Biosynthesis of isonitrile lipopeptides by conserved nonribosomal peptide synthetase gene clusters in Actinobacteria

Nicholas C. Harris[a], Michio Sato[b], Nicolaus A. Herman[c], Frederick Twigg[c], Wenlong Cai[c], Joyce Liu[d], Xuejun Zhu[c], Jordan Downey[c], Ryan Khalaf[e], Joelle Martin[e], Hiroyuki Koshino[f], and Wenjun Zhang[c,g,1]

[a]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [b]Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan; [c]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720; [d]Department of Bioengineering, University of California, Berkeley, CA 94720; [e]Department of Chemistry, University of California, Berkeley, CA 94720; [f]RIKEN Physical Center for Sustainable Resource Science, Wako, Saitama 3510198, Japan; and [g]Chan Zuckerberg Biohub, San Francisco, CA 94158

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved May 26, 2017 (received for review March 27, 2017)

A putative lipopeptide biosynthetic gene cluster is conserved in many species of Actinobacteria, including *Mycobacterium tuberculosis* and *M. marinum*, but the specific function of the encoding proteins has been elusive. Using both in vivo heterologous reconstitution and in vitro biochemical analyses, we have revealed that the five encoding biosynthetic enzymes are capable of synthesizing a family of isonitrile lipopeptides (INLPs) through a thio-template mechanism. The biosynthesis features the generation of isonitrile from a single precursor Gly promoted by a thioesterase and a nonheme iron(II)-dependent oxidase homolog and the acylation of both amino groups of Lys by the same isonitrile acyl chain facilitated by a single condensation domain of a nonribosomal peptide synthetase. In addition, the deletion of INLP biosynthetic genes in *M. marinum* has decreased the intracellular metal concentration, suggesting the role of this biosynthetic gene cluster in metal transport.

biosynthetic enzymes | mycobacteria | metal transport

Small-molecule secondary metabolites are produced by microbes as chemical weapons to combat competing organisms or as communication signals to control complex processes such as virulence, morphological differentiation, biofilm formation, and metal acquisition (1–3). One of the most important classes of secondary metabolites are nonribosomal peptides, which are typically biosynthesized by modular nonribosomal peptide synthetases (NRPSs) in an assembly-line manner (4). Two NRPS-encoding gene clusters (*mbt* and *Rv0096-0101*) have been identified from the genome of *Mycobacterium tuberculosis*, the leading causative agent of tuberculosis that currently infects one-third of the world's population. Although the cluster of *mbt* has been characterized to biosynthesize mycobactin siderophores that form mycobactin-Fe(III) complexes for iron sequestration (5), the role of *Rv0096-0101* remains obscure despite various biological studies that have indicated the production of a virulence factor by this gene cluster (6–14). For example, using transposon-site hybridization, *Rv0098* to *Rv0101* were predicted to be required for *M. tuberculosis* survival in a mouse model of infection (10). Consistently, a transposon insertion of *Rv0097* attenuated *M. tuberculosis* growth and survival in mice (7).

An in silico homology search has revealed that gene clusters homologous to *Rv0096-0101* are conserved in pathogenic mycobacteria, such as *M. bovis*, *M. leprae*, *M. marinum*, *M. ulcerans*, and *M. abscessus* (Fig. 1), but absent in nonpathogenic mycobacteria such as *M. smegmatis*, providing further indication of the virulence-associated nature of the locus product in mycobacteria. Interestingly, in addition to the genus of *Mycobacterium*, related operons are found in the phylum of Actinobacteria across genera including *Streptomyces*, *Kutzneria*, *Nocardia*, and *Rhodococcus* (Fig. 1), suggesting a widespread presence of this cluster. Further bioinformatic analysis has shown that five genes (*Rv0097-0101*) are conserved across all identified gene clusters and that these genes encode proteins homologous to an iron(II) and α-ketoglutarate (α-KG)-dependent oxidase, a fatty acyl-CoA thioesterase, an acyl-acyl carrier protein ligase (AAL), an acyl carrier protein (ACP), and a single- or dimodule NRPS, respectively (Fig. 1 and *SI Appendix*, Fig. S1). Although all of these five proteins are typically involved in secondary metabolite biosynthesis, the identity of the corresponding metabolite and the specific function of these proteins have not yet been fully elucidated.

To better understand the role of this widespread gene cluster, we turned to reconstitution in *Escherichia coli* as a means to quickly and systematically assess the function of the five conserved enzymes through metabolomic exploration. The enzymes from *M. tuberculosis* H37Rv, *M. marinum* strain M (an opportunistic human pathogen), and *Streptomyces coeruleorubidus* NRRL18370 (15) (a known pacidamycin producer) were studied and compared, which revealed similarities and variations in biosynthetic functions (Fig. 1). We discovered that these five conserved enzymes were necessary and sufficient to synthesize a unique group of isonitrile lipopeptides (INLPs). Based on both in vivo reconstitution and in vitro biochemical analysis, we scrutinized the timing and substrate specificity of these enzymes in INLP biosynthesis. Additionally, a mutagenesis study in *M. marinum* suggests that this biosynthetic gene cluster plays a role in metal acquisition.

Significance

***Mycobacterium tuberculosis* is the leading causative agent of tuberculosis, from which millions die annually. A putative lipopeptide biosynthetic gene cluster has been shown to be essential for the survival of this pathogen in hosts, and homologous gene clusters have also been found in all pathogenic mycobacteria and other species of Actinobacteria. We have identified the function of these gene clusters in making a family of isonitrile lipopeptides. The biosynthesis has several unique features, including an unprecedented mechanism for isonitrile synthesis. Our results further suggest that these biosynthetic gene clusters play a role in metal transport and thus have shed light on a metal transport system that is crucial for virulence of pathogenic mycobacteria.**

Author contributions: N.C.H. and W.Z. designed research; N.C.H., M.S., N.A.H., F.T., W.C., J.L., X.Z., J.D., R.K., J.M., and W.Z. performed research; N.C.H., M.S., N.A.H., H.K., and W.Z. analyzed data; and N.C.H. and W.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: MS and MS/MS data have been deposited in the European Molecular Biology Laboratory–European Bioinformatics Institute MetaboLights database (DOI: 10.1093/nar/gks1004; PubMed PMID: 23109552) with the identifier MTBLS454. The complete dataset can be accessed at www.ebi.ac.uk/metabolights/MTBLS454.

[1]To whom correspondence should be addressed. Email: wjzhang@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705016114/-/DCSupplemental.

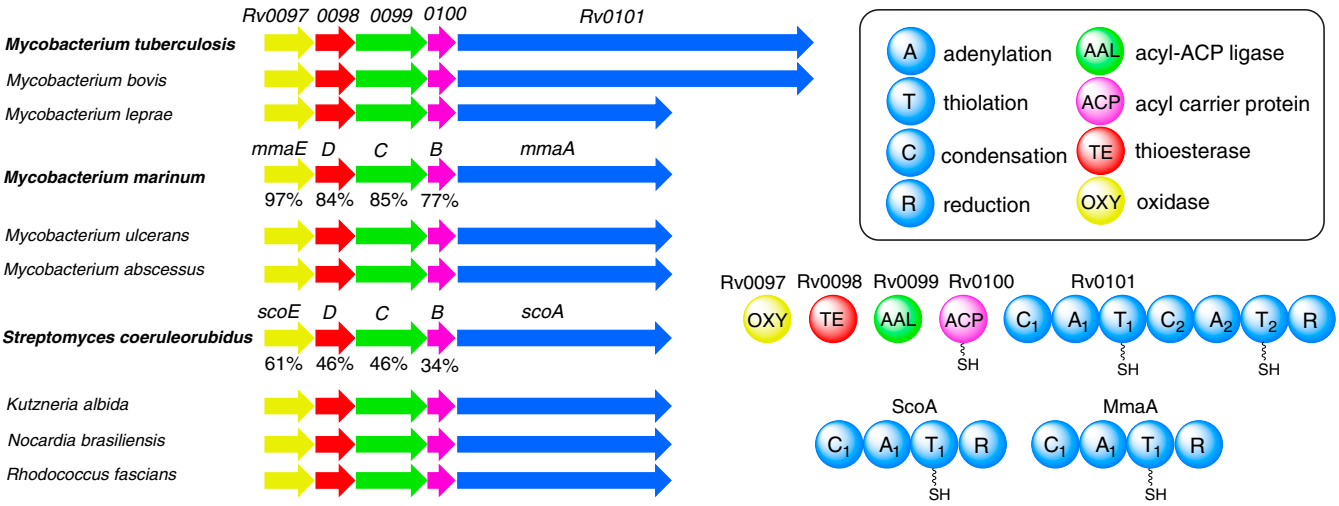


Fig. 1. Schematic of the selected conserved biosynthetic gene clusters and their encoding protein products. Thousands of homologous gene clusters have been identified from published genomes. The similarity of protein homologs from *M. marinum* and *S. coeruleorubidus* to *M. tuberculosis* is shown below the gene clusters. A_1 in Rv0101, MmaA, and ScoA was predicted to activate Lys based on the 10-residue specificity sequence (DIEDVGSVVK, DIEDVGSVVK, and DTEDVGTVVK, respectively). A_2 in Rv0101 was predicted to activate Phe (DAWTVAAICK).

Results and Discussion

Biosynthesis of INLPs by ScoA-E in *E. coli*. Previous biochemical studies of Rv0099-0101 suggested that a lipopeptide might be produced by these enzymes (16, 17). To unveil molecular features of the putative lipopeptide and gain preliminary knowledge of the functions of the five conserved biosynthetic enzymes, *scoA-E* from *S. coeruleorubidus* and *mmaA-E* from *M. marinum* (Fig. 1) were cloned for *E. coli* heterologous expression and functional reconstitution, respectively. A negative control strain transformed with empty vectors was also constructed. Protein expression analysis showed that all 10 proteins were solubly expressed in *E. coli* (*SI Appendix*, Fig. S2). The *E. coli* cultures were extracted with organic solvents, concentrated, and analyzed by liquid chromatography-high resolution mass spectrometry (LC-HRMS) followed by untargeted metabolomics analysis using XCMS (18, 19) for the determination of metabolic profile differences and the identification of new metabolites.

Although no new metabolite was immediately identified in the culture of *E. coli-mmaA-E* compared with the negative control strain, upon coexpression of *scoA-E*, two new metabolites with molecular formulas $C_{18}H_{28}N_4O_4$ (INLP **1**, calculated for $C_{18}H_{29}N_4O_4^+$: 365.2183; found: 365.2185) and $C_{16}H_{26}N_4O_3$ (INLP **2**, calculated for $C_{16}H_{27}N_4O_3^+$: 323.2078; found: 323.2079) that were absent in the negative control were found to be produced (Fig. 2 and *SI Appendix*, Fig. S3). Both metabolites were UV-inactive above 210 nm and acid sensitive. The minor metabolite INLP **2** was further predicted to

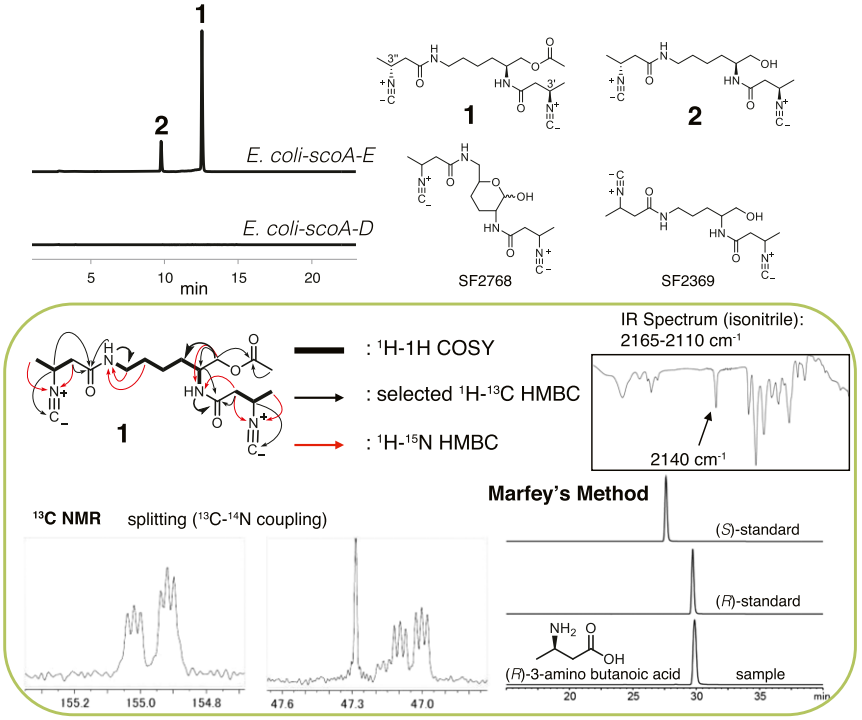


Fig. 2. Biosynthesis of INLPs (INLP **1** and **2**) by ScoA-E in *E. coli*. Extracted ion chromatograms show production of INLP **1** and **2** by *E. coli-scoA-E*. The calculated masses for INLP **1** and **2** with 10-ppm mass error tolerance were used for each trace. The negative control strain with empty vectors and strains containing any of the four gene combinations did not produce INLP **1** and **2**, and only one representative trace is shown here for simplicity. The structural determination of INLP **1** is boxed. Structures of two known Actinomycetes metabolites are also shown here.

be deacetylated INLP **1** based on HRMS/MS analysis (*SI Appendix*, Fig. S3). To purify a sufficient quantity of INLP **1** and **2** for structural elucidation, a total of ∼40 L of *E. coli* culture was prepared and extracted with chloroform, followed by purification via multiple rounds of HPLC using reverse-phase C18 columns. These purification steps yielded ∼3 mg of pure compound INLP **1** and ∼0.5 mg of INLP **2**. NMR spectra, including ^{1}H, ^{13}C, dqf-COSY, HSQC, HMBC, and ROESY spectra, were obtained for compound INLP **1** and used to determine its molecular connectivity (Fig. 2 and *SI Appendix*, Fig. S4 and Table S1). The presence of an isonitrile moiety in INLP **1** was further confirmed by comparison with the reported ^{13}C-^{14}N nuclear spin coupling constants and IR spectroscopy absorption (Fig. 2) (20). The absolute configuration of the C3′ and C3′′ (both *R*) was determined by acid hydrolysis to yield the monomer of 3-amino butyric acid, which was then reacted with Marfey's reagent and compared with the standards (Fig. 2) (21). The NMR spectra of **2** confirmed that it is a C1 deacetylated analog of INLP **1** (Fig. 2 and *SI Appendix*, Fig. S5 and Table S2). The molecular structures of INLP **1** and **2** were revealed to be similar to two known isonitrile antibiotics (SF2768 and SF2369) that were originally isolated from the culture filtrates of different Actinomycetes species (22–24) (Fig. 2), strongly indicating that *E. coli* was a suitable heterologous host for the reconstitution of ScoA-E activities to produce relevant metabolites. Systematic removal of each of the five genes from *E. coli-scoA-E* abolished the production of INLP **1** and **2**, demonstrating the necessity of all five enzymes in synthesizing these two INLPs (Fig. 2).

Proposed Biosynthetic Pathways for INLPs by the Five Conserved Biosynthetic Enzymes. The heterologous production of INLP **1** and **2** in *E. coli* enabled the assignment of function to the five conserved biosynthetic enzymes, especially the role of the two modification enzymes (ScoD and ScoE), in a unique isonitrile moiety synthesis. We have thus proposed putative enzymatic pathways for INLP biosynthesis (Fig. 3). The INLP **2** is presumably generated through a hybrid AAL-NRPS assembly-line–based mechanism. The assembly line starts with the activation of crotonic acid by ScoC through adenylation, and the adenylated acid is loaded onto an ACP (ScoB) for further processing. The α,β-unsaturated fatty acyl-ACP is then modified by a thioesterase homolog (ScoD) and a nonheme iron(II)-dependent oxidase (ScoE) to generate a β-isonitrile fatty acyl-ACP intermediate. Instead of being a thioesterase as suggested by the previous biochemical and structural analyses of Rv0098 (25), we propose ScoD to be a reverse lyase-like enzyme that catalyzes a Michael addition of Gly to the β-position of an α,β-unsaturated fatty acyl-ACP to yield an *N*-carboxymethyl-3-aminoacyl-ACP. ScoE presumably catalyzes the subsequent oxidation and decarboxylation to yield a β-isonitrile moiety. The isonitrile-modified fatty acyl chain is then condensed to both amino groups of Lys promoted by the NRPS (ScoA) and reductively released by the reduction (R) domain of the NRPS to form a terminal alcohol product INLP **2**. This reductive release mechanism is consistent with the identified activity of the R domain in Rv0101, which contains a conserved Ser/Tyr/Lys catalytic triad and was demonstrated to catalyze a four-electron thioester reduction using the purified truncated R domain and a synthetic substrate (17). We propose that the extra acetyl moiety found in INLP **1** is due to the activity of a promiscuous acetyltransferase endogenous to *E. coli* for possible detoxification, and similar acetylation events have been observed in other shunt product biosynthesis (26). An analogous biosynthetic pathway involving MmaA-E has also been proposed, differing in the chain length of the fatty acyl moiety (Fig. 3). Although ScoC seems to prefer a short-chain fatty acid substrate (such as crotonic acid as shown in INLP **1** and **2**), MmaC likely favors a medium-chain fatty acid substrate (such as 2-decenoic acid) because previous biochemical and structural analyses of Rv0099, a homolog of MmaC from the same genus of *Mycobacterium*, showed that this AAL activated fatty acids of medium-chain length (16, 17).

AAL, ACP, and NRPS Promote Diacylated Lipopeptide Biosynthesis. We next performed additional in vivo metabolic analyses using various combinations of biosynthetic genes and in vitro biochemical analyses using purified enzymes to dissect the proposed biosynthetic pathways for INLPs (Fig. 3). The substrate specificity of NRPSs, including ScoA and MmaA, was tested using the classical ATP-[^{32}P]PP$_i$ exchange assay. As expected, both enzymes demonstrated a strong preference for the activation of L-Lys (Fig. 4*A*), which is consistent with the molecular structures of INLP **1** and **2** and supports the assignment of the absolute configuration of C2 in INLP **1** to be 2*S* (Fig. 2). Although L-Ornithine was not activated by ScoA/MmaA, it could be a preferred substrate for other conserved NRPSs based on the structure of SF2369 (Fig. 2). We next probed the fatty acid substrate specificity of AALs encoded by *scoC* and

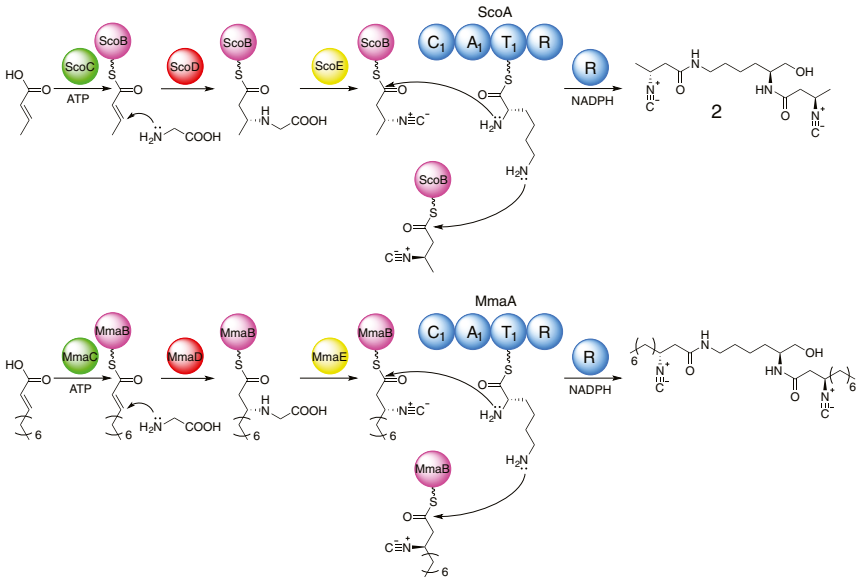


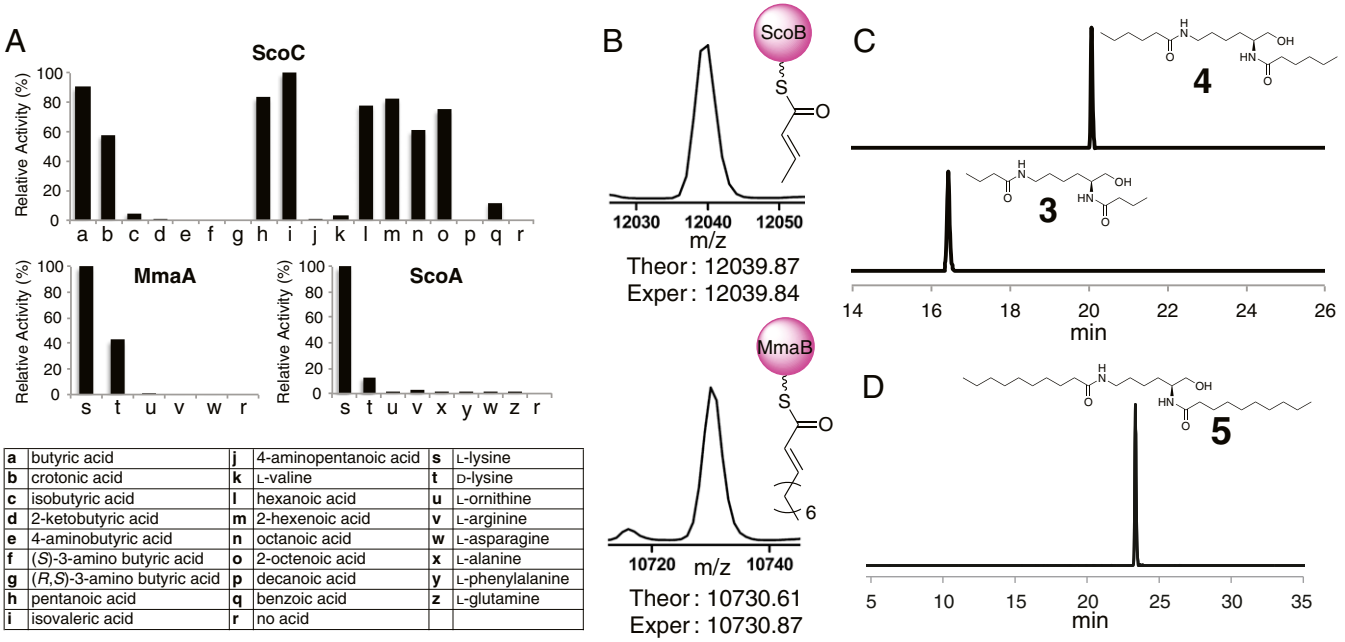
Fig. 3. Proposed function of ScoA-E and MmaA-E in INLP biosynthesis.

mmaC, respectively. The ability of ScoC to reversibly adenylate various acids was tested using the ATP-[^{32}P]PP$_i$ exchange assay. ScoC exhibited a strong preference for the activation of fatty acids with a short-chain length (C4–C8), and as expected, α,β-unsaturated fatty acids were well recognized (Fig. 4*A*). The subsequent loading of selected fatty acyl moieties onto ScoB was further confirmed by HRMS analysis (Fig. 4*B*). MmaC demonstrated an intrinsic ATPase activity in the ATP-PP$_i$ exchange assay that prohibited the determination of substrate specificity using this method. Nonetheless, the direct substrate activation and loading assays confirmed that, similar to Rv0099, MmaC preferentially activates fatty acids of medium-chain length with tolerance toward α,β-unsaturation (Fig. 4*B*).

One of the unusual events in the proposed biosynthetic pathway is the condensation of the same fatty acyl moiety to both amino groups of Lys, presumably promoted by a single C domain of the NRPS (Fig. 3). We hypothesized that this C domain has relaxed substrate specificity, and the product formation assays using ScoA-C or MmaA-C successfully yielded lipopeptides (LP **3–5**) with the amide bond formation at both amino positions (Fig. 4 *C* and *D* and *SI Appendix*, Fig. S6). Negative controls missing any of the enzyme or substrate (fatty acid, Lys, ATP, and NADPH) abolished the production of LP **3–5**. Products with a terminal alcohol were generated in these assays, confirming the four-electron reduction activity of the conserved R domain in ScoA and MmaA.

Homologs of Rv0097 and Rv0098 Promote Isonitrile Biosynthesis through an Unprecedented Mechanism. Although the isonitrile functionality has been found in quite a few natural products, only one biosynthetic pathway has been identified in which one carbon is transferred to an amino group catalyzed by an isonitrile synthase such as IsnA (27–30). Our proposed pathway for isonitrile synthesis is mechanistically distinct from the known pathway by using a different set of enzymes. Rv0098 has previously been shown to be a long-chain fatty acyl-CoA thioesterase, although structurally it lacks a general base or a nucleophile that is conserved in the thioesterase catalytic site. In addition, very low hydrolysis activities were obtained in the biochemical analysis of Rv0098, which solicited further detailed characterization of this hypothetical protein (25). The recent biochemical characterization of CmiS1, a homolog of ScoD (identity/similarity = 47%/56%), showed that CmiS1 catalyzed the Michael addition of Gly to the β-position of a non–2-enoic acid thioester in the biosynthesis of the macrolactam antibiotic cremimycin (31). We thus have proposed that a similar reaction of Michael addition of Gly to the β-position of an α,β-unsaturated fatty acyl-ACP could be promoted by ScoD/MmaD/Rv0098 to yield an *N*-carboxymethyl-3-aminoacyl-ACP (Fig. 3). This is consistent with the failed activation of 3-amino butyric acid by ScoC (Fig. 4*A*), which argued against the known pathway using an isonitrile synthase. To confirm the proposed function of MmaD, a biochemical reaction using MmaB-D, ATP, 2-decenoic acid, and Gly was performed, and the product was analyzed by LC-HRMS after release from the protein by base hydrolysis. This yielded the expected Gly adduct (calculated for $C_{12}H_{24}NO_4^+$: 246.1700; found: 246.1701), and its mass was shifted by +1 using [2-^{13}C]Gly or by +2 using [2-^{13}C, ^{15}N]Gly as an alternative substrate (Fig. 5*A* and *SI Appendix*, Fig. S7). Negative controls omitting any of the three proteins or substrate abolished the production of the Gly adduct, indicating that the formation of the Gly adduct requires the activation of the fatty acid substrate and occurs on a thio-templated assembly line (Fig. 5*A*). Similar reactions were also performed using ScoB-D, and the activity of ScoD in forming a Gly adduct was also confirmed (*SI Appendix*, Fig. S8). We further observed that the Gly adduct on ACP (either MmaB or ScoB) was readily released by hydrolysis in vitro, likely promoted by MmaD/ScoD, consistent with the observed thioesterase activity of Rv0098 and CmiS1 (5, 18). We thus could not reconstitute the activity of MmaE/ScoE, the iron(II), and the α-KG–dependent oxidase homolog in promoting the subsequent oxidation and decarboxylation in vitro, most likely due to substrate limitation.

Because reconstituting the activity of MmaE/ScoE in vitro has not yet been successful, we turned to in vivo reconstitution coupled with feeding of isotope-labeled Gly to further confirm the proposed activity of enzymes in isonitrile biosynthesis. Although coexpression of *scoA-C* in *E. coli* yielded expected diacylated lipopeptides such as



a	butyric acid	**j**	4-aminopentanoic acid	**s**	L-lysine
b	crotonic acid	**k**	L-valine	**t**	D-lysine
c	isobutyric acid	**l**	hexanoic acid	**u**	L-ornithine
d	2-ketobutyric acid	**m**	2-hexenoic acid	**v**	L-arginine
e	4-aminobutyric acid	**n**	octanoic acid	**w**	L-asparagine
f	(*S*)-3-amino butyric acid	**o**	2-octenoic acid	**x**	L-alanine
g	(*R*,*S*)-3-amino butyric acid	**p**	decanoic acid	**y**	L-phenylalanine
h	pentanoic acid	**q**	benzoic acid	**z**	L-glutamine
i	isovaleric acid	**r**	no acid		

Fig. 4. Biochemical analysis of AALs and NRPSs. (*A*) Substrate specificity of ScoA, MmaA, and ScoC determined by ATP-[^{32}P]PP$_i$ exchange assays. (*B*) Detection of crotonyl-*S*-ScoB and decenoyl-*S*-MmaB by MS with maximum entropy deconvolution. (*C*) Extracted ion chromatograms showing the production of LP **3** and **4** in ScoA-C assays using butyric acid and hexanoic acid as a substrate, respectively. (*D*) Extracted ion chromatogram showing the production of LP **5** in the MmaA-C assay using decanoic acid as a substrate. The calculated masses for LP **3–5** with 10-ppm mass error tolerance were used for each trace.

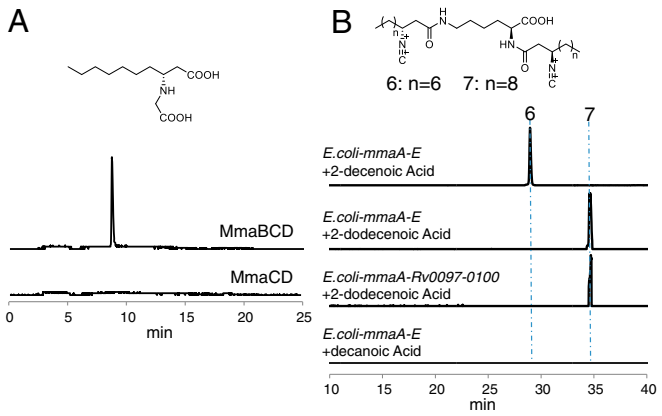


Fig. 5. In vitro and in vivo analysis of isonitrile formation. (*A*) Extracted ion chromatogram showing the production of Gly adduct in the MmaB-D assay. The calculated mass with 10-ppm mass error tolerance was used for each trace. Control assays missing any of the protein or substrate (ATP, 2-decenoic acid, and Gly) abolished the production of Gly adduct, and only one representative trace with no MmaB is shown for simplicity. (*B*) Extracted ion chromatograms showing the *E. coli*-based production of INLP **6** and **7** after feeding of 2-decenoic acid and 2-dodecenoic acid, respectively. The calculated masses for INLP **6** and **7** with 10-ppm mass error tolerance were used for each trace. Strains containing any of the four gene combinations, or feeding of decanoic acid and dodecanoic acid, did not produce INLP **6** and **7**, and only one representative trace with the feeding of decanoic acid to *E. coli-mmaA-E* is shown here for simplicity.

LP **3** and acetylated LP **3**, the addition of *scoD* or *scoE* to *E. coli-scoA-C* did not change the metabolic profile with no new metabolite identified through comparative metabolomics, indicating that the modification of the fatty acyl chain occurs before the biosynthetic intermediate release by the NRPS and that the C or R domain of ScoA is intolerant of a bulky side group on the fatty acyl chain (Gly adduct). Only upon the addition of both *scoD* and *scoE* to *E. coli-scoA-C* were new major products of INLP **1** and **2** identified (Fig. 2). We thus reasoned that ScoD and ScoE function on the assembly line, most likely on the free-standing ACP, ScoB, for isonitrile synthesis. Additionally, feeding of 10 mM of Gly to the culture of *E. coli-scoA-E* boosted the titer of INLP **1** by over 20-fold, and feeding of [2-^{13}C, ^{15}N]Gly showed that C(2)-N of Gly was incorporated into the isonitrile group of INLP **1** and **2**, supporting the proposed role of ScoD and ScoE in isonitrile biosynthesis (*SI Appendix*, Fig. S3).

To further confirm the role of MmaA-E in INLP biosynthesis, we reattempted the in vivo reconstitution of activities of MmaA-E in *E. coli*. Based on the substrate specificity of MmaC, we reasoned that the in vivo substrate limitation could be the reason for the failed INLP production by *E. coli-mmaA-E*. The 2-decenoic acid or 2-dodecenoic acid was then fed to the culture of *E. coli-mmaA-E*, and untargeted metabolomics analysis was performed to search for new metabolites. This led to the identification of a trace amount of two new metabolites with molecular formulas $C_{28}H_{48}N_4O_4$ (INLP **6**, calculated for $C_{28}H_{49}N_4O_4^+$: 505.3748; found: 505.3748) and $C_{32}H_{56}N_4O_4$ (INLP **7**, calculated for $C_{32}H_{57}N_4O_4^+$: 561.4374; found: 561.4376), respectively (Fig. 5*B* and *SI Appendix*, Fig. S9). HRMS analysis suggested that INLP **6** and **7** are INLPs similar to INLP **1** and **2**, but contain two longer fatty acyl chains and a C1 acid moiety (*SI Appendix*, Fig. S9). The presence of the isonitrile was further confirmed by IR spectroscopy to show signature absorption of 2,132 cm^{-1}. As expected, the production of INLP **6** and **7** was dependent on the coexpression of all five genes of *mmaA-E*, and feeding of [2-^{13}C, ^{15}N]Gly to the culture of *E. coli-mmaA-E* demonstrated that C(2)-N of Gly was incorporated into INLP **6** (Fig. 5*B* and *SI Appendix*, Fig. S9). It is notable that feeding of decanoic acid or dodecanoic acid did not lead to the production of INLP **6** and **7**, supporting our hypothesis that the biosynthesis of INLPs requires the activation of an α,β-unsaturated fatty acid (Fig. 3). In addition, replacement of *mmaB-E* by *Rv0097-0100* in *E. coli* produced INLP **7** upon the feeding of 2-dodecenoic acid, demonstrating the interchangeable property of the encoding four enzymes (Fig. 5*B*). The INLP alcohols could not be detected, suggesting that 2-decenoic acid and 2-dodecenoic acid may not be the native substrate for the *M. marinum* proteins that led to the impaired activity of the AAL-NRPS assembly line and low titers of the corresponding INLPs in *E. coli*. Indeed bioinformatic analysis showed that putative fatty acid modification enzymes are encoded in close proximity to *mmaA-E* in the *M. marinum* genome (*SI Appendix*, Fig. S1).

Implications for the Role of the Conserved Gene Clusters in Metal Transport. The isonitrile functionality is known to behave as an electron-rich analog of carbon monoxide and form coordination complexes with most transition metals (32, 33). Metals are required in many life processes, and pathogenic bacteria have addressed the challenge of deficiencies in essential metals or high concentrations of toxic metal cations in the host by evolving metal homeostasis mechanisms that are frequently associated with virulence (34–41). The operon of *Rv0096-0101* has been known to be critical for the virulence of *M. tuberculosis* and essential for the survival of this pathogen in macrophages and mice (7, 9, 10, 27, 32, 33). Based on the metal binding ability of isonitrile, we postulated that this conserved biosynthetic gene cluster plays a role in metal transport and homeostasis. In silico analysis demonstrated that putative heavy metal translocating P-type ATPases such as CtpA and CtpB are encoded in close proximity to the INLP biosynthetic genes in the *M. tuberculosis* and *M. marinum* genomes (*SI Appendix*, Fig. S1), and the expression of CtpB was shown to be coregulated with *Rv0096-0101* in *M. tuberculosis* by transcriptomic analysis (42). In addition, it has been reported that the expression of the *Rv0096-0101* gene cluster was normally repressed in many synthetic media including 7H9 where metal concentrations are high, but was constantly induced in Sauton's medium where metal concentrations are relatively low (6), suggesting the metal uptake function of the cluster. We next constructed a mutant strain of *M. marinum* devoid of *mmaA-E* through allelic exchange according to the reported method (43), and the intracellular metal content of the *M. marinum* wild-type and Δ*mmaA-E* was determined to probe the function of these genes. Deletion of the gene cluster caused a significant decrease in the intracellular accumulation of zinc in Sauton's medium as shown by inductively coupled plasma-optical emission spectrometry (ICP-OES) analysis (Fig. 6). This is consistent with the finding from microarray analysis that the transcription of *Rv0097* was down-regulated in *M. tuberculosis* exposed to excess zinc (38). These results strongly indicate that the conserved gene clusters synthesize a unique family of INLPs that promote metal transport in pathogenic mycobacteria.

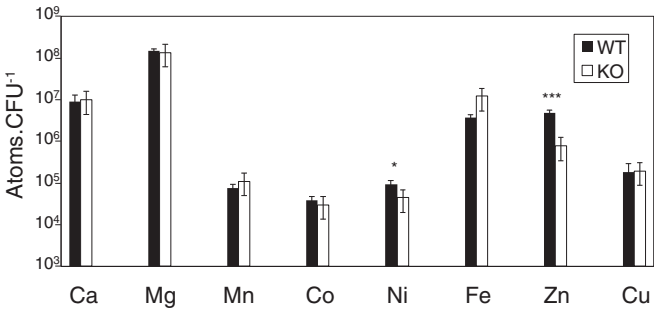


Fig. 6. Effect of *mmaA-E* mutation in metal content. Intracellular metal content of the *M. marinum* wild-type (*Left*) and Δ*mmaA-E* (*Right*) strains grown in Sauton's medium was determined by ICP-OES. Error bars, mean ± SD. *$P \leq 0.05$ and ***$P \leq 0.001$.

Conclusions. We have revealed the function of a biosynthetic gene cluster that is widespread in Actinobacteria including pathogenic mycobacteria such as *M. tuberculosis* and *M. marinum*. Using both in vivo and in vitro analyses, we discovered that the conserved five biosynthetic enzymes were capable of synthesizing a family of INLPs by using a thio-template mechanism. This unusual biosynthetic pathway starts with the activation and loading of an α,β-unsaturated fatty acid onto an ACP by a promiscuous AAL. A Michael addition of Gly to the β-position of the α,β-unsaturated fatty acyl-ACP is then promoted by a thioesterase homolog, followed by oxidation and decarboxylation, presumably catalyzed by a nonheme iron(II)-dependent oxidase, to generate a β-isonitrile fatty acyl-ACP intermediate. This isonitrile intermediate is then condensed to both amino groups of Lys promoted by an NRPS and reductively released to form a terminal alcohol product. The identified isonitrile biosynthetic pathway is distinct from the canonical mechanism in which one carbon is transferred to an amino group to form isonitrile catalyzed by an isonitrile synthase (27–30). In addition, the condensation of an acyl moiety to both amino groups of Lys, presumably promoted by a single C domain of the NRPS, is rare in nonribosomal peptide biosynthesis. Further mutagenesis study in *M. marinum* suggests that this biosynthetic gene cluster plays a role in metal transport. This study thus sheds light on a metal transport system that is critical for the virulence of pathogenic mycobacteria. Future study of native INLP metallophores in mycobacteria and their mode of action during infection may inspire new therapies to combat pathogenic mycobacteria, in particular *M. tuberculosis*, from which millions of people die every year.

Materials and Methods

Full experimental details are available in *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank M. Hadjithomas (Joint Genome Institute) for assistance with bioinformatics search of homologous gene clusters; J. Pelton (University of California, Berkeley) for assistance with NMR spectroscopic analysis, J. Cox (University of California, Berkeley) for providing plasmids for *M. marinum* genetic manipulation; M. Welch (University of California, Berkeley) for providing *M. marinum* strain M; W. Yang (University of California, Berkeley) for assisting with ICP-OES analysis; S. Stanley and K. Sogi (University of California, Berkeley) for providing assistance with *M. marinum* genetic manipulation; and S. Bauer (University of California, Berkeley) for assisting with LCMS analysis. N.C.H. is funded by the National Science Foundation Graduate Research Fellowship Program. W.Z. is a Chan Zuckerberg Biohub investigator. This research was financially supported by the Pew Scholars Program and the National Institutes of Health (Grant DP2AT009148).

1. Romero D, Traxler MF, López D, Kolter R (2011) Antibiotics as signal molecules. *Chem Rev* 111:5492–5505.
2. O'Brien J, Wright GD (2011) An ecological perspective of microbial secondary metabolism. *Curr Opin Biotechnol* 22:552–558.
3. Camilli A, Bassler BL (2006) Bacterial small-molecule signaling pathways. *Science* 311: 1113–1116.
4. Hur GH, Vickery CR, Burkart MD (2012) Explorations of catalytic domains in nonribosomal peptide synthetase enzymology. *Nat Prod Rep* 29:1074–1098.
5. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* 5:631–645.
6. Giovannini D, et al. (2012) A new Mycobacterium tuberculosis smooth colony reduces growth inside human macrophages and represses PDIM Operon gene expression. Does an heterogeneous population exist in intracellular mycobacteria? *Microb Pathog* 53:135–146.
7. Dhar N, McKinney JD (2010) Mycobacterium tuberculosis persistence mutants identified by screening in isoniazid-treated mice. *Proc Natl Acad Sci USA* 107:12275–12280.
8. Zheng H, et al. (2008) Genetic basis of virulence attenuation revealed by comparative genomic analysis of Mycobacterium tuberculosis strain H37Ra versus H37Rv. *PLoS One* 3:e2375.
9. Hotter GS, et al. (2005) Transposon mutagenesis of Mb0100 at the ppe1-nrp locus in Mycobacterium bovis disrupts phthiocerol dimycocerosate (PDIM) and glycosylphenol-PDIM biosynthesis, producing an avirulent strain with vaccine properties at least equal to those of M. bovis BCG. *J Bacteriol* 187:2267–2277.
10. Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* 100:12989–12994.
11. Cappelli G, et al. (2006) Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: Upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Res Microbiol* 157:445–455.
12. Rengarajan J, Bloom BR, Rubin EJ (2005) Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc Natl Acad Sci USA* 102:8327–8332.
13. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77–84.
14. Lamichhane G, et al. (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: Application to Mycobacterium tuberculosis. *Proc Natl Acad Sci USA* 100:7213–7218.
15. Zhang W, Ostash B, Walsh CT (2010) Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *Proc Natl Acad Sci USA* 107: 16828–16833.
16. Liu Z, Ioerger TR, Wang F, Sacchettini JC (2013) Structures of Mycobacterium tuberculosis FadD10 protein reveal a new type of adenylate-forming enzyme. *J Biol Chem* 288:18473–18483.
17. Chhabra A, et al. (2012) Nonprocessive [2 + 2]e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proc Natl Acad Sci USA* 109:5681–5686.
18. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: A web-based platform to process untargeted metabolomic data. *Anal Chem* 84:5035–5039.
19. Saghatelian A, et al. (2004) Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* 43:14332–14339.
20. Stephany RW, Bie MJAD, Drenth W (1974) C-13-Nmr and Ir study of isocyanides and some of their complexes. *Org Magn Resonance* 6:45–47.
21. Bhushan R, Brückner H (2004) Marfey's reagent for chiral amino acid analysis: A review. *Amino Acids* 27:231–247.
22. Amano SI, et al. (2011) A cryptic antibiotic triggered by monensin. *J Antibiot (Tokyo)* 64:703.
23. Tabata Y, Hatsu M, Amano S, Shimizu A, Imai S (1995) SF2768, a new isonitrile antibiotic obtained from *Streptomyces*. *Sci Rep Meiji Seika Kaisha* 34:1–9.
24. Sasaki T, et al. (1987) A new antibiotic SF2369 produced by Actinomadura. *Sci Rep Meiji Seika Kaisha* 26:10–16.
25. Wang F, Langley R, Gulten G, Wang L, Sacchettini JC (2007) Identification of a type III thioesterase reveals the function of an operon crucial for Mtb virulence. *Chem Biol* 14:543–551.
26. Huang W, Kim SJ, Liu J, Zhang W (2015) Identification of the polyketide biosynthetic machinery for the indolizidine alkaloid cyclizidine. *Org Lett* 17:5344–5347.
27. Brady SF, Clardy J (2005) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. *Angew Chem Int Ed Engl* 44:7063–7065.
28. Clarke-Pearson MF, Brady SF (2008) Paerucumarin, a new metabolite produced by the pvc gene cluster from Pseudomonas aeruginosa. *J Bacteriol* 190:6927–6930.
29. Drake EJ, Gulick AM (2008) Three-dimensional structures of Pseudomonas aeruginosa PvcA and PvcB, two proteins involved in the synthesis of 2-isocyano-6,7-dihydroxycoumarin. *J Mol Biol* 384:193–205.
30. Micallef ML, et al. (2014) Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol* 14:213.
31. Amagai K, Takaku R, Kudo F, Eguchi T (2013) A unique amino transfer mechanism for constructing the β-amino fatty acid starter unit in the biosynthesis of the macrolactam antibiotic cremimycin. *ChemBioChem* 14:1998–2006.
32. Mokhtarzadeh CC, et al. (2015) Synthesis and protonation of an encumbered iron tetraisocyanide dianion. *Inorg Chem* 54:5579–5587.
33. Malatesta L (1959) Isocyanide complexes of metals. *Progress in Inorganic Chemistry*, ed Cotton FA (Interscience Publishers, Geneva), Vol 1.
34. Ma Z, Jacobsen FE, Giedroc DP (2009) Coordination chemistry of bacterial metal transport and sensing. *Chem Rev* 109:4644–4681.
35. Remy L, et al. (2013) The Staphylococcus aureus Opp1 ABC transporter imports nickel and cobalt in zinc-depleted conditions and contributes to virulence. *Mol Microbiol* 87:730–743.
36. Ghssein G, et al. (2016) Biosynthesis of a broad-spectrum nicotianamine-like metallophore in Staphylococcus aureus. *Science* 352:1105–1109.
37. De Voss JJ, et al. (2000) The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages. *Proc Natl Acad Sci USA* 97:1252–1257.
38. Botella H, et al. (2011) Mycobacterial p(1)-type ATPases mediate resistance to zinc poisoning in human macrophages. *Cell Host Microbe* 10:248–259.
39. Wolschendorf F, et al. (2011) Copper resistance is essential for virulence of Mycobacterium tuberculosis. *Proc Natl Acad Sci USA* 108:1621–1626.
40. Chaturvedi KS, Hung CS, Crowley JR, Stapleton AE, Henderson JP (2012) The siderophore yersiniabactin binds copper to protect pathogens during infection. *Nat Chem Biol* 8:731–736.
41. Johnstone TC, Nolan EM (2015) Beyond iron: Non-classical biological functions of bacterial siderophores. *Dalton Trans* 44:6320–6339.
42. Rustad TR, et al. (2014) Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biol* 15:502.
43. Bardarov S, et al. (1997) Conditionally replicating mycobacteriophages: A system for transposon delivery to Mycobacterium tuberculosis. *Proc Natl Acad Sci USA* 94:10961–10966.