Bioinformatic and functional evaluation of actinobacterial piperazate metabolism

Yifei Hu[#1,2], Yunci Qi[#1], Spencer D Stumpf[1], John M D'Alessandro[1,3], Joshua A.V. Blodgett[1]*

[1]Department of Biology, Washington University in St Louis, 1 Brookings Dr. St Louis, Missouri 63130, USA

[#]These authors contributed equally to this work

*Corresponding author: jblodgett@wustl.edu

Current addresses:

[2]Pritzker School of Medicine, The University of Chicago, Chicago, Illinois, 60637, USA

[3]Department of Bioengineering, University of Missouri, Columbia, Missouri, 65203, USA

ABSTRACT

Piperazate (Piz) is a nonproteinogenic amino acid noted for its unusual N-N bond motif. Piz is a proline mimic that imparts conformational rigidity to peptides. Consequently, piperazyl molecules are often bioactive and desirable for therapeutic exploration. The *in vitro* characterization of *Kutzneria* enzymes KtzI and KtzT recently led to a biosynthetic pathway for Piz. However, Piz anabolism *in vivo* remained completely uncharacterized. Herein, we describe the systematic interrogation of actinobacterial Piz metabolism using a combination of bioinformatics, genetics, and select biochemistry. Following studies in *Streptomyces flaveolus*, *Streptomyces lividans*, and several environmental *Streptomyces* isolates, our data suggest that KtzI-type enzymes are conditionally dispensable for Piz production. We also demonstrate the feasibility of Piz monomer production using engineered actinobacteria for the first time. Finally, we show that some actinobacteria employ fused KtzI-KtzT chimeric enzymes to produce Piz. Our findings have implications for future piperazyl drug discovery, pathway engineering, and fine chemical bioproduction.

INTRODUCTION

New drugs to treat antibiotic-resistant infections are urgently needed[1]. Piperazate (Piz) is a non-proteinogenic amino acid and a largely overlooked pharmacophore, Originally discovered as a building block of the antibacterial monamycins in 1971[2], several piperazyl antibiotics have since been discovered from filamentous actinomycete

bacteria. These include the matlystatins (peptide deformylase inhibitors)[3, 4], the luzopeptins (antitumor agents)[5], the kutznerides (antifungals)[6], and the sanglifehrins (antivirals and immunosuppressants)[7, 8] (Figure 1A). Piz imparts intramolecular rigidity and prolyl mimicry[9] to parent molecules, making it particularly desirable in the context of peptidyl turn analogs and other peptidomimetics[10, 11].

Until recently, the biosynthesis of Piz and its N-N bond remained enigmatic. Following the biochemical characterization of two *Kutzneria* sp. strain 744 enzymes, KtzI[12] and KtzT[13], a two-step pathway converting L-ornithine (L-Orn) to L-Piz (Figure 1B) was revealed. Specifically, KtzI first converts L-Orn to its *N*-hydroxyl derivative N^5-OH-L-Orn, which is then cyclized to L-Piz by the hemoenzyme KtzT. An important advance in N-N bond biocatalysis, this model also suggested a new path for genomics-guided piperazyl molecule discovery.

However, much remains unknown about Piz biosynthesis. Importantly, its production has never been investigated *in vivo*. This limits the bioengineering potential of this important drug building block. Further, a family of proteins annotated as PaiB-type regulators[13] share significant homology with KtzT. But the functional relationship of PaiB proteins and KtzT-type enzymes, and how this impacts Piz production, remained unclear. Understanding these relationships are key for revealing the extent of Piz metabolism in nature and for creating refined genome-mining methods to discover new piperazyl molecules. To further understand Piz biosynthesis, we utilized a combination of bioinformatics, genetics, and biochemistry to investigate production in several *Streptomyces* species.

Several lines of evidence were acquired that suggest *in vivo* Piz production is more nuanced than initially expected. Contrary to the KtzI/KtzT model, we found that the KtzI ortholog SfaB is dispensable during sanglifehrin production in *Streptomyces flaveolus* and that expressing the strain's KtzT ortholog, SfaC, is sufficient for Piz monomer production in *Streptomyces lividans*. These observations strongly suggest piperazyl-molecule biosyntheses likely benefit from N^5-OH-L-Orn production encoded outside of their cognate genomic loci; a trend supported by heterologous Piz production tests in several environmental *Streptomyces* stains. Following extensive phylogenetic analysis of KtzT/SfaC enzymes, we also discovered certain actinomycetes carry out Piz biosynthesis by chimeric enzymes featuring fused KtzI- and KtzT-type domains. Two chimeras were tested, revealing one from *Micromonospora tulbaghiae* with significant activity. Finally, our bioinformatic efforts also revealed specific residue patterns that

differentiate PaiB and KtzT family proteins, and these patterns were experimentally validated using SfaC point mutants. Together, our data expand upon the current biosynthetic model and reveal new opportunities for piperazate pathway recognition and bioengineering.

RESULTS AND DISCUSSION

Analysis of Piz metabolism during sanglifehrin production.

The current Piz biosynthetic model is based on enzymes sourced from the kuzneride biosynthetic locus of *Kutzneria* sp. strain 744. However, this strain is challenging to transform, leading us to initiate our Piz metabolic studies in *Streptomyces flaveolus* DSM 9954. The characterized producer of the sanglifehrins, a family of therapeutically interesting piperazyl antivirals, *S. flaveolus* is readily transformed via intergeneric conjugation[14].

The sanglifehrin biosynthetic locus was sequenced previously[14], and the cluster encodes homologs of *ktzI* and *ktzT* (*sfaB* and *sfaC,* respectively). Based on the KtzI/KtzT-based Piz pathway, we predicted deleting either gene should abrogate sanglifehrin production via Piz precursor loss. *S. flaveolus* is known to produce >20 sanglifehrin congeners[14], with therapeutically-relevant sanglifehrin A (*m/z* 1090.7) being among the most highly produced[15]. LC/MS/MS detected sanglifehrin A production in both wild-type (JV270) and *rpsL* (JV571) strain extracts (Figures 2A, S1). In contrast, our unmarked Δ*sfaC* (JV575) deletion mutant was deficient for sanglifehrin A production in addition to several other peaks that likely correspond to other sangifehrin-family compounds. Suggesting *sfaC* is indeed involved in Piz production, we found that either feeding exogenous L-Piz (0.25 mM) or ectopic *sfaC* expression (JV577) both restored sanglifehrin production in the mutant background. Together these findings are consistent with the current Piz biosynthetic model and confirm the *ktzT* homolog *sfaC* enzyme is essential for *S. flaveolus* Piz production.

In contrast with *sfaC*, our *sfaB* (JV574) deletion mutant retains sanglifehrin production (Figure 2A). This was unexpected because *sfaB* is a *ktzI* homolog, and thus should be required for the Piz intermediate N^5-OH-L-Orn. Based on this observation, we surmised the *S. flaveolus* genome must encode additional functionally redundant *N*-hydroxylases. Lacking a sequenced genome to verify this, we designed degenerate PCR primers to detect such genes using several published Orn *N*-OHase sequences (see

Supporting Information). This revealed the presence of least one additional flavin-dependent *N*-OHase encoded within the genome of the *S. flaveolus* Δ*sfaB* strain.

Heterologous L-Piz production in multiple *Streptomyces* spp.

To further investigate Piz anabolism via SfaB/SfaC, we carried out heterologous expression experiments in *Streptomyces lividans* strain 66. This commonly-used heterologous host lacks apparent *ktzT*/*sfaC* homologs in its genome and is not known to produce piperazyl compounds. *S. lividans* was transformed with integrative plasmids expressing either *sfaB* (pYH014), *sfaC* (pYH015), or a combination of both plasmids. Resulting Piz production was monitored by LC-MS. As expected, wild-type *S. lividans* and the *sfaB* transgenic strain (JV593) lacked detectable Piz. In contrast, *S. lividans* expressing both *sfaB* and *sfaC* (JV596) or *sfaC* alone (JV594) produced Piz (Figure 2B). Finding that *S. lividans* can produce Piz via *sfaC* in the absence of *sfaB* echoed our *S. flaveolus* observations. While N^5-OH-L-Orn metabolism in *S. lividans* is poorly documented, the strain's genome[16] harbors multiple *ktzI*/*sfaB* homologs that may support its production (e.g. EFD65609.1, EFD71462.1, the latter encoded within a predicted hydroxamate-type siderophore biosynthetic locus).

The above experiments revealed that Piz, a desirable drug building block, can be produced by engineered actinobacteria for the first time. They further suggested that N^5-OH-L-Orn metabolism, which remains poorly explored in the literature, may be fairly active within most Streptomycetes. To further understand this, we assayed several environmental *Streptomyces* isolates for the ability to produce Piz following transformation with pYH015, a *sfaC* expression plasmid (Table S1). Of 22 obtained transformants, 12 produced plasmid-dependent Piz (Figure 2C). A single strain, SP18CS05, also produced Piz in the absence of the plasmid. Further investigation by degenerate PCR with primers designed to amplify *sfaC/ktzT* homologs confirmed that this organism harbors a native copy. Thus, our summed observations reveal N^5-OH-L-Orn metabolism is indeed a fairly common and potentially exploitable feature of many *Streptomyces*.

Bioinformatic analysis of KtzT/SfaC type enzymes and PaiB homologs

Prior to the biochemical characterization of KtzT, KtzT/SfaC homologs encoded within sequenced piperazyl-molecule biosynthetic loci were often mis-annotated as PaiB-family transcriptional regulators[13]. PaiB- and KtzT/SfaC- type proteins show significant amino acid similarity and likely share similar protein structures[13] (Figure S8). Following KtzT's recent functional assignment as a hemoenzyme, it remained unclear

whether these protein families have identical or distinct catalytic functions. Indeed, the recent revelation that *Geobacillus stearothermophilus* PaiB also binds indicates PaiB-family proteins could also function outside of transcriptional control[13].

To begin addressing if PaiB and KtzT/SfaC proteins are functionally distinct, or are in fact identical, we first assembled a collection of high-confidence SfaC/KtzT ortholog sequences. This was done by screening genomes for co-localized *ktzI/sfaB* and *ktzT/sfaC* homologs, and then scanning nearby gene neighborhoods for non-ribosomal peptide synthase (NRPS) genes having Piz adenylation domains. (See Methods; for an NRPS review, see[17]). Because most resulting hits lacked functional locus names, we annotated putative KtzI/SfaB and KtzT/SfaC orthologs as PzbA's and PzbB's (for **p**ipera**z**ate **b**iosynthesis A & B), respectively. This effort revealed >80 predicted Piz biosynthetic loci (Table S4). To date, all known natural piperazyl molecules are produced by filamentous actinomycetes. Perhaps unsurprisingly, the vast majority of the identified putative piperazyl loci were discovered within actinobacterial genomes., However, our PzbA/B co-localization and PzbB sequence analyses suggest Piz metabolism likely extends to the proteobacteria as well, identifying certain *Collimonas fungivorans* and *Photorhabdus luminescens* isolates as candidate piperazyl-molecule producers (Table S4, Figure 3).

To identify and analyze PaiB-type proteins, we assembled a representative set of homologs mined from public databases (see methods) that were not associated with *pzbA* and/or NRPS genes. Whereas the vast majority of candidate *pzbB* genes were found in actinobacteria (all except 2; Figure 3), we found putative *paiB* genes to be more broadly distributed among a diversity of microorganisms including firmicutes, proteobacteria, as well as some fungi Notably, a maximum-likelihood phylogeny clearly separates the manually- curated PzbB sequences from the other PaiB-type homologs indicating PzbB's are distinct and suggest the potential for different biological roles between the two protein clades (Figure 3).

Sequence diversity among PzbB proteins

After determining that that PaiB and PzbB proteins are clearly sequence-distinct, we further noted significant sequence variation within the PzbB group itself. For example, SfaC is 89.6% identical to PzbB from *Streptomyces* sp. strain PBH53, but only 33.6% identical to PzbB from *Actinoalloteichus cyanogriseus*. With PzbBs ranging from only 200-252 AA in length, this raised the question if all members of the PzbB clade are indeed capable of Piz catalysis.

To test if sequence diversity within the PzbB group affects Piz catalysis, we expressed a panel of diverse PzbB homologs (Figure 3) cloned from several actinomycetes (*Lentzea flaviverrucosa*, *Streptomyces himastatinicus*, *Kutzneria* sp. strain 744, and *Actinoalloteichus cyanogriseus*) in *S. flaveolus* and *S. lividans.* Although all are weakly conserved with SfaC (33.6 - 40.8% amino acid identity), we found each tested PzbB is functional in Piz biogenesis in both heterologous hosts (Figures 4A and B). In contrast, PaiB cloned from *Bacillus subtilis* strain 168 (24.5% amino acid identity with SfaC) failed to rescue the Δ*sfaC* mutation (not shown). Along with the phylogenetic analysis above, these data further support the PzbB group as a functionally conservative clade, and further suggest PaiB proteins are functionally distinct (analyzed further below).

Discovery and characterization of chimeric *pzbAB* genes and enzymes

Our data-mining efforts also led to the discovery of a subgroup of piperazyl molecule biosynthetic loci that harbor apparent PzbAB chimeric proteins. These enzymes are characterized by a linkerless amino-terminal PzbA domain fused to a carboxyl PzbB terminus (Figures 4C, S8). PzbB domains extracted from these chimeras form a distinct clade within the PzbB tree and have greater sequence similarity with PaiBs than their stand-alone counterparts (Figure 3). By comparing the sequences of 83 stand-alone PzbA and PzbB pairs against 11 available chimeric PzbAB's, we found that the fusion proteins are distinguished by unique position-specific amino acid utilization patterns compared to stand-alone homologs (Figure 4C).

To test if these chimeras catalyze Piz production, we expressed *pzbAB* genes cloned from *Amycolatopsis alba* and *Micromonospora tulbaghiae* in *S. flaveolus* Δ*sfaC*. Both chimeras restored sanglifehrin, but to a lesser extent than stand-alone PzbBs tested in Figure 4A. The *A. alba* chimera barely complemented the mutant, while the *M. tulbaghiae* gene restored ~75% of WT activity. (Figure 4A). To test if both of the PzbA and PzbB of the fused proteins domains are functional (as the chimeric PzbB domains might have sourced N^5-OH-L-Orn from background *S. flaveolus* metabolism) we expressed truncations of both *pzbAB* chimeras in *S. flaveolus* Δ*sfaC*. The truncated gene encompassing the *M. tulbaghiae* PzbB domain (AA's 453-672) also rescued sanglifehrin production, retaining ~40% of the activity of the full-length pzbAB gene. Similar to the full-length *pzbAB* from *A. alba,* the cloned *pzbB* (AA's 456-661) domain showed little activity. Why the chimeras of *A. alba* and *M. tulbaghiae* show such distinct activity differences in our heterologous assay remains unclear. Despite this, our data importantly demonstrate that PzbAB from *M. tulbaghiae* is catalytically competent and its

PzbB domain supports Piz catalysis. Other newly- identified PzbAB proteins documented in this work (Figure 3, Table S8) may be may be similarly active.

Following the above *in vivo* assays, we employed *in vitro* assays to confirm that purified PzbAB from *M. tulbaghiae* catalyzes Piz catalysis from both N^5-OH-L-Orn and L-Orn (Figures 5A and 5B). For these experiments, we employed SfaC or SfaB/SfaC coupled reactions as controls. SfaB, SfaC, and PzbAB were purified by Ni^{+2} -affinity chromatography after expression in *E. coli* (Supporting Methods, Figures S2, S6, S7). From prior KtzT observations, we expected both SfaC and PzbAB should be hemoproteins. We confirmed this by direct prosthetic group analysis (Figures S2, S7) and catalytic rate calculations (below). Once heme B binding was established for both proteins, the transformation of synthetic N^5-OH-L-Orn to L-Piz from heme B-replete SfaC and *M. tulbaghiae* PzbAB was monitored by LC-MS/MS, HRMS and ^{1}H NMR (Figures 5a, S3, S4, S5). We confirmed both proteins produced L-Piz under these conditions. While this activity was expected from SfaC (a KtzT ortholog), these assays conclusively demonstrate exogenous N^5-OH-L-Orn can enter the PzbB active site of the intact *M. tulbaghiae* PzbAB chimera.

Building on these results and prior coupled KtzI-KtzT studies[13], we predicted a single-pot reaction containing either SfaB and SfaC or PzbAB should generate L-Piz from L-Orn in the presence of O_2, NADPH, FAD, and b-heme. For the SfaB-SfaC coupled assay, purified SfaB was first tested for activity as described for KtzI. NADPH consumption in the presence of L-Orn, FAD and O_2 indicated a k_{cat} of ~7.7 min^{-1} (Figure S6), comparable to KtzI (k_{cat} = 7.0 min^{-1}). Subsequent assays testing SfaB + SfaC or *M. tulbaghaiae* PzbAB revealed Piz production from L-Orn (Figure 5B), confirming the activity of both *M. tulbaghiae* enzyme domains. Taken together, our *in vivo* and *in vitro* experiments reveal *M. tulbaghiae* PzbAB as a versatile catalyst capable of transforming both L-Orn and N^5-OH-L-Orn to L-Piz. This work also highlights PzbAB proteins for future substrate channeling, domain-domain interaction, and evolutionary covariation investigations.,

Identifying hallmark PzbB sequence signatures.

Finally, following our *in vivo* and *in vitro* explorations of several *pzbB* and *pzbAB* genes, we initiated combined bioinformatic and mutational studies to differentiate piperazate-linked *pzbB* genes from apparent *paiB'*s. To do this, 85 PzbB and 51 PaiB protein sequences were aligned, revealing a highly conserved sequence motif shared in both protein groups (KLSQ, corresponding to SfaC residues 183-186, Figures S8, 5C

and 5D). Scrutinizing the amino acid differences surrounding this common motif revealed two residue position-identity differences between PzbBs and PaiBs. At position 181, PzbBs typically have methionine, while PaiBs have lysine. At position 187, PzbB homologs have aspartate or glutamate, while PaiB homologs typically have asparagine.

We then tested the effects of substituting PaiB-type residues at these positions, into SfaC. Thus, SfaC M181K and E187N mutant proteins were purified and assayed for Piz production. Assuming Michaelis- Menten kinetics, WT heme-replete SfaC progress curves indicate an apparent K_m of 360 μM (Figure S9). In assays supplemented with N^5-OH-L-Orn substrate at ~10X K_m, WT SfaC (+hemin) k_{cat} was found to be 8000 ± 500 min^{-1} (-hemin 88 ± 2 min^{-1}). In contrast, the SfaC M181K (+hemin 99±8 min^{-1}; -hemin 0.41±0.01 min^{-1}) and E187N (+hemin 110±10 min^{-1}; -hemin 0.89±0.02 min^{-1}) mutants had turnover numbers multiple orders of magnitude lower than WT.

Both mutant proteins purified with absorbance maxima at 411 nm, suggesting lowered activity is independent of heme loss. Attempts to directly measure mutant K_m values were hindered by visible precipitation caused by elevated N^5-OH-L-Orn. Despite this, the observed reduction of L-Piz catalysis by the mutant enzymes clearly demonstrates a negative correlation with PaiB position-specific residues. Based on these observations and our phylogenetic analyses above, our work contributes further evidence that PaiB proteins likely function outside of piperazate metabolism. These findings are potentially important for accurate recognition of piperazyl-molecule biosynthetic loci during genome mining efforts and future mechanistic inquiries.

Methods

Reagents, media, and enzymes. All chemicals and media components were purchased from Sigma-Aldrich, Fisher Scientific, or Santa Cruz Biotechnology unless otherwise noted. N^5-OH-L-Ornithine hydrochloride (>90%) was synthesized by and purchased from AKos Consulting & Solutions GmbH. L-Piperazic acid dihydrochloride (>95%) was synthesized by and purchased from WuXi AppTec. DL-Piz racemate was purchased from BOC Sciences. Restriction endonucleases, T4 DNA ligase, and Taq polymerase were purchased from New England BioLabs. DNA purification kits were purchased from Qiagen. KOD Hot Start DNA Polymerase (EMD Millipore) and FailSafe PCR 2X

PreMixes (Epicentre) were used for PCR from *Streptomyces* genomic DNA. Oligonucleotides (Table S2) were purchased from Integrated DNA Technologies (IDT).

Instrumentation. Unless otherwise specified, LC-MS/MS analysis was performed on an Agilent 1260 Infinity HPLC connected to an Agilent 6420 Triple-Quadrupole mass spectrometer with electrospray ionization (ESI) source. Resulting data were analyzed offline with Agilent MassHunter software. UV Spectroscopy was performed using a Shimadzu UV-1800 spectrophotometer.

Bacterial growth and actinomycete isolations. All strains used in this work are listed in Table S1 and plasmids are listed in Table S3. Actinomycetes were routinely propagated on ISP2 (Difco) agar plates or in Trypticase Soy Broth (TSB, Difco) at 28˚C. *E. coli* strains were grown on standard Luria-Bertani (LB) plates or in equivalent liquid media. Sterile glass beads (6 mm) were added to actinomycete liquid cultures to disrupt mycelial clumps. PCR templates for verifying transgenic or mutant actinomycetes were prepared by grinding cells in 100 µL DMSO, similar to Van Dessel et al[18]. Several *Streptomyces* strains were isolated using standard enrichment methods[19] from soils collected at Tyson Valley Research Center in Eureka, Missouri. Isolates were confirmed as *Streptomyces* by 16S rDNA amplicon sequencing as described elsewhere[20] (Tables S1 and S6). Prior to conjugation with pYH015, each isolate was screened for native *pzbB*s. This was done using degenerate primers designed using BLOCKMAKER[21] and CODEHOP[22] software (Table S2).

Actinomycete conjugations. *Streptomyces* conjugations were performed essentially as described previously[23]. *S. flaveolus* and *S. lividans* were grown on ISP2 agar (Difco) and ISP-S agar (Difco malt extract, 1.5%; Difco soluble starch, 0.5%; Difco yeast extract, 0.5%; $CaCO_3$, 0.3%; Bacto agar, 2%; pH 7.2~7.5), respectively, for sporulation. Spores were harvested using TX Buffer[24]. *E. coli* strain JV36 was used as the conjugal donor. Exconjugants were selected with 50 µg/mL apramycin. After conjugation, JV36 was selected against with either 15 µg/mL polymyxin B or 30 µg/mL colistin. For plasmids using ΦC31 integration, multiplex PCR was used to verify correct *attB* insertion as previously described[23].

***S. flaveolus* gene deletions.** Genes of interest were removed by double-homologous recombination essentially as previously described using pJVD52.1[25]. Streptomycin-

resistant (Str^R) mutants necessary for counterselection[26] were isolated on ISP2 agar with 100 µg/mL of streptomycin. *S. flaveolus rpsL*(K43R) was chosen for further sanglifehrin analysis because its production parallels wild-type. Deletion constructs were transformed into *S. flaveolus rpsL*(K43R) and exconjugants were selected via apramycin resistance. Double recombinants were identified via counterselection after 3 days of growth at 37°C in TSB. Gene deletions were confirmed by PCR.

***S. flaveolus* sanglifehrin detection.** *S. flaveolus* spores were used to inoculate 15 mL of TSB in a 125 mL Erlenmeyer flask at 28˚C with vigorous shaking. Glass beads were added to disrupt mycelial clumps. After overnight growth, the culture was spread for confluence to sanglifehrin production medium[14] with 2% agar and incubated at 28˚C for 3 days. The agar with adherent cells was subsequently diced and immersed in ethyl acetate overnight. The resulting extract was evaporated at low pressure and resuspended in 500 µL of HPLC-grade methanol. Analysis was performed using a Phenomenex Luna C18 column (75 x 3 mm, 3 µm) using the following method: *T* = 0, 10% B; *T* = 5, 10% B; *T* = 25, 100% B; *T* = 27, 100% B, *T* = 29, 10% B, *T* = 30, 10% B; A: water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid; 0.6 mL/min. Authentic sanglifehrin A was used to develop the mass transition *m/z* 1090.7 -> 294.2 (fragmentor = 135 V and collision energy = 40 V) for single-reaction monitoring using MassHunter Optimizer (Agilent). For quantification, a standard curve was created with using concentrations ranging from 100 ng/mL to 1 mg/mL.

Heterologous Piz production. *Streptomyces* strains expressing SfaC were plated on YEME[27]-NSG (Yeast Extract-Malt Extract-No Sucrose or Glycine: Difco yeast extract, 0.3%; Difco Bacto-peptone, 0.5%; Difco malt extract, 0.3%; dextrose, 1.0%; Bacto agar, 2%; after autoclaving, $MgCl_2$ (2.5 M), 0.2%) supplemented with 0.5 mg/mL L-Orn and incubated at 28˚C for three days. The agar was subsequently diced and immersed in methanol overnight. The methanol fraction was evaporated to dryness and resuspended in 500 µL of LC-MS grade water plus 500 µL 6% sulfosalicylic acid. Samples were clarified by centrifugation (21000 x *g*, 5 min.) and 0.2 µm filtration prior to LC/MS/MS.

Piz detection via LC-MS and quantification via LC-MS/MS. Analysis was performed using an Imtakt Intrada Amino Acid column (50 x 3 mm, 3 µm) using the following

method: *T* = 0, 86% B; *T* = 3, 86% B; *T* = 10, 0% B; *T* = 11, 0% B; *T* = 12, 86% B; *T* = 14, 86% B; where A= water + 100 mM ammonium formate and B= MeCN + 0.1% formic acid; 0.6 mL/min, 35˚C column. The flow was routed to waste between *T* = 7.6 and *T* = 7.9 to avoid buffer salts. L-Piz was detected at *T* = 5.8, *m/z* 131.1 $[M+H]^+$. For Piz quantification, an SRM transition (*m/z* 131.1 => 56.3; source voltage, 86 V; collision energy, 37 V) was monitored, and a standard curve (second order polynomial, R^2 = 0.9996) was generated between 0.1 µM and 100 µM L-Piz-2HCl standard.

PzbB/PaiB sequence mining and phylogeny. Potential PzbB orthologs were identified in the JGI database using BLASTp with SfaC as the query. The following criteria were used to identify potential PzbB orthologs: the presence of a PzbA homolog and/or an NRPS having predicted Piz-adenylation domain signatures within ~15 kb. Hits having low identity scores (<25%) or E-values (>0.01) were excluded and likely Piz loading modules were identified by PRISM[28] or AntiSmash[29]. Identities of PzbBs used to create the PzbB phylogeny in Figure 3 are found in Table S4. BLASTp 2.6.0+ was used to calculate e-values, coverage, and protein identity. PaiB homologs were found via InterPro, using the term “Transcriptional regulator PAI 2-type”. PaiB from *Geobacillus stearothermophilus*[30] (PDB 2OL5_A) was selected as a structural archetype for PaiB/PzbB ESPRIPT[31] alignment. Efforts were taken to sample PaiB sequences from diverse microorganisms. To create a phylogeny of PzbB and PaiB proteins, the amino acid sequences of 85 putative PzbB, 51 PaiB, and the C-terminal domains of 10 putative PzbAB chimeras were aligned with ClustalOmega. Highly similar PzbB sequences from overrepresented species were omitted and the alignment was manually de-gapped. A maximum likelihood phylogeny was estimated using the WAG substitution model with 500 bootstrap simulations. Low confidence branches (<50% of bootstrap replicates) were collapsed in CLC Main Workbench Ver 8.1. Sequence logos were generated from alignments at: https://weblogo.berkeley.edu

SfaB, SfaC and PzbAB purification. 5-mL cultures of *E. coli* Rosetta 2(DE3) carrying expression plasmids, pYH004, pYH048 or pYH089 were grown overnight and used to inoculate 1 L of LB containing ampicillin (100 µg/mL) and chloramphenicol (12 µg/mL), in 2.8-L baffled flasks. After shaking at 30°C to mid-log phase, cultures were cold shocked in an ice-water bath and induced with IPTG (Isopropyl β-D-1-thiogalactopyranoside, 100 µM final). After returning to an 18˚C shaker overnight, induced cells were pelleted, decanted, and suspended in 50 mL of lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 25

mM imidazole, pH 8.0). The resulting suspension was passed through an Avestin-High Pressure Homogenizer (>16000 psi) and the lysate was cleared via centrifugation at 13,000 × *g* for 30 min. For SfaC-His, ammonium sulfate (40% final concentration) precipitation was additionally required to enrich the soluble fraction before prior to affinity purification. Cleared lysates were incubated with lysis buffer-equilibrated Ni-NTA resin for 30 min at 4˚C. After loading to a gravity column, the resin was washed with five column-volumes of wash buffer (50 mM HEPES, 500 mM NaCl, 50 mM imidazole, pH 8.0), 3X. For His_6-SfaB, the column turned noticeably yellow. For SfaC-His_6, the column turned noticeably brownish-red. Bound proteins were eluted with 5 mL of elution buffer (50 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 8.0). After analysis by SDS-PAGE, the elution fraction was concentrated (Pierce Concentrators, 9K MWCO, 7 mL) to ~0.5 mL and subsequently desalted into 50 mM HEPES, pH 8.0 using Zeba Spin Desalting Columns (7K MWCO, 2 mL). Protein concentrations were measured via Bradford assay with a BSA standard curve. Enzyme fractions estimated to be >90% pure following SDS-PAGE were immediately assayed or were frozen on dry ice and stored at -80˚C. See Supplementary for enzyme cofactor characterization and SfaB assay details.

SfaC kinetic assay. A 1 mM hemin stock solution was prepared in 20 mM NaOH. Kinetic assays were set up with the following conditions: SfaC-His_6 (0.05 μM to 10 μM), N^5-OH-L-ornithine-HCl (5 mM, saturating), hemin (0.5 μM to 100 μM), and Tris-HCl (50 mM, pH 8.0). Hemin to enzyme ratio was maintained in each assay at 10:1. The reaction was incubated at 22˚C for 10 minutes before addition of substrate. At 30, 60, and 90 s, an aliquot was removed from the mixture and quenched with an equal volume of 6% sulfosalicylic acid for protein precipitation and sample acidification. The cloudy mixture was centrifuged at 21000 x *g* for 5 minutes, and the supernatant was analyzed by LC-MS/MS for Piz as described above. The rate of Piz formation was divided by the enzyme concentration to obtain k_{cat}. Apparent K_m was calculated from reaction progress curves using DynaFit[32] software (Figure S9A).

Coupled PzbA/B end-point assays. SfaB/C enzyme assays were set up with the following components: His_6-SfaB (10 μM), SfaC-His_6 (50 nM), L-ornithine-HCl (1 mM), hemin (0.5 μM), FAD (50 μM), NADPH (2 mM) and Tris-HCl (50 mM, pH 8.0). PzbAB (*M. tulbaghiae*) enzyme assays were set up with the following components: PzbAB-His_6 (5 μM), L-Ornithine-HCl (1 mM), hemin (50 μM), FAD (50 μM), NADPH (2 mM) and Tris-HCl (50 mM, pH 8.0). Prior to setting up these assays, the reaction buffer was vigorously

shaken just prior to use to increase dissolved oxygen. Reactions were incubated at 22˚C for 10 minutes before addition of substrate to allow cofactor incorporation. Fifteen minutes after initiating catalysis with L-Orn, reactions were stopped with an equal volume of 6% sulfosalicylic acid. Mixtures cleared via centrifugation at 21000 x *g* for 5 minutes, and supernatants were analyzed by LC-MS as described above.

Characterization of enzymatically produced L-Piz. 200 μL SfaC reactions were set up as follows: 50 mM phosphate buffer, pH 8.0; SfaC-His_6 (50 nM); hemin (500 nM); and N^5-OH-L-ornithine-HCl (5 mM). After mixing and allowing reaction to proceed for ~15 minutes, an equal volume of 6% sulfosalicylic acid was added to precipitate the protein and quench the reaction. The solution was centrifuged at 21,000 x *g* for 5 minutes. The supernatant was diluted 1:10 with 20 mM acetic acid. For cation-exchange purification, a Discovery DSC-SCX SPE column (500 mg, Sigma) was conditioned with 2 mL of methanol followed by 2 mL of 20 mM acetic acid. The sample was applied to the column and washed with 2 mL of methanol. L-Piz was eluted with 2 mL of 5% ammonium hydroxide in 80% methanol. The elution fraction was evaporated under vacuum and resuspended in 100 μL LC-MS-grade water for HRMS analysis (Figure S3).

Marfey's analysis. Enzymatic Piz was derivatized with 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA, Marfey's reagent) to determine stereochemistry. 50 μL of 1% FDAA in acetone was added to 100 μL aqueous sample. The resulting yellow solution slowly became cloudy, and following incubation at 50˚C for 1 hour, the solution turned orange. 100 μL of 1 M HCl was added to quench the reaction. Finally, 100 μL of MeCN was added to dissolve the precipitate. The supernatant was filtered (Agilent Captiva Econo Filter, 0.2 μm) before LC-MS/MS. Analysis was performed using a Phenomenex Luna C18 column (75 x 3 mm, 3 μm) using the following method: *T* = 0, 10% B; *T* = 5, 10% B; *T* = 25, 100% B; *T* = 27, 100% B, *T* = 29, 10% B, *T* = 30, 10% B; A: water + 0.1% formic acid, B: MeCN + 0.1% formic acid; 0.6 mL/min. Stereochemistry was determined against FDAA-derivatized authentic L-Piz and DL-Piz racemate controls via LC-MS at $[M + H]^+$ *m/z* 383.1 and 340 nm (Figure S5).

L-Piz NMR. ^{1}H NMR spectroscopy was carried out at the WUSTL Chemistry NMR Facility. Spectra were measured using an Agilent 600 MHz instrument with DD2 console and Agilent 600 HCN cold probe. L-Piz was purified from *in vitro* reactions via SAX chromatography prior to NMR and dissolved in D_2O (Figure S4).

Supporting information

The Supporting Information is available free of charge on the ACS Publications website at http://pubs.acs.org.

Tables S1-S6, Figures S1-S9, and Supplementary methods.

Acknowledgments

We acknowledge R. Kranz and M. Sutherland (WUSTL) for valuable assistance with heme protein characterization. We also thank P. Kuzmic (BioKin, Ltd) for assistance with DynaFit software, M. Singh (WUSTL NMR labs) for assistance with spectral analyses, and A. Bose (WUSTL) for helpful suggestions. We also thank former WUSTL Bio3493 students P. Felder, C. Stump, B. Burger, C. Martini, M. Rickles-Young, L. Malcolm, E. Song and J. McMullen for assistance isolating, identifying, and transforming environmental *Streptomyces* strains used in this study. HRMS measurements were made at the WUSTL Biomedical Mass Spectrometry Resource, funded by NIH 8P41GM103422. This work was supported by WUSTL Startup Funds and National Science Foundation CAREER award 1846005 to J. Blodgett and a BioSURF Fellowship to Y. Hu. We are indebted to S. Javadov (U. Puerto Rico School of Medicine) for a sample of authentic sanglifehrin A for production quantification.

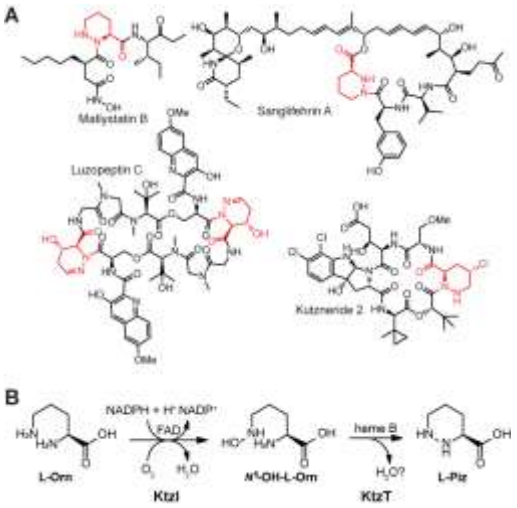

Figure 1. Piz Natural Products and Biosynthesis. (**A**) Piperazic acid (Piz, in red) is incorporated into structurally diverse peptidyl natural products. (**B**) Current biosynthetic pathway for Piz production based on *Kutzneria* sp. strain 744 KtzI and KtzT characterization.

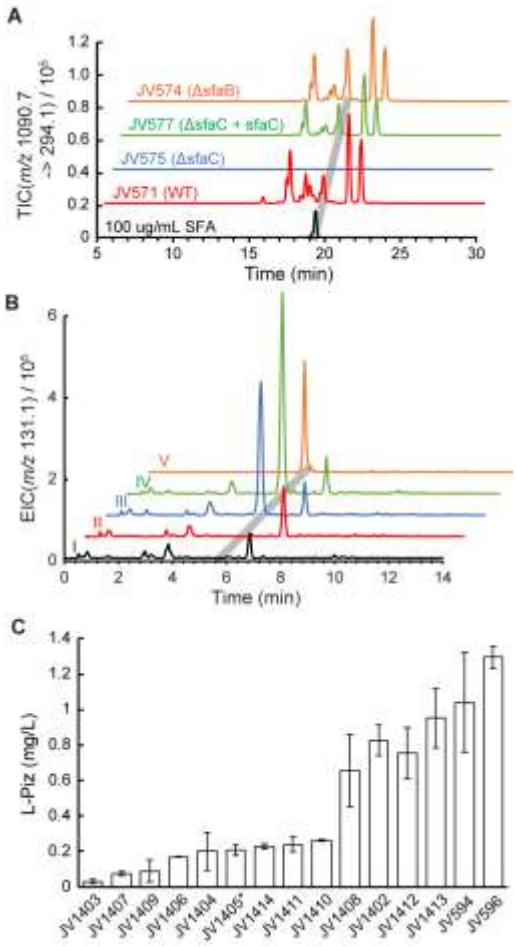

Figure 2. Genetic analysis of *pzbB* in Piz production *in vivo.* (**A**) LC-MS/MS detection of sanglifehrins (SRM *m/z* 1090.7 -> 294.1) produced by: *S. flaveolus rpsL*^{K43R} (JV571); *S. flaveolus rpsL*^{K43R} Δ*sfaC* (JV575); *S. flaveolus rpsL*^{K43R} Δ*sfaC* +*sfaC* (JV577); and *S. flaveolus rpsL*^{K43R} Δ*sfaB* (JV574). Grey bar highlights the retention time for an authentic standard of sanglifehrin A. (**B**) Piz detection in transgenic *S. lividans* by LC-MS (*m/z* 131.1). (I) Wild-type; (II) + *sfaB* (JV593); (III) + *sfaC* (JV594); (IV) + *sfaB and sfaC* (JV596); and (V) 50 μM L-Piz standard. Grey bar indicates position of authentic Piz (**C**) LC-MS/MS quantification of L-Piz production in *sfaC*-transformed *Streptomyces* species. Strains JV1403-JV1413 are environmental isolates; JV1405 (starred) produces similar quantities of Piz in the absence of *sfaC* (see text). Error bars indicate standard deviations between biological triplicates. JV594 and JV596 are *S. lividans* transformants in (**B**).

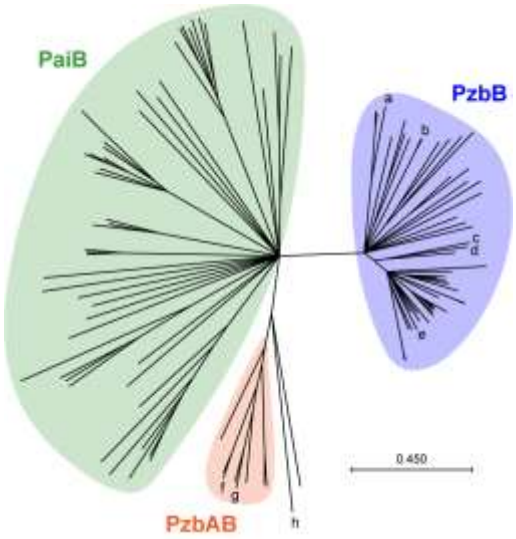

Figure 3. Maximum-Likelihood phylogeny of PzbB homologs. PzbB enzymes (blue) are closely related to PaiB proteins (green) that are found in multiple bacterial phyla and some fungi. This analysis reveals a clade of PzbAB fusion enzymes (orange) whose PzbB domains are sequence-distinct from stand-alone PzbBs. Nodes a-e correspond to PzbBs tested *in vivo* in Figs. 2 and 4 (from *S. flaveolus, A. cyanogriseus, L. flaviverrucosa* and *Kutzneria* sp. 744 and *S. himastatinicus,* respectively). f and g correspond to PzbAB fusion proteins investigated herein from *M. tulbaghiae* and *A. alba*. All PzbAB and all except for two PzbBs are sourced from actinobacteria. Node h corresponds to a putative stand-alone PzbB protein from the proteobacterium *Photorhabdus*; another proteobacterial PzbB from *Collimonas fungivorans* groups within the actinobacterial PzbB tree (unmarked). A single stand-alone PzbB from actinobacterium *Mycobacterium xenopi* RIVM700367 clades within in the PzbAB group (unmarked).

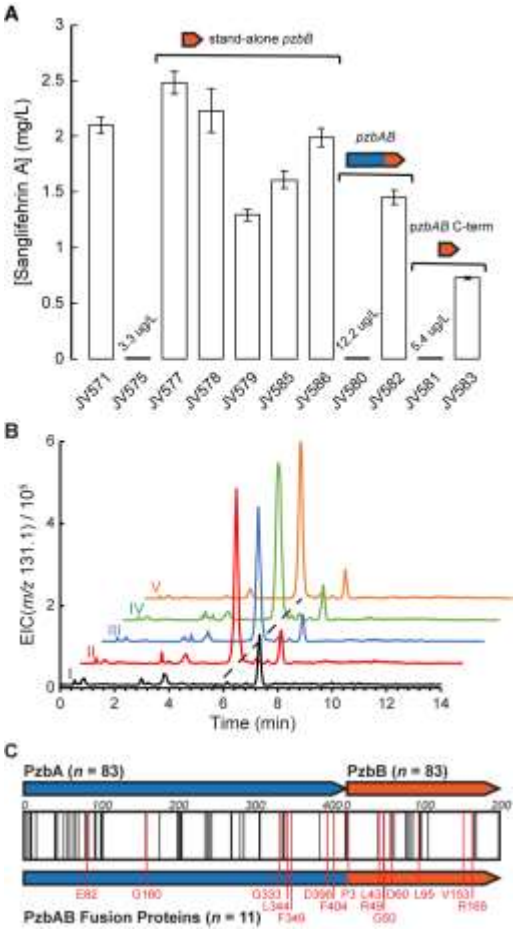

Figure 4. Piz production by PzbB orthologs. (**A**) LC-MS/MS quantification of sanglifehrin A production in *S. flaveolus rpsL*^{K43R} (JV571); *rpsL*^{K43R} Δ*sfaC* (JV575); and Δ*sfaC* complemented with sequence-diverse *pzbB*s: *sfaC* (JV577); *ktzT* (JV578); *hmtC* (JV579); *pzbB*$_{A.\ cyanogriseus}$ (JV585); *pzbB*$_{L.\ flaviverrucosa}$ (JV586). JV575 was also complemented with PzbAB chimeras of *A. alba* and *M. tulbaghiae* (JV580 and JV582, respectively) as well as the *pzbB* domains cloned from the same genes (JV581 and JV583, respectively). Error bars indicate standard deviations between biological triplicates. (**B**) LC-MS chromatograms for Piz production in (I) *S. lividans* +*sfaB* (JV593) expressing: (II) *hmtC*$_{S.\ himistatinicus}$ (JV598); (III) *pzbB*$_{A.\ cyanogriseus}$ (JV599); (IV) *pzbB*$_{L.\ flaviverrucosa}$ (JV600); and (V) *ktzT* (JV597). (**C**) PzbAB fusion proteins have an N-terminal PzbA (ornithine hydroxylase) domain and C-terminal PzbB domain. Conserved residues are indicated by black bars. Residues conserved in only PzbAB sequences are indicated by red bars.

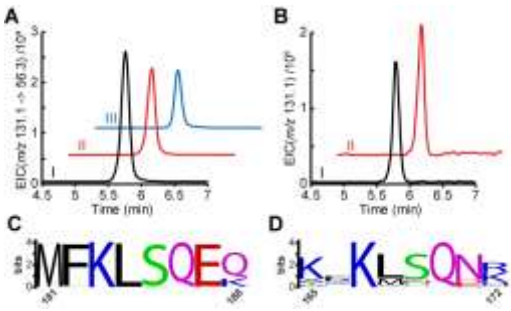

Figure 5. *In vitro* PzbB and coupled PzbA/B assays. (**A**) LC-MS chromatograms for (**I**) 50 μM L-Piz standard; and *in vitro* endpoint products of (II) SfaC and (III) PzbAB$_{M.\ tulbaghiae}$ using L-*N*^{5}-hydroxy ornithine as the substrate. (**B**) LC-MS chromatograms of *in vitro* reaction endpoints of (I) coupled SfaB/SfaC and (II) PzbAB$_{M.\ tulbaghiae}$ using L-ornithine as the substrate. (**C** and **D**) Logo plots of residues surround the conserved KLSQ motifs of PzbB and PaiB proteins, respectively. The numbers below the first and last amino acids correspond to their respective positions in *S. flaveolus* SfaC and *B. subtilis* 168 PaiB.

REFERENCES

[1] Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America, *Clin. Infect. Dis. 48*, 1-12.

[2] Hassall, C., Morton, R., Ogihara, Y., and Phillips, D. (1971) Amino-acids and peptides. Part XII. The molecular structures of the monamycins, cyclodepsipeptide antibiotics, *J. Chem. Soc. C. 0*, 526-532.

[3] Ogita, T., Sato, A., Enokita, R., Suzuki, K., Ishii, M., Negishi, T., Okazaki, T., Tamaki, K., and Tanzawa, K. (1992) Matlystatins, new inhibitors of typeIV collagenases from *Actinomadura atramentaria*. I. Taxonomy, fermentation, isolation, and physico-chemical properties of matlystatin-group compounds., *J. Antibiot. 45*, 1723-1732.

[4] Madison, V., Duca, J., Bennett, F., Bohanon, S., Cooper, A., Chu, M., Desai, J., Girijavallabhan, V., Hare, R., and Hruza, A. (2002) Binding affinities and geometries of various metal ligands in peptide deformylase inhibitors, *Biophys. Chem. 101*, 239-247.

[5] Huang, C.-H., Mirabelli, C. K., Mong, S., and Crooke, S. T. (1983) Intermolecular cross-linking of DNA through bifunctional intercalation of an antitumor antibiotic, luzopeptin A (BBM-928A), *Cancer Res. 43*, 2718-2724.

[6] Broberg, A., Menkis, A., and Vasiliauskas, R. (2006) Kutznerides 1-4, depsipeptides from the actinomycete *Kutzneria* sp. 744 inhabiting mycorrhizal roots of *Picea abies* seedlings., *J. Nat. Prod. 69*, 97-102.

[7] Sanglier, J. J., Quesniaux, V., Fehr, T., Hofmann, H., Mahnke, M., Memmert, K., Schuler, W., Zenke, G., Gschwind, L., Maurer, C., and Schilling, W. (1999) Sanglifehrins A, B, C and D, novel cyclophilin-binding compounds isolated from *Streptomyces* sp. A92-308110. I. Taxonomy, fermentation, isolation and biological activity., *J. Antibiot. 52*, 466-473.

[8] Pua, K. H., Stiles, D. T., Sowa, M. E., and Verdine, G. L. (2017) IMPDH2 is an intracellular target of the Cyclophilin A and Sanglifehrin A complex, *Cell Rep. 18*, 432-442.

[9] A Ciufolini, M., and Xi, N. (1998) Synthesis, chemistry and conformational properties of piperazic acids, *Chem. Soc. Rev. 27*, 437-439.

[10] Oelke, A. J., France, D. J., Hofmann, T., Wuitschik, G., and Ley, S. V. (2011) Piperazic acid -containing natural products : Isolation, biological relevance and total synthesis *Nat. Prod. Rep. 28*, 1445-1471.

[11] Handy, E. L., and Sello, J. K. (2015) Structure and Synthesis of Conformationally Constrained Molecules Containing Piperazic Acid, in *Peptidomimetics I* (Lubell, W., Ed.) pp 1-29, Springer Berlin Heidelberg, Berlin, Heidelberg.

[12] Neumann, C. S., Jiang, W., Heemstra, J. R., Gontang, E. A., Kolter, R., and Walsh, C. T. (2012) Biosynthesis of Piperazic Acid via N^5 - Hydroxy - Ornithine in *Kutzneria* spp. 744, *ChemBioChem 13*, 972-976.

[13] Du, Y.-L., He, H.-Y., Higgins, M. A., and Ryan, K. S. (2017) A heme-dependent enzyme forms the nitrogen–nitrogen bond in piperazate, *Nat. Chem. Bio. 13*, 836-838.

[14] Qu, X., Jiang, N., Xu, F., Shao, L., Tang, G., Wilkinson, B., and Liu, W. (2011) Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor, *Mol. BioSyst. 7*, 852-861.

[15] Fehr, T., Kallen, J., Oberer, L., and Sanglier, J. J. (1999) Sanglifehrins A, B, C and D, novel cyclophilin-binding compounds isolated from *Streptomyces* sp. A92-308110. II. Structure elucidation, stereochemistry and physico- chemical properties, *J. Antibiot. 52*, 474-479.

[16] Rückert, C., Albersmeier, A., Busche, T., Jaenicke, S., Winkler, A., Friðjónsson, Ó. H., Hreggviðsson, G. Ó., Lambert, C., Badcock, D., and Bernaerts, K. (2015) Complete genome sequence of *Streptomyces lividans* TK24, *J. Biotechnol. 199*, 21-22.

[17] Fischbach, M. A., and Walsh, C. T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms., *Chem. Rev. 106*, 3468-3496.

[18] Van Dessel, W., Van Mellaert, L., Geukens, N., and Anné, J. (2003) Improved PCR-based method for the direct screening of *Streptomyces* transformants, *J. Microbiol. Meth. 53*, 401-403.

[19] Blodgett, J. A. V., Oh, D.-C., Cao, S., Currie, C. R., Kolter, R., and Clardy, J. (2010) Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria., *Proc. Natl. Acad. Sci. U S A 107*, 11692-11697.

[20] Stach, J. E. M., Maldonado, L. A., Ward, A. C., Goodfellow, M., and Bull, A. T. (2003) New primers for the class Actinobacteria: application to marine and terrestrial environments., *Env. Microbiol*. *5*, 828-841.

[21] Henikoff, S., Henikoff, J., Alford, W., and Pietrokovski, S. (1995) Automated construction and graphical presentation of protein blocks from unaligned sequences, *Gene 163*, GC17-26.

[22] Rose, T., Schultz, E., Henikoff, J., Pietrokovski, S., McCallum, C., and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences, *Nucl. Acids Res. 26*, 1628-1635.

[23] Qi, Y., Ding, E., and Blodgett, J. A. (2017) Native and engineered clifednamide biosynthesis in multiple *Streptomyces* spp, *ACS Syn. Bio. 7*, 357-362.

[24] Hirsch, C., and Ensign, J. (1976) Heat activation of *Streptomyces viridochromogenes* spores, *J. Bacteriol. 126*, 24-30.

[25] Blodgett, J. A. V., Thomas, P. M., Li, G., Velasquez, J. E., van der Donk, W. A., Kelleher, N. L., and Metcalf, W. W. (2007) Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide, *Nat. Chem. Biol. 3*, 480-485.

[26] Hosted, T., and Baltz, R. (1997) Use of *rpsL* for dominance selection and gene replacement in *Streptomyces roseosporus*, *Journal of Bacteriology 179*, 180-186.

[27] Keiser, T., Bibb, M., Buttner, M., Chater, K., and Hopwood, D. (2000) in *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich.

[28] Skinnider, M. A., Dejong, C. A., Rees, P. N., Johnston, C. W., Li, H., Webster, A. L., Wyatt, M. A., and Magarvey, N. A. (2015) Genomes to natural products prediction informatics for secondary metabolomes (PRISM), *Nucl. Acids Res. 43*, 9645-9662.

[29] Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers, *Nucl. Acids Res. 41*, W204-212.

[30] Filippova, E. V., Brunzelle, J. S., Cuff, M. E., Li, H., Joachimiak, A., and Anderson, W. F. (2011) Crystal structure of the novel PaiB transcriptional regulator from *Geobacillus stearothermophilus*, *Proteins 79*, 2578-2582.

[31] Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins, *Nucl. Acids Res*. *31*, 3320-3323.

[32] Kuzmič, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase, *Anal. Biochem. 237*, 260-273.