineered promoters in natural product discovery. Nat. Prod. Rep 33, 1006–1019. [PubMed: 27438486]

Newman DJ, and Cragg GM (2016). Natural products as sources of new drugs from 1981 to 2014. J. Nat. Prod 79, 629–661. [PubMed: 26852623]

Olano C, Lombó F, Méndez C, and Salas JA (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. Metab. Eng 10, 281–292. [PubMed: 18674632]

Omura S, et al. (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. U. S. A 98, 12215–12220. [PubMed: 11572948]

Osbourn A (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet. TIG 26, 449–457. [PubMed: 20739089]

Perego M, Glaser P, Minutello A, Strauch MA, Leopold K, and Fischer W (1995). Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis* identification of genes and regulation. J. Biol. Chem 270, 15598–15606. [PubMed: 7797557]

Pierson LS, and Pierson EA (2010). Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. Appl. Microbiol. Biotechnol 86, 1659–1670. [PubMed: 20352425]

Price-Whelan A, Dietrich LEP, and Newman DK (2006). Rethinking "secondary" metabolism: physiological roles for phenazine antibiotics. Nat. Chem. Biol 2, 71–78. [PubMed: 16421586]

Ramos I, Dietrich LEP, Price-Whelan A, and Newman DK (2010). Phenazines affect biofilm formation by *Pseudomonas aeruginosa* in similar ways at various scales. Res. Microbiol 161, 187–191. [PubMed: 20123017]

Ray L, Yamanaka K, and Moore BS (2016). A peptidyl-transesterifying type I thioesterase in salinamide biosynthesis. Angew. Chem. Int. Ed Engl 55, 364–367. [PubMed: 26553755]

Rowe CJ, Cortés J, Gaisser S, Staunton J, and Leadlay PF (1998). Construction of new vectors for high-level expression in actinomycetes. Gene 216, 215–223. [PubMed: 9714812]

Rui Z, Ye M, Wang S, Fujikawa K, Akerele B, Aung M, Floss HG, Zhang W, and Yu T-W (2012). Insights into a divergent phenazine biosynthetic pathway governed by a plasmid-born esmeraldin gene cluster. Chem. Biol 19, 1116–1125. [PubMed: 22999880]

Rutledge PJ, and Challis GL (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol 13, 509–523. [PubMed: 26119570]

Schoenafinger G, Schracke N, Linne U, and Marahiel MA (2006). Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. J. Am. Chem. Soc 128, 7406–7407. [PubMed: 16756271]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. [PubMed: 14597658]

Shao Z, Zhao H, and Zhao H (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. 37, e16. [PubMed: 19074487]

Siegl T, Tokovenko B, Myronovskyi M, and Luzhetskyy A (2013). Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab. Eng 19, 98–106. [PubMed: 23876413]

Sievers F, and Higgins DG (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol 1079, 105–116. [PubMed: 24170397]

Smith S, and Tsai S-C (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat. Prod. Rep 24, 1041–1072. [PubMed: 17898897]

Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, and Moore BS (2015). Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem. Biol 10, 2841–2849. [PubMed: 26458099]

Trischman JA, Tapiolas DM, Jensen PR, Dwight R, Fenical W, McKee TC, Ireland CM, Stout TJ, and Clardy J (1994). Salinamides A and B: Anti-inflammatory depsipeptides from a marine streptomycete. J. Am. Chem. Soc 116, 757–758.

Voggu L, Schlag S, Biswas R, Rosenstein R, Rausch C, and Götz F (2006). Microevolution of cytochrome *bd* oxidase in *Staphylococci* and its implication in resistance to respiratory toxins released by *Pseudomonas*. J. Bacteriol 188, 8079–8086. [PubMed: 17108291]

Wang M, et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol 34, 828–837. [PubMed: 27504778]

Weber T, et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243. [PubMed: 25948579]

Wilson MC, and Moore BS (2012). Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. Nat. Prod. Rep 29, 72–86. [PubMed: 22124767]

Wu Y, and Seyedsayamdost MR (2017). Synergy and target promiscuity drive structural divergence in bacterial alkylquinolone biosynthesis. Cell Chem. Biol 24, 1437–1444. [PubMed: 29033316]

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, and Moore BS (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl. Acad. Sci. U. S. A 111, 1957–1962. [PubMed: 24449899]

Yang Z, Jin X, Guaciaro M, and Molino BF (2012). Asymmetric synthesis and absolute configuration of streptophenazine G. J. Org. Chem 77, 3191–3196. [PubMed: 22432723]

Highlights

- Streptophenazine biosynthetic pathway (*spz*) was cloned from marine *Streptomyces*
- Refactoring the *spz* pathway led to tremendous chemical diversity (112 compounds)
- Streptophenazines with *N*-formylglycine moiety show potent antibiotic activity
- Discrete adenylation protein is required to produce *N*-formylglycine analogs

Significance

This work presents the discovery, engineering, and heterologous expression of the hybrid streptophenazine (*spz*) biosynthetic gene cluster from marine *Streptomyces* that unites phenazine, polyketide, and nonribosomal peptide biochemistry in an unprecedented expansion on the biosynthetic and chemical logic of the phenazine class of metabolites. We developed a set of promoter cassettes compatible with the widely used pCAP series of shuttle vectors to successfully link the cryptic *spz* BGC to its molecular products, including *N*-formylglycine analogs that possess superior antibiotic activity. We identified one gene, *spz*15, that encodes a discrete adenylation protein as a genetic search hook into analogous systems, and propose its involvement in the attachment of the rare *N*-formylglycine moiety. This discovery encourages future work to reveal the mechanism behind this biosynthetic transformation. Complete understanding of the biosynthesis of diverse streptophenazines opens an opportunity for engineering to improve the bioactivities of these compounds. Ultimately, using biology to inform chemistry in this way to reliably regulate expression of BGCs is fundamental in discovering, understanding, and governing the chemistry of our natural world.

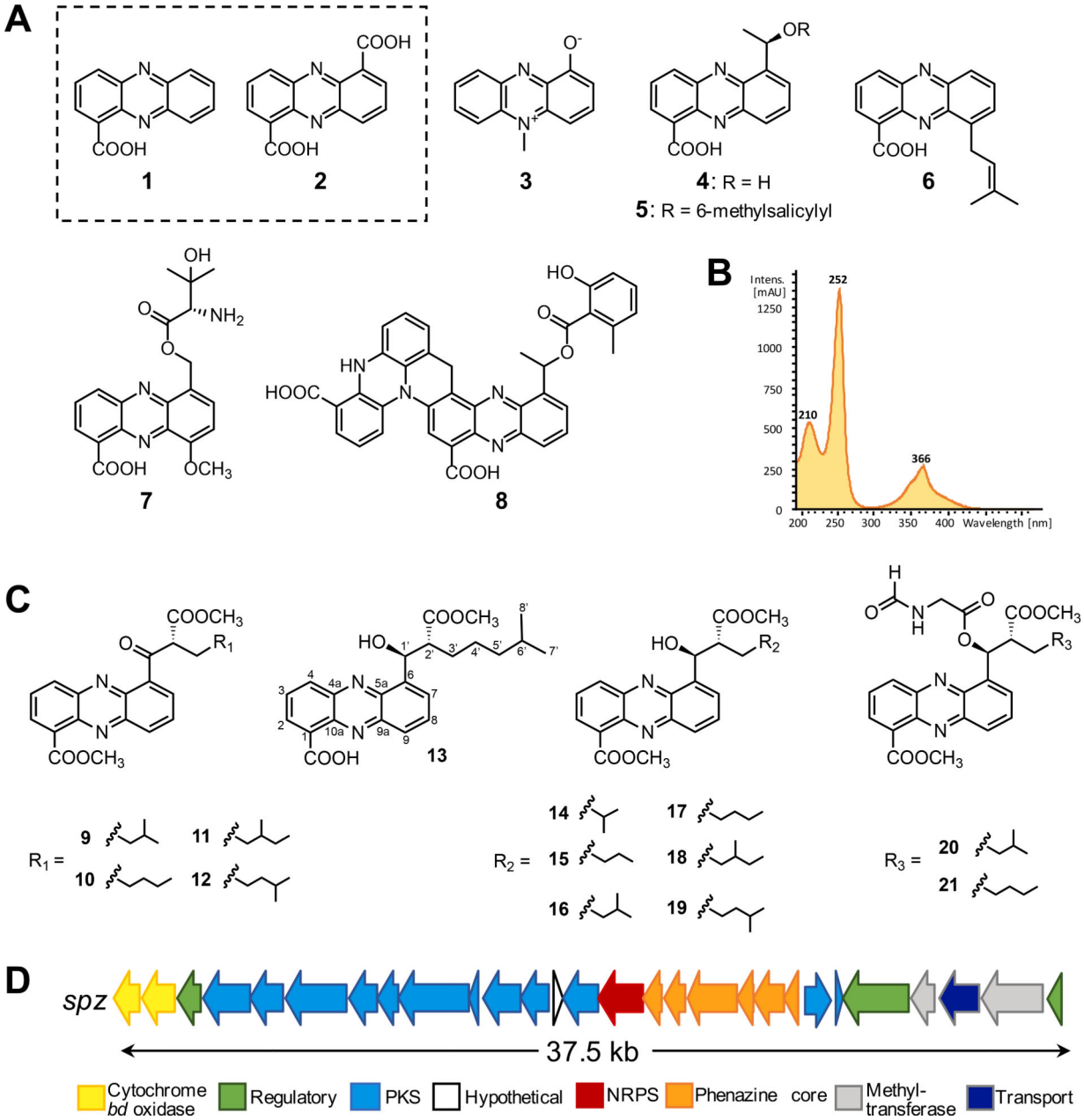


Figure 1. Chemical structures of selected phenazine metabolites and gene organization of the *spz* biosynthetic gene cluster from *Streptomyces* sp. CNB091.

A. Phenazine compounds isolated from *Streptomyces* and *Pseudomonas* bacteria: phenazine-1-carboxylic acid (PCA) (**1**), phenazine-1,6-dicarboxylic acid (PDC) (**2**), pyocyanin (**3**), saphenic acid (**4**), saphenamycin (**5**), endophenazine A (**6**), pelagiomicin A (**7**), esmeraldin B (**8**). **B.** Characteristic UV absorption spectrum of streptophenazines (λ_{max} 252 nm, 366 nm). **C.** Chemical structures of streptophenazine metabolites associated with the *spz* BGC from *S.* sp. CNB-091. Previously reported compounds include streptophenazines A (**16**), B (**17**), C (**13**), D (**14**), F (**19**), G (**18**). Compounds isolated in this work include oxo-streptophenazines A (**9**), B (**10**), G (**11**), F (**12**), and streptophenazines P

(**15**), Q (**20**), R (**21**). **D.** Gene arrangement of the *spz* BGC from *S.* sp. CNB-091. See also Figures S1, S2, S5.

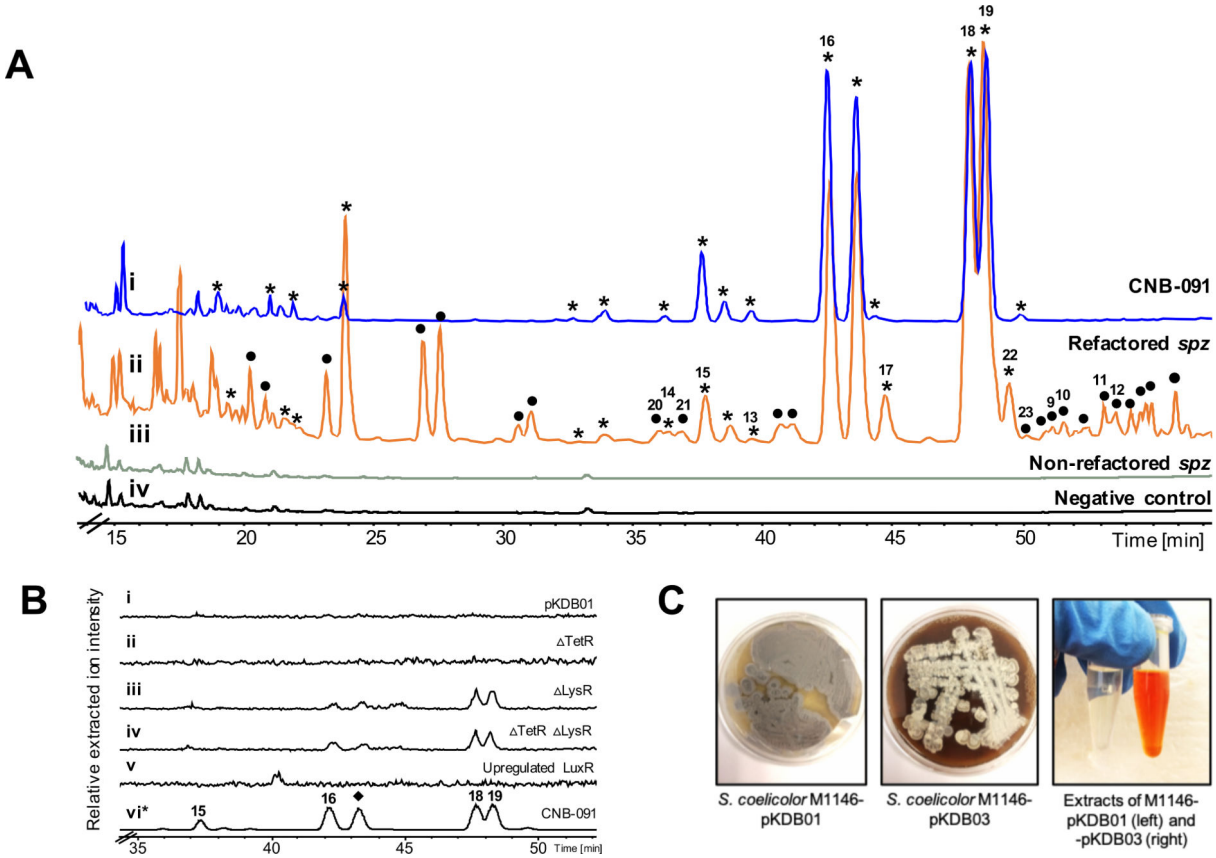


Figure 2. Analysis of the crude extracts from the engineered *spz* cluster in the heterologous host *S. coelicolor* M1146 and the wild type producer *S.* sp CNB-091.

A. HPLC chromatograms (254 nm) of ethyl acetate extracts of: **i** – CNB-091, * indicates peaks with streptophenazine UV profile; **ii** – M1146-pKDB03 (refactored *spz*), • indicates peaks with streptophenazine UV profile appearing in refactored strain; **iii** – M1146-pKDB01 (non-refactored *spz*); **iv** – M1146-pCAP03 (negative control). **B.** Extracted LCMS chromatograms (*m/z* 425.2, 439.2, corresponding to streptophenazines A, B, F, G) of the organic extracts from regulatory gene mutant strains: **i** – pKDB01 (non-refactored *spz*); **ii** – pKDB02 (non-refactored *spz*ΔTetR); **iii** – pKDB01ΔLysR, **iv** – pKDB02ΔLysR; **v** – pKDB01-*ermE**p-LuxR, **vi** – CNB-091 (*trace is minimized 100x for scale). ◆ indicates peak that was not isolated. **C.** Production of streptophenazines on solid media by *S. coelicolor* M1146-pKDB03 and corresponding ethyl acetate extracts of strains with non-refactored (pKDB01) and refactored (pKDB03) *spz* cluster. Bright orange color is due to the abundant production of streptophenazines. See also Figures S2 and S3.

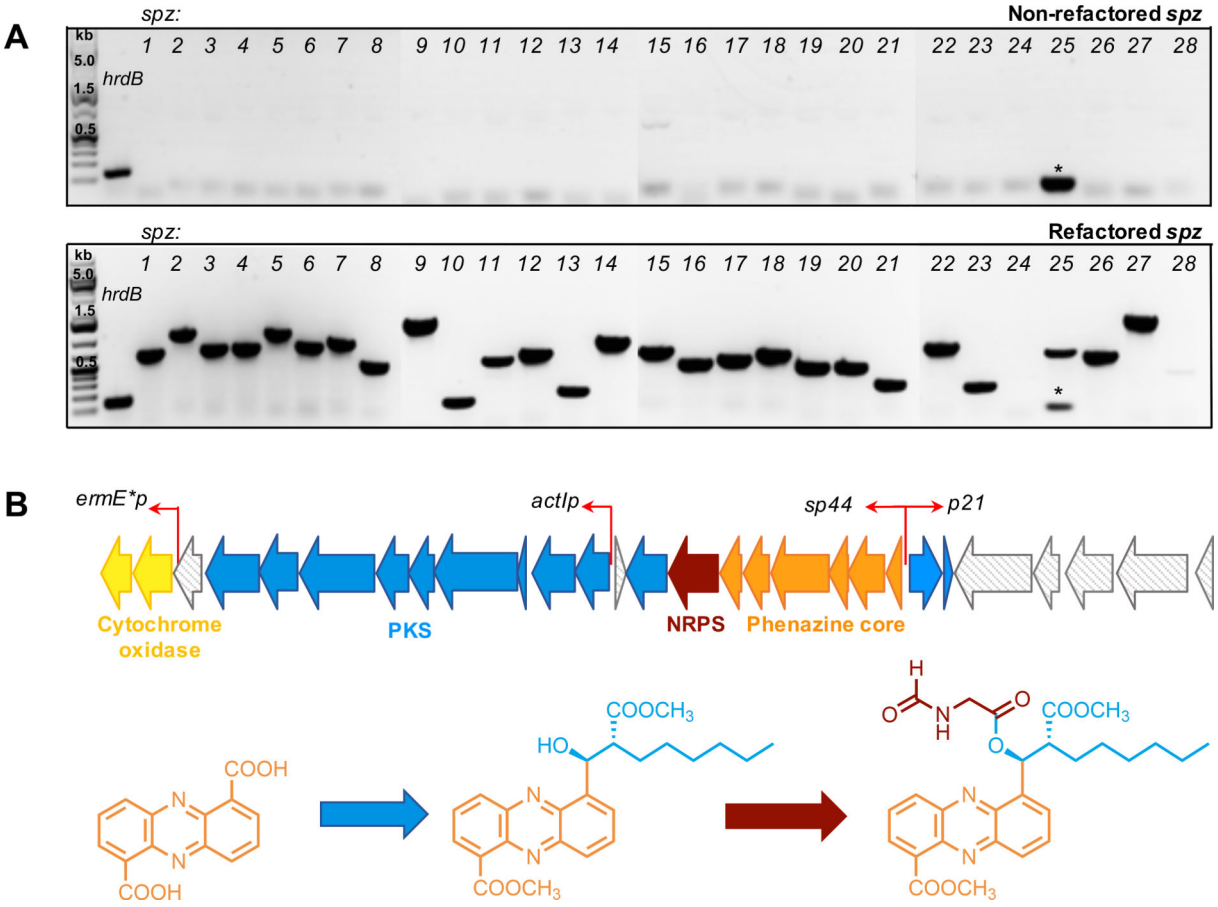


Figure 3. Refactoring the *spz* biosynthetic gene cluster.

A. RT-PCR analysis of expression of biosynthetic genes in the *spz* BGC before and after refactoring. Gene encoding a sigma-like transcription factor of *S. coelicolor, hrdB,* was used as a positive control. Strains *S. coelicolor* M1146-pKDB01 (non-refactored *spz*) and *S. coelicolor* M1146-pKDB03 (refactored *spz*) were used for analysis. Expected sizes (bp) of amplified fragments as follows: *hrdB* – 132, *spz* genes: *spz*1 – 679; 2 – 1135; 3 – 731; 4 – 750; 5 –1093; 6 – 800; 7 – 946; 8 – 544; 9 – 1134; 10 – 180; 11 – 587; 12 – 689; 13 – 268; 14 – 963; 15 – 768; 16 – 532; 17 – 611; 18 – 722; 19 – 515; 20 – 540; 21 – 328; 22 – 727; 23 – 242; 24 – 1339; 25 – 667; 26 – 623; 27 – 1499; 28 – 410. * indicates low molecular weight band corresponding to primer dimer. Expression of the gene *spz*24 (encodes LuxR-type regulator) was not observed for both versions of the *spz* cluster, refactored and non-refactored. **B.** Schematic illustration of the refactoring strategy indicating locations and directions of the promoters introduced (red arrows). See also Figure S2 and Table S1.

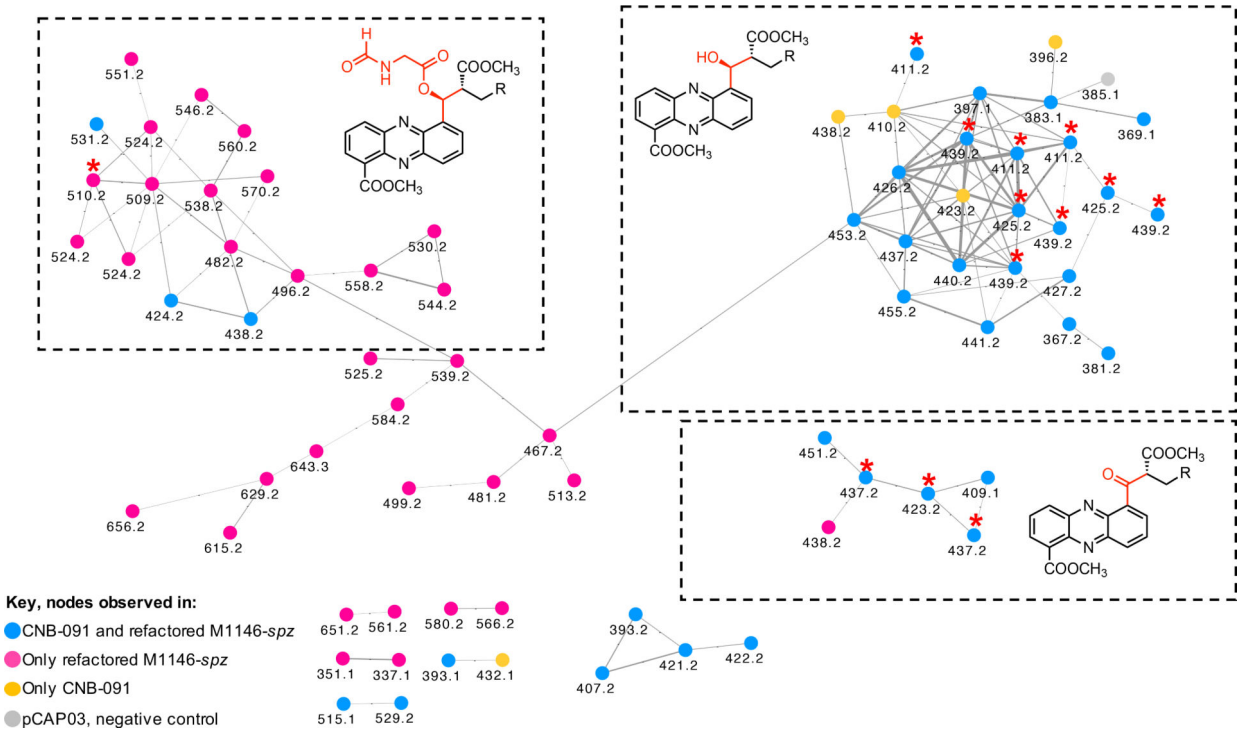


Figure 4. Streptophenazine cluster of nodes from the molecular network.

MS/MS data of extracts of CNB-091, M1146-pKDB01 (non-refactored *spz*), M1146-pKDB03 (refactored *spz*), and pCAP03 (negative control) are included and color coded as per key. All nodes are labeled with the corresponding precursor ion mass. Isolated compounds are marked with red stars. See also Figures S3 and S4 and Table S2.

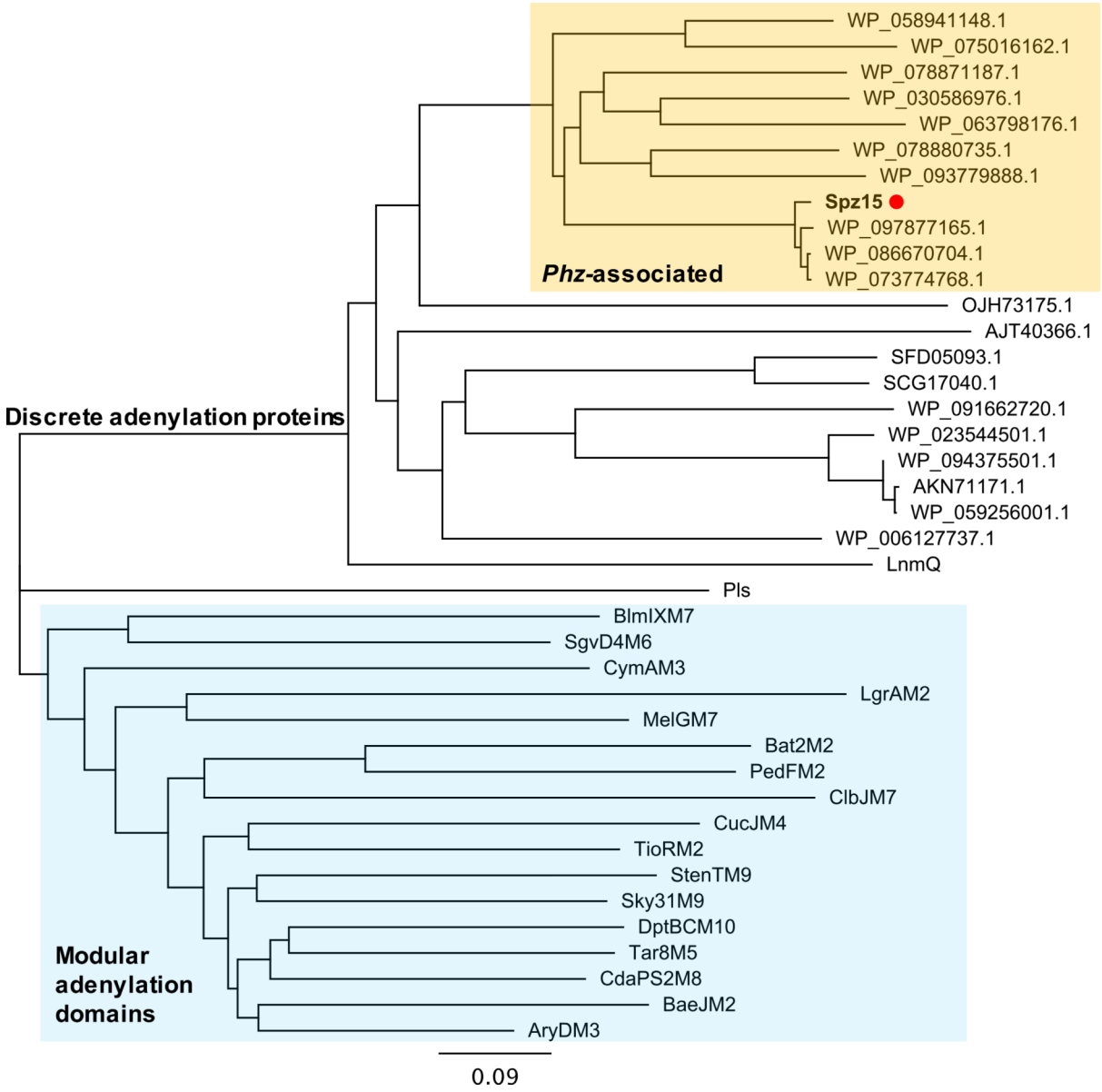


Figure 5. Phylogenetic analysis of Spz15 adenylation protein.

Phylogenetic tree of Spz15 with characterized adenylation domains and NCBI blastp homologs of Spz15 was constructed using Geneious software. Spz15 (marked with orange dot) forms a separate clade from canonical modular characterized adenylation domains (highlighted in blue). Proteins that clade with Spz15 (highlighted in orange) are found within putative phenazine type (*phz*) BGCs. The remaining protein sequences are discrete standalone adenylation enzymes. Protein abbreviations as follows: LnmQ – leinamycin, Pls – ε-poly-L-lysine, NRPS: BlmIXM7 – bleomycin module 7, SgvD4M6 – griseoviridin module 6, CymAM3 – cyclomarin module 3, LgrM2 – linear gramicidin module 2, MelGM7 – melithiazol module 7, Bat2M2 – batumin module 2, PedFM2 – pederin module 2, ClbJM7 – colibactin module 7, CucJM4 – cupriachelin module 4, TioRM2 – thiocoraline module 2, StenTM – stenothricin module 9, Sky31M9 – skyllamycin module 9, DptBCM10

– daptomycin module 10, Tar8M5 – taromycin module 5, CdaPS2M8 – calcium dependent antibiotic module 8, BaeJM2 – bacillaene module 2, AryDM3 – arylomycin module 3; all sequences were downloaded from NCBI protein database. See also Figures S6 and S7.

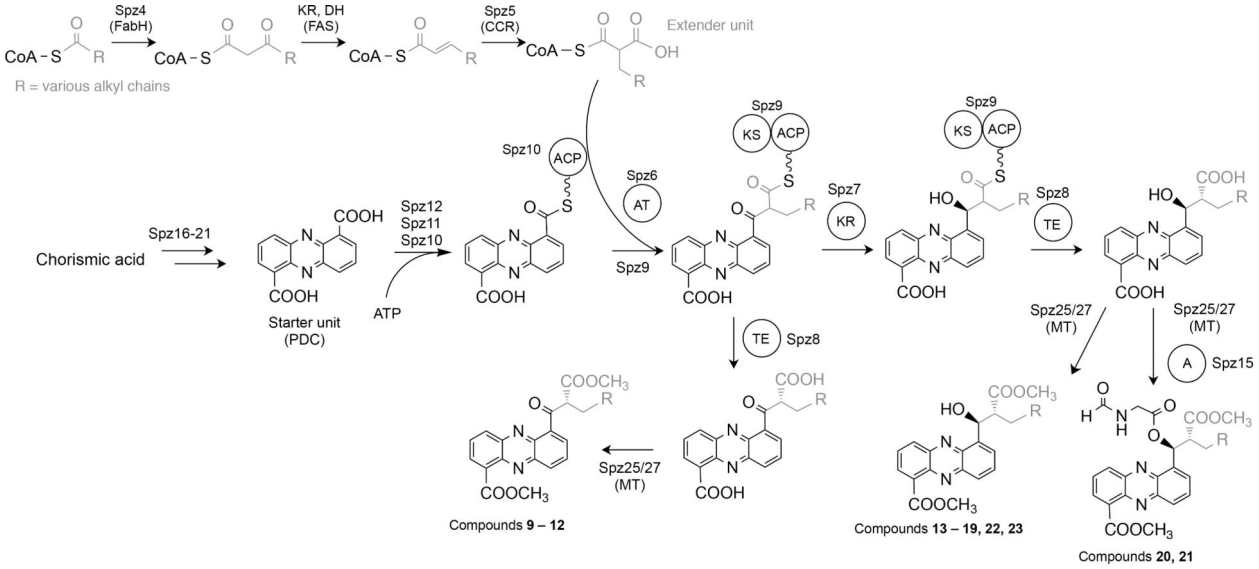


Figure 6. Proposed biosynthesis of streptophenazines from *Streptomyces* sp. CNB-091.

R denotes alkyl substituents for the side chain. A variety of polyketide extender units arise from diverse substituted malonyl-CoA derivatives from primary metabolism through catalysis by a crotonyl-CoA carboxylase/reductase (Spz5).

Table 1.

Open reading frames (ORFs) in the *spz* biosynthetic gene cluster and their proposed roles.

Proposed function	Protein	aa	BLAST closest homolog	%Cover/ Identity	Homolog accession number
Cytochrome *bd* ubiquinol oxidase II	Spz1	329	*Streptomyces* sp. ms184	100/98	WP_097875779.1
Cytochrome *bd* ubiquinol oxidase I	Spz2	439	*Streptomyces albovinaceus*	100/96	WP_086670717.1
LysR-type regulator	Spz3	293	*Streptomyces* spp.	100/97	WP_073774781.1
Acyl-CoA synthetase	Spz4	600	*Streptomyces albovinaceus*	100/98	WP_086670715.1
Crotonyl-CoA carboxylase/reductase	Spz5	416	*Streptomyces* spp.	100/99	WP_073774779.1
Type I PKS (KS*, AT)	Spz6	800	*Streptomyces albovinaceus*	100/97	WP_086670713.1
Ketoreductase (KR)	Spz7	351	*Streptomyces* spp.	100/99	WP_073774774.1
Thioesterase (TE)	Spz8	264	*Streptomyces* sp. ms184	100/97	WP_097876860.1
Type I PKS (KS, ACP)	Spz9	914	*Streptomyces* sp. ms184	100/98	WP_097876858.1
Acyl carrier protein (ACP)	Spz10	109	*Kitasatospora purpeofusca*	95/61	WP_078880679.1
Aldehyde dehydrogenase	Spz11	481	*Streptomyces albovinaceus*	100/97	WP_086670706.1
PDC adenylase	Spz12	365	*Streptomyces* sp. ms184	100/98	WP_097876855.1
Hypothetical protein	Spz13	113	*Streptomyces* sp. TSRI0445	100/96	OKI73654.1
FAD-binding oxidoreductase	Spz14	455	*Streptomyces* spp.	100/98	WP_073776179.1
D-alanine-poly(phosphoribitol) ligase	Spz15	545	*Streptomyces* sp. ms184	100/98	WP_097877165.1
Pyridoxamine 5'-phosphate oxidase (PhzG)	Spz16	220	*Streptomyces* sp. ms184	100/99	WP_097877165.1
2,3-Dihydro-3-hydroxylanthranilate isomerase (PhzF)	Spz17	283	*Streptomyces* spp.	100/98	WP_073774764.1
Anthranilate synthase (PhzE)	Spz18	627	*Streptomyces* sp. ms184	99/98	WP_097877170.1
Isochorismatase (PhzD)	Spz19	207	*Streptomyces* sp. ms184	100/99	WP_097877172.1
Phospho-2-dehydro-3-deoxyheptonate aldolase (DAHP synthetase) (PhzC)	Spz20	391	*Streptomyces* sp. ms184	100/98	WP_097877172.1
Phenazine biosynthesis protein (PhzA/B)	Spz21	166	*Streptomyces* sp. ms184	100/98	WP_097877172.1
3-Oxoacyl-ACP synthase III (FabH)	Spz22	344	*Streptomyces* sp. ms184	100/99	WP_097877172.1
Aryl carrier domain protein	Spz23	97	*Kitasatospora purpeofusca*	92/61	WP_030398483.1
LuxR-type regulator	Spz24	846	*Kitasatospora purpeofusca*	96/55	WP_030398483.1
Class I SAM-dependent methyltransferase	Spz25	289	*Streptomyces* sp. TSRI0445	96/99	WP_079198635.1
MFS transporter	Spz26	516	*Streptomyces* sp. ms184	100/97	WP_097874479.1
HemK family methylase	Spz27	781	*Streptomyces kanasensis*	100/70	WP_058941165.1
TetR/AcrR regulator	Spz28	207	*Streptomyces griseus*	100/98	WP_030809832.1

* lacking catalytic cysteine and histidine residues