Structural and biosynthetic analysis of the fabrubactins, unusual siderophores from *Agrobacterium fabrum* strain C58

Vladimir Vinnik[1,5], Fan Zhang[2], Hyunjun Park[1,3], Taylor B. Cook[4], Kurt Throckmorton[1], Brian F. Pfleger[4], Tim S. Bugni[2], Michael G. Thomas[1*]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

[2]Pharmaceutical Sciences Division, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA

[3]Current address: CATALOG, Boston, Massachusetts 02129, USA

[4]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

[5]Current address: Department of Chemistry, University of Utah, Salt Lake City, Utah, 84112

*Correspondence: michael.thomas@wisc.edu (M.G.T)

KEY WORDS

Siderophore, nonribosomal peptide synthetase, polyketide synthase, natural product, *Agrobacterium*, monooxygenase, peroxidase, anachelin

ABSTRACT

Siderophores are iron-chelating molecules produced by microorganisms and plants to acquire exogenous iron. Siderophore biosynthetic enzymology often produces elaborate and unique molecular through unusual reactions to enable specific recognition by the producing organisms. Herein, we report the structure of two siderophore analogs from *Agrobacterium fabrum* strain C58, which we named fabrubactin (FBN) A and FBN B. Additionally, we characterized the substrate specificities of the NRPS and PKS components. The structures suggest unique Favorskii-like rearrangements of the molecular backbone that we propose are catalyzed by the flavin-dependent monooxygenase, FbnE. FBN A and B contain a 1,1-dimethyl-3-amino-1,2,3,4-tetrahydro-7,8-dihydroxy-quinolin (Dmaq) moiety previously seen only in the anachelin cyanobacterial siderophores. We provide evidence that Dmaq is derived from L-DOPA and propose a mechanism for the formation of the mature Dmaq moiety. Our bioinformatic analyses suggest that FBN A and B and the anachelins belong to a large and diverse siderophore family widespread throughout the *Rhizobium*/*Agrobacterium* group, α-proteobacteria, and cyanobacteria.

INTRODUCTION

Siderophores are high-affinity iron-chelating small molecules made by plants, fungi, and prokaryotes, that are used to acquire exogenous iron. Biosynthesis of siderophores comes at a significant metabolic cost to the producing organism, which is further exacerbated by competitors capturing the ferric-siderophore before it can be absorbed by the producer. As a result, siderophore biosynthesis is an evolutionary battleground of enzymology that leads to a high diversity of elaborate structures[1-4]. One such example are the pyoverdines, made by fluorescent pseudomonads. Apart from a characteristic dihydroxyquinoline chromophore present in all of these molecules, pyoverdine biosynthetic enzymology produces remarkable structural diversity by varying the number, identities, and cyclization of amino acids incorporated into these natural products[5]. To date there are more than 60 structurally unique pyoverdines[6]. Thus, exploring siderophore biosynthesis can facilitate the discovery of unusual natural product enzymology and provide insights into how Nature generates natural product diversity.

We have previously reported that *Agrobacterium fabrum* strain C58, formerly known as *A. tumefaciens* strain C58[7, 8], produces a siderophore assembled by a hybrid nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS)[9]. This conclusion was based on the associated biosynthetic gene cluster (BGC) encoding two PKS modules, one fatty acyl-CoA ligase module, and seven NRPS modules. This enzymology did not align with the structure of any known siderophore from *Agrobacterium* spp. or related bacteria. Interestingly, some of the enzymes encoded by this BGC have homologs

in the cyanobacterium *Anabaena* sp. strain PCC7120[9], and it was subsequently shown that the BGC from this cyanobacterium encodes enzymes that assemble a siderophore of unknown structure[10]. More recent bioinformatic analyses of BGCs in cyanobacteria suggest the BGC in PCC7120 is involved in the production of an anachelin analog[11].

In our previous study, we were able to purify the siderophore, but did not perform any structural analyses. Efforts to use biochemical characterization of the biosynthetic enzymes to guide structural analysis were thwarted by our inability to obtain soluble NRPS components when heterologously produced in *Escherichia coli*. Since our initial work, we, and others, discovered that small proteins belonging to the MbtH-like protein (MLP) superfamily dramatically influence the solubility and activity of associated NRPSs[12-14]. Genes encoding MLPs are commonly found in association with genes encoding NRPSs and hybrid NRPS/PKSs[15], as is the case with the *A. fabrum* strain C58 siderophore BGC.

Herein we report the structure of two siderophore analogs from *A. fabrum* strain C58 that we refer to as fabrubactin (FBN) A and B and the biochemical characterization of the NRPS and PKS components that produce the siderophore backbones. These siderophores contain a lipid tail at one end and are capped by a 1,1-dimethyl-3-amino-1,2,3,4-tetrahydro-7,8-dihydroxy-quinolin (Dmaq) at the other (Figure 1). This Dmaq moiety is rare in natural products and has previously only been found in three analogs of the cyanobacterial siderophore anachelin[16, 17]. We provide biochemical evidence that levodopa (L-DOPA) is a precursor for the formation of the Dmaq moiety and is produced

by a heme peroxidase. We hypothesize that this heme peroxidase is bifunctional and catalyzes L-DOPA formation and *aza*-annulation to form the Dmaq moiety. FBN A and B are unusual in that they contain two backbone quaternary carbons each with a free hydroxyl and carboxyl group. We propose these moieties are formed by two Favorskii-like rearrangements catalyzed by a flavin-dependent monooxygenase. Finally, the formation of the two analogs suggests the potential for an unusual NRPS module that first selectively activates β-Alanine (β-Ala), covalently tethering the amino acid to its thiolation domain, and then activates either another β-Ala (FBN A) or captures an activated L-Ala from a neighboring NRPS for amide bond formation (FBN B), generating a dipeptide on the thiolation domain. The structure of the FBNs and the unusual enzymology provide more support for siderophores and their biosyntheses being rich sources of novel chemistry and enzymes. Additionally, bioinformatics suggests the FBNs and anachelins are a large class of Dmaq-containing siderophores found in many bacteria.

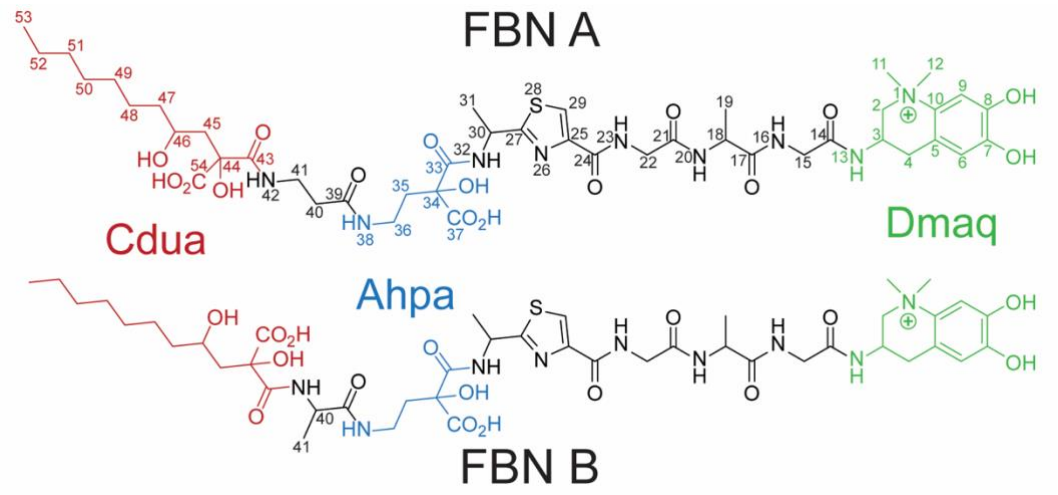


Figure 1. Structures of fabrubactin (FBN) analogs A and B. Unique partial structures 2-(2-aminoethyl)-2-hydroxy-propanedioic acid (Ahpa; shown in blue), 2-C-carboxy-2,4-dihydroxy-undecanoic acid (Cdua; shown in red), and 1,1-dimethyl-3-amino-1,2,3,4-tetrahydro-7,8-dihydroxy-quinolin (Dmaq; shown in green) are indicated. Only atoms unique to FBN B are numbered on the structure of this analog.

RESULTS AND DISCUSSION

Structure elucidation of FBN A and B. Growth conditions in Tris-based minimal medium containing the iron chelator dipyridyl were optimized for siderophore production. The siderophores were purified as described in the Experimental Procedures. Fractions containing the siderophore were identified by reactivity with chromeazurol S reagent[18, 19]. During growth and purification, a parallel culture of an *A. fabrum* strain C58 disrupted in siderophore production (an *fbnN* deletion) was used as a negative control. From eight liters of culture, five mg of FBN A and B were obtained for structural studies.

The structures of the two siderophores, FBN A and B (Figure 1), were determined using a combination of MS and NMR. The purified compounds had an *m*/z of 1008.4334 (M^+), showing the appropriate M+1, M+2, and M+3 isotopic patterns (Figure S1A). A metabolite with this *m*/z was not observed in the negative control. The molecular formula of FBN A and B was established as $C_{44}H_{66}N_9O_{16}S$ based on high-resolution electrospray ionization mass spectrometry (HR-ESI-MS) data. The two analogs are identical except for β-Ala in FBN A and Ala in FBN B corresponding to backbone atoms numbered 39-42. Therefore, FBNs will henceforth be used to refer to both analogs unless otherwise indicated. Additionally, ultra-high-resolution Fourier transform ion cyclotron resonance (FT-ICR) MS was also employed to obtain the exact molecular formula information from its isotopic fine patterns, which clearly suggested the presence of one sulfur (Figure S1B, S1C).

Analysis of the ^{13}C-^{13}C COSY data provided most of the backbone carbons of the structure (i.e., without configurational assignment) (Figure 2, Table S1, and Table S2). Analysis of the 1D NMR, COSY, and HMBC data of FBN A identified three standard amino acids including one Ala and two glycines (Gly), and one β-Ala. Similarly, two Ala and two Gly amino acids were identified in FBN B. HMBC correlations from H-29 (s, δ_H 7.88) to C-25 (δ_C 148.8) and C-27 (δ_C 174.5) suggested the presence of a 2-substituted thiazole (Thz) ring[20]. The connectivity between C-27 and C-30 and between C-30 and C-31 completed the Ala-Thz fragment (Figure 2, Table S1, and Table S2).

The HMBC correlations from the proton signals of the two N-methyl groups (H_3-11 and H_3-12) to C-2 and C-10 enabled a linkage of C-2 and C-10 via a nitrogen atom. C-7 and

C-8 were assigned as two phenolic carbons based on the chemical shift and were also confirmed by NOESY correlations between OH-8 and H-9 and between H-6 and OH-7. Therefore, one of the remaining unassigned moieties was associated with the Dmaq unit (Figure 2, Table S1, and Table S2)[16, 17]. The counterion for the Dmaq moiety was trifluoroacetate, used in the HPLC purification. The catecholate in the Dmaq moiety of anachelin is one of the functional groups involved in binding iron[17]. The presence of Dmaq in the FBNs provides structural support for their role as siderophores. Mixing a solution of $FeCl_3$ and siderophore gave a peak at 1043 *m*/z corresponding to a 1:1 complex as $[M+Fe^{3+}-3H-H_2O]^+$, determined by ESI-MS (Figure S2).

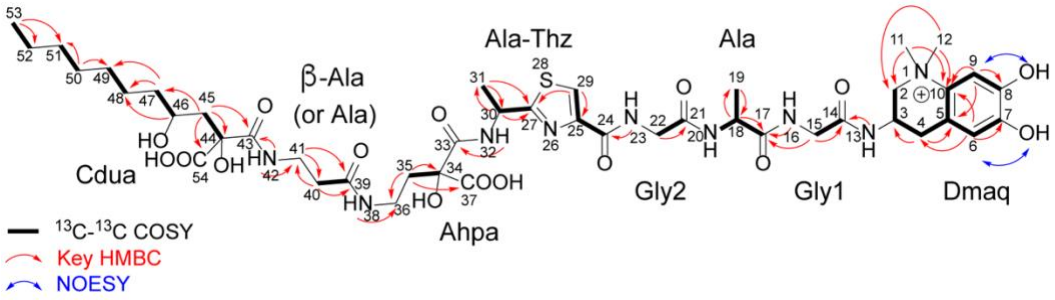


Figure 2. Select 2D NMR analysis for FBN A. 2D NMR correlations to determine FBN A (β-Ala) and FBN B (Ala) structures. NMR data is provided in Tables S1 and S2.

Assembly of the last partial structures, Ahpa and 2-C-carboxy-2,4-dihydroxy-undecanoic acid (Cdua) was accomplished by ^{1}H-^{1}H COSY and HMBC. Since both C-34 (δ_C 78.0) and C-44 (δ_C 77.6) were sp3 quaternary carbons, one hydroxy and one carboxylic acid group was attached to each to satisfy the molecular formula (Figure 2, Table S1, and Table S2). The presence of two carboxylic acid groups were further confirmed by ESI-

MS^2 and FT-MS^2 data (Figure S3A), showing two significant peaks $[M-COOH]^+$ and $[M-2*COOH]^+$ at 964.4371 *m*/z and 920.4474 *m*/z, respectively.

Attempts to confirm the connectivity of the aforementioned partial structures using FT-MS^2 experiments were conducted concurrently with the NMR-based structure elucidation. Accurate mass measurements and fine structure isotopic patterns resulted in the assignment of exact molecular formulas of fragments (Figure S3B)[21, 22]. The fragment ion at *m*/z 548.2283 could be assigned the elemental composition $C_{24}H_{34}N_7O_6S^+$, corresponding to Dmaq−Gly1−Ala−Gly2−Thz-Ala. Detection of the ion at *m*/z 764.3029 ($C_{32}H_{46}N_9O_{11}S^+$), resulted in the sequence assignment of a β-Ala−Ahpa after the Ala-Thz residue. The neutral loss between *m*/z 764.3029 and parent ion was produced through cleavage of the amide bond of the Cdua residue (Figure S3B). Observation of HMBC correlations from NH-42 and H_2-45 to C-43 suggested the connectivity between Cdua and β-Ala, thus completing the structure of the FBNs without configurational assignment.

Substrate specificity of the initiation module, FbnG. The unusual structures of the FBNs raised questions about how these siderophores are biosynthesized. We previously identified a large portion of the FBN BGC[9]. Here we extended the BGC to additional ORFs to account for the biosynthesis of necessary precursors, transcriptional regulation, FBN transport across the inner and outer membranes, and the FBN biosynthetic enzymes (Table S3). At the core of FBN biosynthesis is an NRPS/PKS hybrid megasynthase (Figure 3A) and understanding the substrate-specificities of these enzymes would provide insights into how these siderophores are assembled. Each of the adenylation (A) domain

and acyltransferase (AT) domain-containing modules were overproduced and purified in *E. coli* as either N- or C-terminal histidine-tag fusions and assayed for substrate specificity. The likely initiating module was predicted to be FbnG based on it containing an N-terminal acyl-CoA-ligase domain (AL). The purified, histidine-tagged version of this protein was assayed for fatty acid recognition using standard ATP/PP_i exchange assays. FbnG was found to preferentially activate decanoate over octanoate, nonanoate, and 3-hydroxydecanoate (Figure S4A). This result was unexpected since the FBN structures suggest that a hydroxylated fatty acid would be the initial substrate. The BGC codes for a flavin-dependent monooxygenase (FMO, FbnE) that may catalyze the hydroxylation of the thioesterified decanoate; however, as discussed below, we hypothesize that FbnE catalyzes the C2 hydroxylation of thioesterified intermediates on the PKS T domains during Favorskii-like rearrangement reactions. Currently, it is not clear if FbnE is involved in modification of decanoate while tethered to FbnG or whether FbnG has a different substrate specificity *in vivo*.

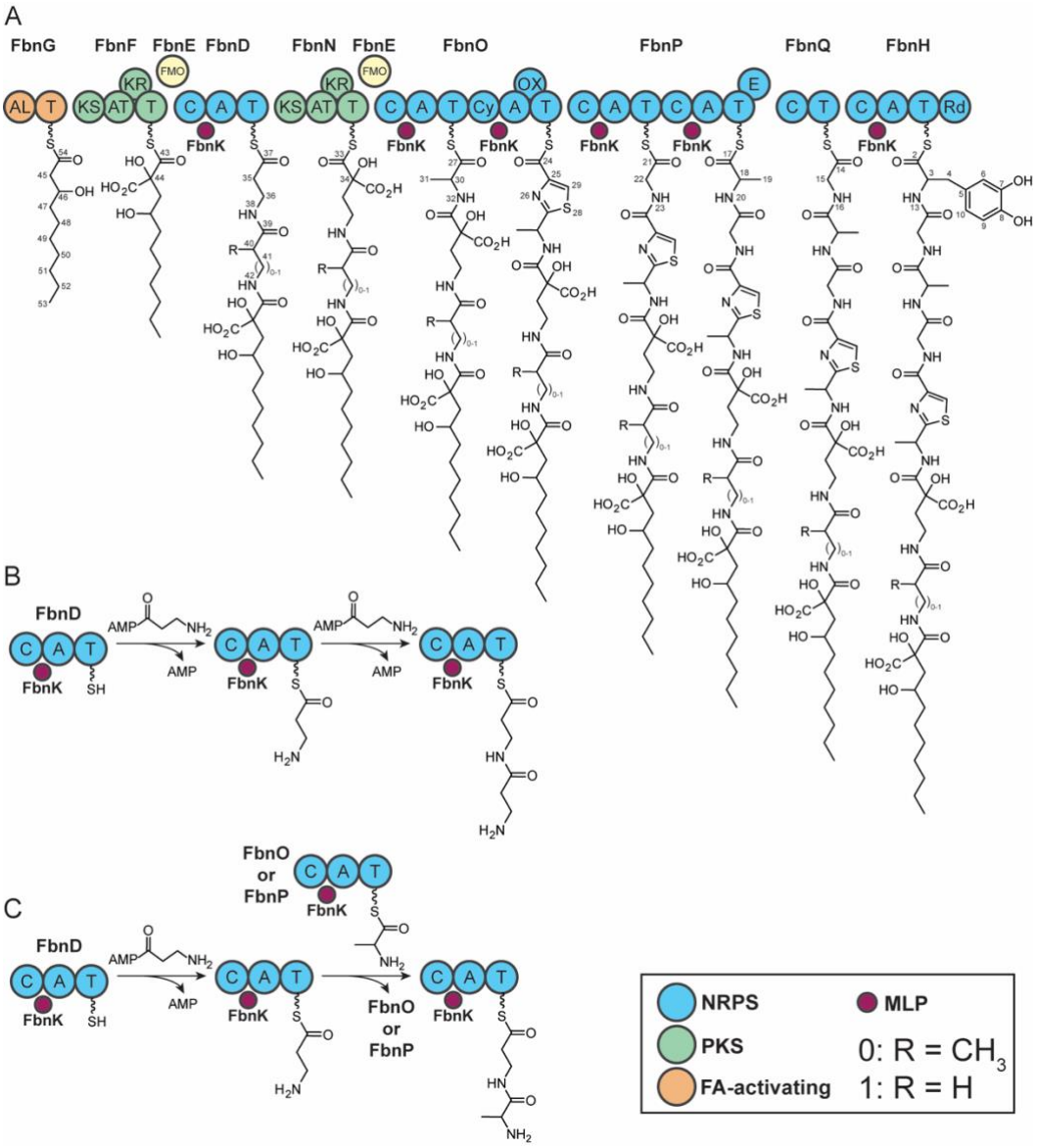


Figure 3. Biosynthetic scheme of the NRPS and PKS components. A) Hybrid polyketide synthase/nonribosomal peptide synthetase encoded by the *fbn* BGC. Substrates incorporated by each module are displayed as the numbered chemical moiety that is added to the preceding structure (left to right) and thioesterified to the protein. B) Proposed mechanism of putative iterative NRPS FbnD leading to FBN A formation. C) Proposed mechanism of capture of L-Ala thioesterified to the T domain of either FbnO or FbnP leading to FBN B formation. Domains are abbreviated as follows: AL, acyl-CoA ligase; T, thiolation KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; FMO, flavin monooxygenase; C, condensation; A. adenylation; Cy, cyclase; OX, oxidase; E, epimerase, Rd, reductase; MLP, MbtH-like protein. R is

conditional upon the presence of the β-carbon (shown in parentheses as 0 or 1): if 1, R = H (FBN A), if 0, R = CH_3 (FBN B).

Substrate specificities of the adenylation domains. All the NRPS modules were insoluble when heterologously overproduced in *E. coli* unless they were co-overproduced with the MLP FbnK; thus, we conclude every A domain is MLP-dependent. Each of the purified modules was assayed for amino acid activation. According to the solved FBN structures, the amino acids that must be accounted for in the biosynthesis are: two glycine residues, one L-cysteine (L-Cys) residue, two (FBN A) or three L-Ala (FBN B) residues, one (FBN B) or potentially two (FBN A) β-Ala residues, and likely an L-DOPA residue (Figure 1). Interestingly, the megasynthase only contains seven NRPS modules for eight amino acids, with one of the remaining modules (FbnQ) lacking an A domain (Figure 3A).

FbnH was hypothesized to be the terminal module since it contains a C-terminal reductase (Rd) domain to reductively release the product from the megasynthase. Based on the tetrahydroquinolone termini of the FBNs, FbnH was predicted to activate L-DOPA. FbnH was assayed with all 20 proteinogenic amino acids and L-DOPA, with L-DOPA being the only amino acid that stimulated significant ATP/PP_i exchange (Figure S4B); thus, FbnH is the terminal module of the megasynthase.

FbnO is a two-module NRPS subunit. Screening the A domain of the first module for amino acid activation using the standard 20 proteinogenic amino acids detected the

highest activation with L-Ala, but also activation of L-Cys, Gly, and L-Ser (Figure S4C). To determine the preferred substrate, the specificity constant (V_{max}/K_m) was determined for each amino acid (Table S4). These data support L-Ala as the preferred substrate for the first module of FbnO. The second module of FbnO contains both a cyclization (Cy) domain and an oxidase (Ox) domain. This domain organization suggested the second module of FbnO would be involved in forming the thiazole rings of the FBNs. Consistent with this hypothesis, the A domain of this module was specific for L-Cys (Figure S4D).

FbnP has two modules, each containing an A domain. We were unable to generate a construct that enabled soluble production of the N-terminal module to be produced, but we could overproduce and purify both the full-length FbnP and the C-terminal module. The C-terminal A domain was specific for L-Ala, while the full-length FbnP activated both L-Ala and Gly (Figure S4E, S4F). The presence of a C-terminal epimerase (E) domain suggests the stereochemistry of this residue in the FBNs will be D-Ala. These data are consistent with FbnP functioning immediately after FbnO to form the central region of the FBNs.

FbnD was the only A domain-containing module remaining. Of the 20 proteinogenic amino acids tested, FbnD was found to weakly activate only Gly. The BGC codes for a homolog of aspartate α-decarboxylase (FbnA) that converts L-Asp to β-Ala, and the FBNs contain β-Ala moieties. Based on these observations, it is reasonable to propose that FbnD activates β-Ala. ATP/PP_i assays using Gly, L-Ala, D-Ala, β-Ala, and β-Ala-Gly determined the A domain was specific for β-Ala activation (Figure S4G).

Based on the combined ATP/PP_i assay results and our structural analysis, it was reasonable to place FbnQ, which lacks an A domain, as the module between FbnP and the terminal module, FbnH. The structures of the FBNs suggest the penultimate NRPS module would introduce Gly. Since the N-terminal module of FbnP is the only one specific for Gly, we propose that it loads Gly onto the T domain of FbnQ *in trans* (Figure 3A).

FbnD is an unusual NRPS. The distinction between FBN A and B is the presence of a β-Ala or L-Ala linking the Cdua and Ahpa portions of the siderophores (Figure 1). As discussed below, the Ahpa moiety is likely to be derived from β-Ala and malonyl-CoA loaded onto FbnD and FbnN, respectively. The functions of all the other modules are accounted for, so the introduction of the linking β-Ala or L-Ala must come from an unusual mechanism. β-Ala may be introduced by an iterative function of FbnD. Briefly, β-Ala would be activated and tethered onto the T domain of FbnD, forming a β-Ala-*S*-FbnD intermediate. The A domain of this module would then work a second time to form a β-Ala-AMP intermediate, but since there is only a single T domain, amide bond formation would occur when the amino group of the thioesterfied β-Ala attacks the activated carbonyl of the β-Ala-AMP, thereby forming β-Ala-β-Ala-*S*-FbnD (Figure 3B). Our proposed mechanism is analogous to streptothricin biosynthesis wherein a β-lysyl polymer is formed by an iterative module[23], with the exception that two distinct A domains are used; one forms the thioesterified β-lysyl unit and a second activates the remaining β-lysyl-AMP intermediates for capture by the growing thioesterified β-lysyl polymer. Here, we propose a single A domain will perform both activities. This is a reasonable hypothesis

based on the observations that a single NRPS module can have both the thioesterified amino acid and an amino acyl-AMP intermediate present simultaneously[24, 25]. However, at this time we cannot eliminate the possibility that a second FbnD module provides the second β-Ala *in trans*.

The formation of the L-Ala bridge observed in FBN B is more challenging to address. We tested three hypotheses. First, we assessed whether FbnD was able to activate L-Ala after forming a β-Ala-*S*-FbnD intermediate. This was tested by first forming the β-Ala-*S*-FbnD intermediate then assaying whether the A domain gained the ability to activate L-Ala. The results from this experiment were negative, with L-Ala activation by β-Ala-*S*-FbnD not being statistically different than the control reactions (Figure S5). Next, we investigated whether β-Ala-*S*-FbnD was able to capture L-Ala from either L-Ala-AMP or L-Ala thioesterified to a T domain from either FbnO or FbnP (Figure 3C), the two modules that activate L-Ala. To test this we formed β-Ala-*S*-FbnD, added [^{3}H]-L-Ala, ATP, and either apo- or holo-FbnO or FbnP, and investigated whether [^{3}H]-L-Ala-β-Ala-*S*-FbnD was formed by scintillation counting the FbnD band excised from an SDS-PAGE gel. No radiolabel was detected on β-Ala-*S*-FbnD in either experiment nor in control experiments where FbnO and FbnP were absent. The lack of detectable radiolabel on β-Ala-*S*-FbnD in the control experiments also eliminated the possibility that L-Ala was activated by FbnD at some low level and then tethered to the β-Ala-*S*-FbnD. These series of negative results suggest the relationship between FbnD, FbnO, and FbnP will require further investigations to decipher how FBN A and B are assembled.

Substrate specificities of the PKSs FbnF and FbnN. The PKSs were incubated with [^{14}C-C2]-malonyl-CoA or [^{14}C-C2]-(*R,S*)-methylmalonyl-CoA and subjected to SDS-PAGE followed by phosphorimaging to check for [^{14}C]-labeling. Both FbnF and FbnN were radiolabeled when incubated with [^{14}C-C2]-malonyl-CoA but not when they were incubated with [^{14}C-C2]-(*R,S*)-methylmalonyl-CoA, indicating that malonyl-CoA is the preferred substrate of both (Figure S6).

Given the gene organization in the BGC, we predict FbnF would function after FbnG and FbnN would function after FbnD (Figure 3A). However, it is unclear how the available enzymology would account for the quaternary carbons, C-34 and C-44, or the lack of hydroxyl groups on C-35 and C-45, which would be expected due to the presence of ketoreductase (KR) domains in FbnN and FbnF. Additionally, assuming canonical NRPS-PKS enzymology, the number of backbone carbons in both Cdua and Ahpa, assembled in part by the PKSs, is one less than expected.

FbnE is a putative flavin-dependent monooxygenase. Taken together, the aforementioned data suggest the FBNs undergo two Favorskii-like rearrangements to excise two carbons, C-37 and C-54, from the backbone as carboxylic acids. We propose in each case, the rearrangement is catalyzed by the putative flavin-dependent monooxygenase (FMO), FbnE, stimulated by a double oxidation of the α-carbons C-34 and C-44 from the acyl groups introduced by FbnF and FbnN (Figure 4A, 4B). Currently, it is unclear what role the KR domains play since both the solved FBN structures and our proposed mechanism for backbone contraction by Favorskii-like rearrangements are

incompatible with established KR function. To reveal the origins of the excised carbons, we attempted ^{13}C-acetate feeding studies with C-1, C-2, and both carbons labeled; however, the addition of acetate to the growth media dramatically decreased FBN production and, as a result, the degree of ^{13}C incorporation was too low to be interpretable. Repeated attempts to find the appropriate time for ^{13}C-acetate addition were unsuccessful, with disruption of FBN production occurring whenever acetate was added to the growth medium.

Favorskii-like rearrangements are commonly observed in dinoflagellate natural products, such as the brevetoxins, and structural evidence of similar enzymology is seen in fungal aspyrones[26, 27]. In bacteria, there are two known FMOs involved in Favorskii-like rearrangements of polyketides, EncM and AmbI from enterocin and ambruticin biosynthesis, respectively[28-30]. Additionally, the structure of myxoprincomide from *Myxococcus xanthus* DK1622 suggested to us the FMO domain associated with MXAN_3779 is likely to catalyze a Favorskii-like rearrangement[31].

FbnE is a homolog of the MXAN_3779 FMO domain (43% identical), yet it is not homologous to either EncM or AmbI, which in turn are also not homologous to each other. One key difference in the reactions catalyzed by these enzymes is the fate of the excised backbone carbon. In enterocin, it is retained via intra-molecular esterification, while in ambruticin and myxoprincomide, it is lost via decarboxylation. Based on the FBN structures, we propose the reactions catalyzed by FbnE retain the backbone carbon as a

carboxylic acid thereby creating the unusual quaternary carbons C-34 and C-44 that are likely to be essential for iron chelation (Figure 4A, 4B).

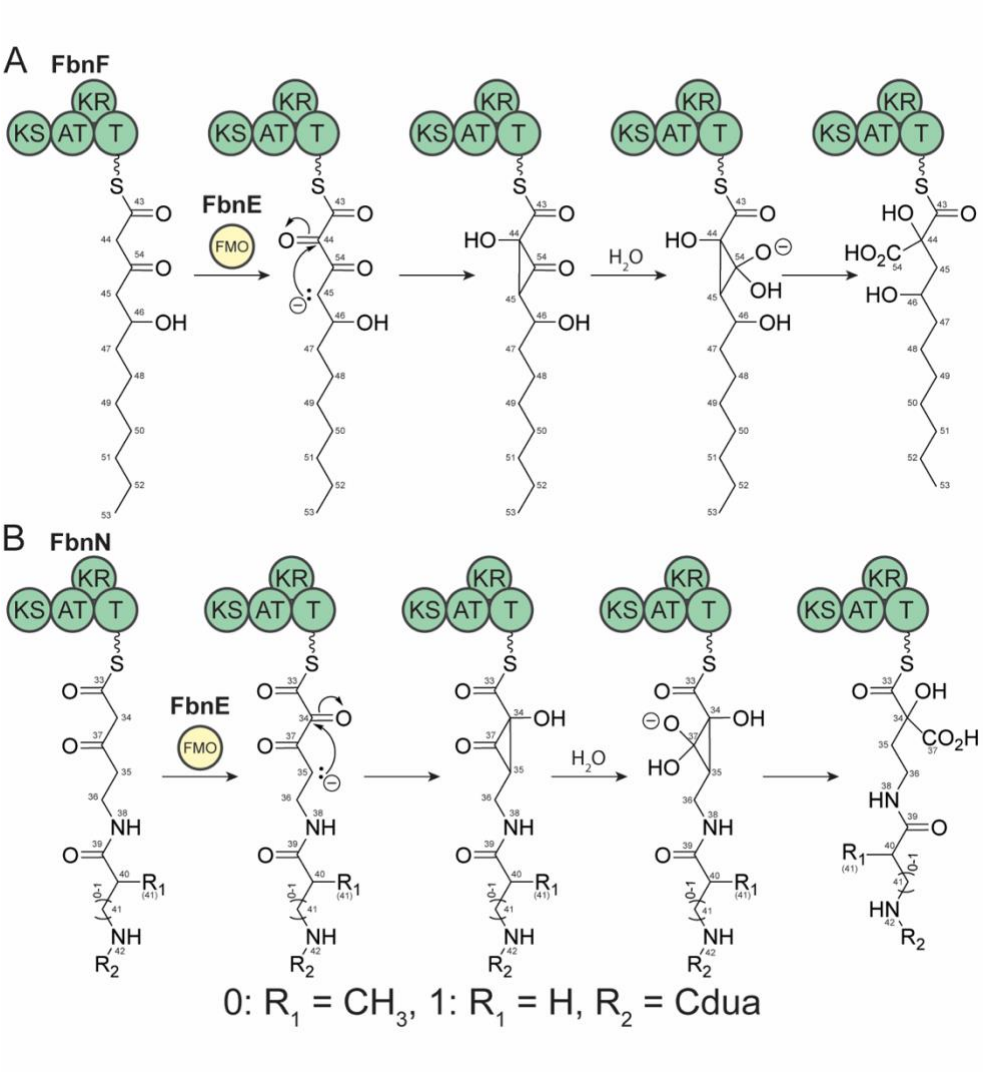


Figure 4. Proposed Favorskii rearrangement mechanisms of FBNs. Favorskii rearrangements are catalyzed by flavin monooxygenase FbnE when intermediates are thioesterified to FbnF (A) or FbnN (B). R_1 is conditional upon the presence of the β-carbon (shown in parentheses as 0 or 1): if 1, R_1 = H (FBN A); if 0, R_1 = CH_3 (FBN B).

Formation of L-DOPA. We established FbnH is selective for L-DOPA (Figure S4B). We hypothesized that L-DOPA is made from L-Tyr by FbnM prior to loading onto the megasynthase. FbnM is weakly homologous to SfmD (28% identical), a member of a recently discovered group of heme-containing peroxidases that catalyze the conversion of 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in the tetrahydroisoquinoline family of natural products[32]. Additionally, FbnM contains the HXXXC motif found in SfmD thought to be required for heme binding. Initial attempts to overproduce soluble FbnM in *E. coli* were unsuccessful. We noticed FbnM shows homology to only the C-termini half of SfmD. We hypothesized that the gene immediately upstream of *fbnM* may code for a protein that complexes with FbnM based on its genomic location and our finding of FbnL/FbnM homolog fusions in GenBank (e.g. WP_096621997.1 from *Microchaete diplosiphon*).

FbnL and FbnM were found to be soluble when the encoding genes were co-expressed and the proteins co-purified (Figure 5A). The UV-Vis spectra of the FbnL/M pair is consistent with other heme-containing enzymes (Figure 5B)[33]. Additionally, the presence of both genes on a single plasmid resulted in production of a brown pigment that was not present when the genes were expressed independently (Figure S7). Connor *et al.* reported dark brown pigmentation when overproducing Orf13, an unrelated heme peroxidase confirmed to catalyze the *ortho*-hydroxylation of L-Tyr to L-DOPA[34]. It is the subsequent polymerization of L-DOPA, through a dopaquinone intermediate, that produces the pigment[35]. Thus, it is reasonable that one role of FbnL/M is to hydroxylate L-Tyr, using H_2O_2 as the oxidant, thereby forming L-DOPA, which is loaded onto the NRPS FbnH.

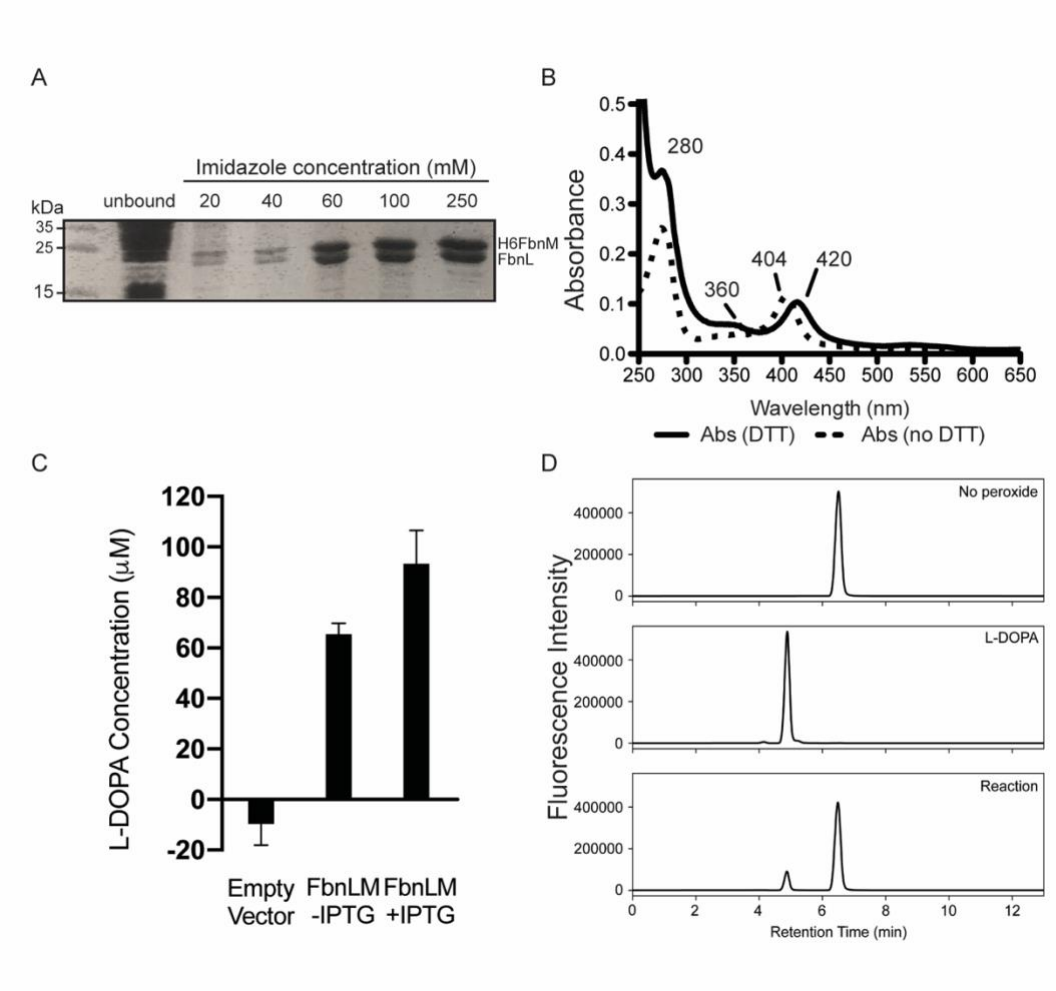


Figure 5. FbnL/M is a heme peroxidase that converts L-Tyr to L-DOPA. A) SDS-PAGE/Coomassie Blue staining of co-overproduced and Ni-NTA-purified FbnL/M. Imidazole concentrations used for protein elution are noted above each lane. B) UV-vis spectrum of 10 μg of purified FbnL/M complex in 120 μL of buffer with (solid line) or without (dashed line) the addition of 100 μM DTT. The soret band typical of heme B shifts from 404 nm to 420 nm signifying the ferric to ferrous iron conversion. C) Detection of L-DOPA in culture supernatants. Samples were induced with 100 μM IPTG where indicated. The averages of three independent reactions with standard deviations are shown. D) Three representative HPLC chromatograms of reactions (top to bottom): no hydrogen peroxide added, L-DOPA added rather than L-Tyr, and L-Tyr plus hydrogen peroxide added.

We tested this hypothesis using *in vivo* and *in vitro* approaches. We employed a colorimetric assay based on an L-DOPA dioxygenase from *Mirabilis jalapa* (mjDOD)[36] to determine whether *fbnLM*-expressing *E. coli* were exporting excess L-DOPA. MjDOD converts L-DOPA to betalamic acid, which spontaneously reacts with amino acids to form yellow pigmented betaxanthins. A greater betaxanthin signal was detected when *E. coli* BL21(DE3) contained the *fbnLM* overexpression vector compared to the empty vector control (Figure 5C). The L-DOPA concentration also increased when *fbnLM* expression was induced with IPTG. We also detected L-DOPA formation from L-Tyr by purified FbnL/M using HPLC (Figure 5D). The low level of turnover detected is not unexpected based on the well-established understanding of how heme peroxidases self-inactivate during catalysis with H_2O_2[37]. The detection of L-DOPA in these experiments confirm the heme peroxidase FbnL/M can convert L-Tyr to L-DOPA and is the likely source of L-DOPA for incorporation into the FBNs by FbnH.

Proposed biosynthetic mechanism of the Dmaq heterocycle. To form the Dmaq moiety, FbnL/M hydroxylates L-Tyr (**1**) forming L-DOPA (**2**). The L-DOPA is loaded onto the NRPS FbnH (**3**) and condensed with the upstream backbone of FBN (**4**) donated by FbnQ (Figure 6). To set up the *aza*-annulation, the FBN backbone must be reductively released from FbnH to form the semi-aldehyde (**5**) by the reductase domain of FbnH. An amino group is then transferred to the carbonyl carbon by FbnJ, a homolog of glutamate-1-semialdehyde amino transferases (HemL superfamily, COG0001) to form **6**. It has been proposed that Dmaq formation in the anachelins occurs similarly to cyclo-DOPA, i.e. through a tyrosinase-catalyzed oxidation of either L-Tyr or L-DOPA to dopaquinone,

which enables the *aza*-annulation[38]. However, a tyrosinase is not encoded in the genome of *A. fabrum* C58, thus the ring closure to form FBN must be accomplished by other means. We propose the FbnL/M pair performs a second function by catalyzing the oxidation of the catechol to a quinone (**7**). In the following non-enzymatic steps, as have been previously proposed[38], an intra-molecular Michael addition yields **8**, which tautomerizes resulting in the *aza*-annulation product (**9**). This is followed by double *N*-methylation by FbnI (pfam 13649, class I SAM-dependent methyltransferase) to form the mature Dmaq moiety (**10**), although it is possible this step precedes quinone formation (Figure 6).

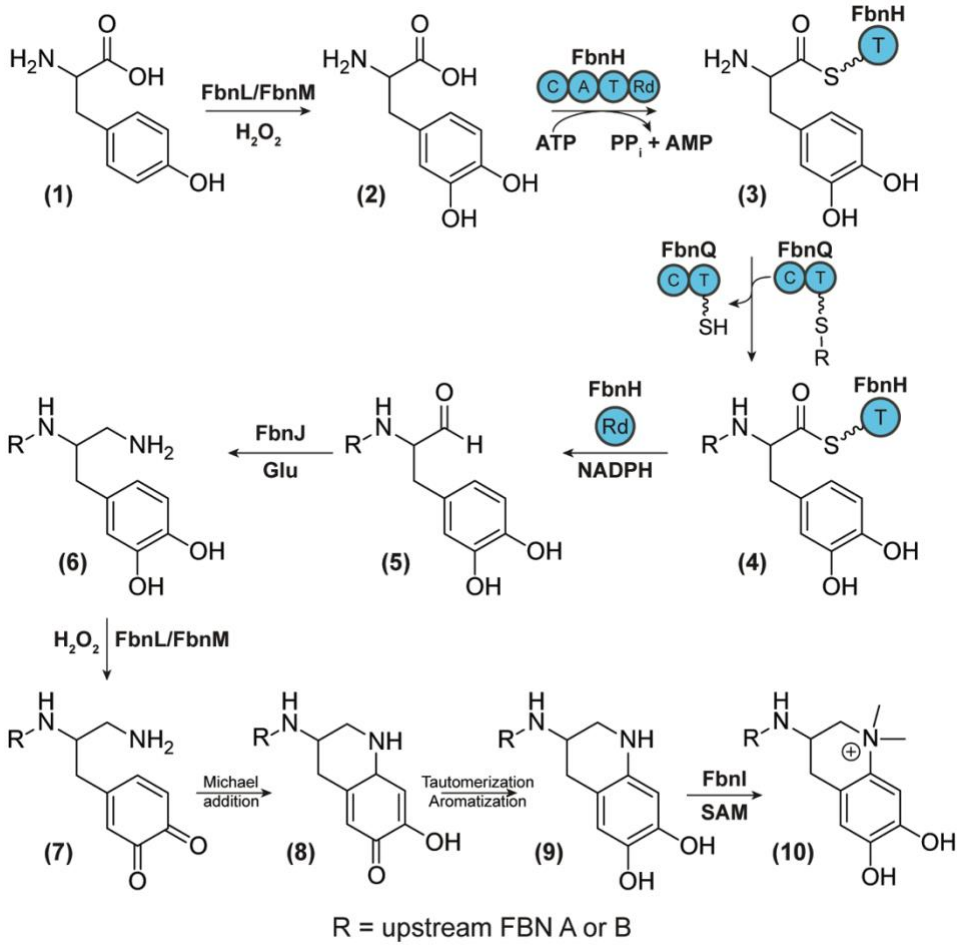


Figure 6. Proposed mechanism for the biosynthesis of the Dmaq moiety.

Fabrubactins and anachelins belong to a large siderophore family. The Dmaq moiety has only been observed in the anachelins and the FBNs. The biosynthetic formation of Dmaq is defined by the presence of FbnH-M homologs, with the occasional absence of an FbnK homolog (the MbtH-like protein). The organisms encoding homologs of this proposed Dmaq biosynthetic "cassette" are widespread in the *Rhizobium*/*Agrobacterium* group, α-proteobacteria, and cyanobacteria (Figure 7). To identify BGCs likely to be involved in the production of the Dmaq family of siderophores, we focused our initial attention on finding FbnL and FbnM homologs since these are the most distinct proteins in FBN biosynthesis. We also searched for FbnH-K homologs, NRPS and PKS enzymology, and proteins involved in siderophore transport (e.g. TonB, outer-membrane siderophore transporters). We did not fully analyze partial BGCs from incompletely assembled (meta)genomes; their includion would add at least double the number of BGCs of interest.

From this analysis we identified 40 BGCs, in three groups, that we predict to be involved in the production of a Dmaq-containing siderophore (Figure 7, Table S5). The largest group (Group I in Figure 7) includes the FBN BGC, many members of the *Rhizobium*/*Agrobacterium* group, and other α-proteobacteria. Based on the similarity of these BGCs, we anticipate all will produce FBN analogs. Furthermore, homology between NRPS components provided clear indications of where insertions or deletions of modules has occurred. Group II includes three cyanobacteria BGCs, one of which is the putative anachelin BGC from *Anabaena cylindrica* PCC7122[11] . These do not show the same

similarity in NRPS/PKS components shared in Group I and only *Nostoc* sp. '*Lobaria pulmonaria*' codes for a homolog of FbnE; thus, these will likely share the Dmaq moiety, but will be structurally distinct from each other. Group III BGCs have the necessary components for generating a Dmaq moiety, but the NRPSs associated with L-DOPA activation and reductive release of the siderophore backbone are split onto two distinct proteins. Interestingly, five of the six members of this group have a significant modification to the NRPS component for Dmaq formation. These five have the reductase domain on a separate polypeptide than the C-A-T domains, and these proteins contain an N-terminal non-heme β-hydroxylase domain (CmlA_N, pfam18456), a central UlaG-like domain (COG2220), and TubC-N-terminal NRPS docking domain (pfam18563). This domain organization suggests the Dmaq moiety itself will be structurally distinct, potentially with a β-hydroxylated L-DOPA incorporated, resulting in a hydroxyl group on the C4 of the Dmaq moiety. The sixth member of this Dmaq siderophore Group III is found in a Planctomcetaceae isolate and while it has a split FbnH homolog, it lacks the additional domains seen in the other five members of the group and likely will form a standard Dmaq moiety.

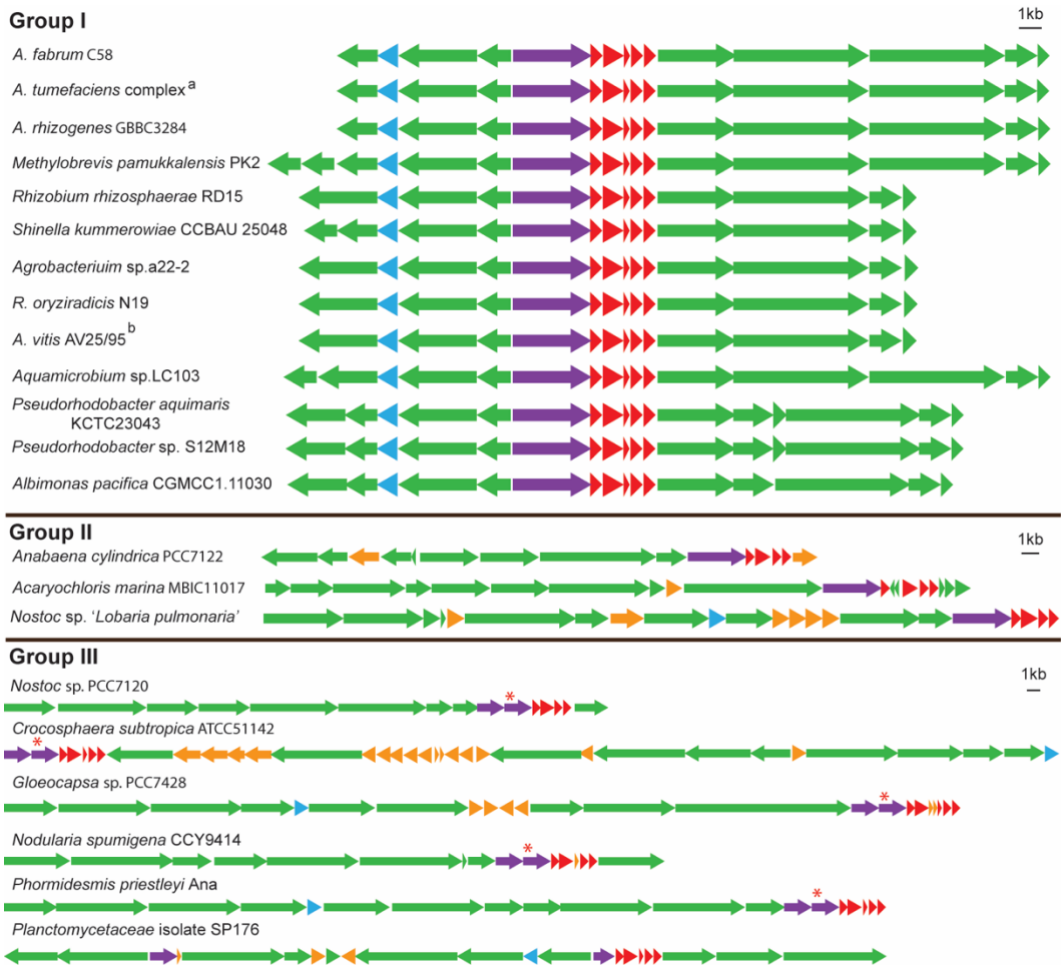


Figure 7. BGCs for the production of Dmaq-containing siderophores. Schematic representations of the BGCs proposed to be involved in the production of Dmaq-containing siderophores. Arrow colors represent genes coding for the following: Green (NPRS or PKS enzymology), Blue (Favorskii-like rearrangement FMOs), Purple (L-DOPA-recognizing NRPS involved in Dmaq formation); Purple with Red asterisk (L-DOPA-recognizing NRPS with additional domains prior to reductase domain); Red (Dmaq-forming enzymes); Orange (ORFs coding for other functions). [a], representative of BGCs in *A. fabrum* (strains 12D13, 1D132, ATCC 31749, ARqua1, and Arqua), *A. tumefaciens* (strains 15955, 15-174, CFBP5877, CFBP5499, LBA4404, F4), *Rhizobium* species. (NFIX01, NFIX02, UGM030330-04); [b], representative of BGCs in *A. vitas* (strains SZ1, AB3, AB4, 15-172, 1D1609), *A. tumefaciens* strains 15-174 and 1D1609, *R. rhizosphaerae* MH17). Identifications of the locus tags for the 5' and 3' genes for each BGC are listed in Table S5.

CONCLUSIONS

We have solved the structures of two siderophores from *A*. *fabrum* C58, FBN A and B, and characterized the specificities of the core enzymes of the NRPS/PKS megasynthase previously linked to the production of these molecules[9]. Taken together, our data suggest the involvement of unusual enzymology in the assembly of the FBNs. We identified an unusual NRPS module, FbnD, that first activates and tethers β-Ala to its T domain, and then captures a second molecule of β-Ala from itself (FBN A) or L-Ala from some other source (FBN B) to form the bridge between CduA and Ahpa components of the FBNs. We determined the FBN siderophores contain an unusual C-terminal Dmaq moeity previously only seen in the anachelin siderophores[16, 17]. We have provided biochemical evidence that a two-protein heme peroxidase (FbnL/M) catalyzes the conversion of L-Tyr to L-DOPA and propose this protein complex also catalyzes the formation of a dopaquinone intermediate the enables *aza*-annulation to form one of the rings of Dmaq. Due to its high affinity for iron, Dmaq, when synthetically conjugated to a polyethylene glycol, has been shown to bind surface-adhered metal-oxides and display protease-resistant properties[39]. This application has the potential to attach a desired natural product onto a metal-coated surface by using a combinatorial biosynthesis approach to selectively add Dmaq to the molecule. A better understanding of the Dmaq-producing suite of biosynthetic enzymes is needed to facilitate such combinatorial approaches.

The structures of the FBNs suggest the involvement of a Favorskii-like rearrangement-catalyzing FMO, FbnE, distinct from the other two known bacterial FMOs that carry out

similar chemistry[28, 29]. Favorskii rearrangements in bacterial NRPS- and PKS-made natural products are rare and little is known about the substrate selectivity of the FMOs. A deeper understanding of these enzymes could allow for unprecedented backbone-carbon manipulations of nonribosomal peptides or polyketides. We propose FbnE is responsible for the formation of the two quaternary carbons, C-34 and C-44 (Figure 4), that make the FBNs unique among bacterial polyketide-associated natural products. We have identified related hybrid NRPS/PKS-encoding BGCs in multiple rhizobacterial species that we propose also make FBN analogs containing the Dmaq moiety and quaternary backbone carbons (Figure 7). Finally, we provide bioinformatic evidence that the enzymology for Dmaq formation is found in many bacteria, suggesting there are still many members of this siderophore family yet to be discovered.

METHODS

Details of Experimental Methods are provided in the Supporting Information.

SUPPORTING INFORMATION

The supporting Information is available free of charge at http://pubs.acs.org. The contents of the supporting information include detailed experimental procedures, MS analyses, results from A and AT domain biochemical assays, NMR tables, annotation of the ORFs in the *fbn* gene cluster, related siderophore biosynthetic gene cluster locus tags, and lists of primers, strains, and plasmids used in this study.

AUTHOR CONTRIBUTIONS

V.V., H. P., F. Z., T. B. C., K. T., B. F. P, T. S. B., and M.G.T. designed the project and analyzed data. V.V., H.P., T. B. C., and M. G. T. conducted biochemical characterization of the biosynthetic enzymes. V.V. purified the siderophore. F.Z. and T.S.B determined the structure. K. T. and M. G. T. performed bioinformatic analyses. V.V., H.P., F.Z., and M.G.T. wrote the manuscript. All authors contributed to proofreading the manuscript.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (NSF 1716594 to M.G.T.) and the National Institutes of Health (GM100346 and AI065850 to M.G.T.). V.V. was supported in part by the Jerome J. Stefaniak Predoctoral Fellowship. F.Z. was supported by the National Institutes of Health grants R01GM104192, U19AI109673 and U19AI142720. We thank E. Schmidt (University of Utah) for insightful discussions regarding the potential Favorskii-like rearrangements catalyzed by the FMO FbnE. We thank P. Romero and M. Politz (UW-Madison) for discussions and conditions for the colorimetric L-DOPA assay. We thank A. Podevels for technical assistance in the cloning of *fbnD* and *fbnG*. The NMR study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41GM103399 (NIGMS), old number: P41RR002301. Equipment was purchased with funds from the University of Wisconsin-Madison, the NIH P41GM103399, S10RR02781, S10RR08438, S10RR023438, S10RR025062, S10RR029220), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA. The authors thank the Analytical Instrumentation Center (AIC) at the

University of Wisconsin-Madison School of Pharmacy for the facilities to acquire spectroscopic data, especially HRESIMS data. The authors would like to acknowledge the UW-Madison Human Proteomics Program Mass Spectrometry Facility (initially funded by the Wisconsin partnership funds) for support in obtaining mass spectrometry data (FR-ICR MS) and NIH S10OD018475 for the acquisition of ultra-high resolution mass spectrometer.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

[1] Crosa, J. H., and Walsh, C. T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria, *Microbiol. Mol. Biol. Rev. 66*, 223–249.

[2] Neilands, J. B. (1995) Siderophores: structure and function of microbial iron transport compounds, *J. Biol. Chem. 270*, 26723–26726.

[3] Winkelmann, G. (2002) Microbial siderophore-mediated transport, *Biochem. Soc. Trans. 30*, 691–696.

[4] Winkelmann, G. (1991) *CRC Handbook of Microbial Iron Chelates*, CRC Press, Boca Raton, FL.

[5] Meyer, J. M. (2000) Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent Pseudomonas species, *Arch. Microbiol. 174*, 135–142.

[6] Meyer, J. M., Gruffaz, C., Raharinosy, V., Bezverbnaya, I., Schäfer, M., and Budzikiewicz, H. (2008) Siderotyping of fluorescent Pseudomonas: molecular mass determination by mass spectrometry as a powerful pyoverdine siderotyping method, *Biometals 21*, 259–271.

[7] Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A., and Sawada, H. (2001) A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*, *Int. J. Syst. Evol. Microbiol. 51*, 89–103.

[8] Young, J. M., Kuykendall, L. D., Martinez-Romero, E., Kerr, A., and Sawada, H. (2003) Classification and nomenclature of Agrobacterium and Rhizobium – a reply to Farrand *et al*. *In. J. Syst. Evol. Microbiol. 53*, 1689–1695.

[9] Rondon, M. R., Ballering, K. S., and Thomas, M. G. (2004) Identification and analysis of a siderophore biosynthetic gene cluster from *Agrobacterium tumefaciens* C58, *Microbiology 150*, 3857–3866.

[10] Jeanjean, R., Talla, E., Latifi, A., Havaux, M., Janicki, A., and Zhang, C. C. (2008) A large gene cluster encoding peptide synthetases and polyketide synthases is involved in production of siderophores and oxidative stress response in the cyanobacterium *Anabaena* sp. strain PCC 7120, *Environ. Microbiol. 10*, 2574–2585.

[11] Calteau, A., Fewer, D. P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C. A., Sivonen, K., Piel, J., and Gugger, M. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria, *BMC Genomics 15*, 977.

[12] Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases, *Biochemistry 49*, 8815–8817.

[13] Imker, H. J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C. T. (2010) N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF, *Chem. Biol. 17*, 1077-1083.

[14] Zhang, W., Heemstra, J. R., Walsh, C. T., and Imker, H. J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins, *Biochemistry 49*, 9946–9947.

[15] Baltz, R. H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery, *J. Ind. Microbiol. Biotechnol. 38*, 1747–1760.

[16] Beiderbeck, H., Taraz, K., Budzikiewicz, H., and Walsby, A. E. (2000) Anachelin, the siderophore of the cyanobacterium *Anabaena cylindrica* CCAP 1403/2A, *Z. Naturforsch. C. 55*, 681–687.

[17] Itou, Y., Okada, S., and Murakami, M. (2001) Two structural isomeric siderophores from the freshwater cyanobacterium *Anabaena cylindrica* (NIES-19), *Tetrahedron 57*, 9093–9099.

[18] Alexander, D. B., and Zuberer, D. A. (1991) Use of the chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria, *Biol. Fertil. Soils 12*, 39-45.

[19] Schwyn, B., and Neilands, J. B. (1987) Universal chemical assay for the detection and determination of siderophores, *Anal. Biochem. 160*, 47–56.

[20] Dalisay, D. S., Rogers, E. W., Edison, A. S., and Molinski, T. F. (2009) Structure elucidation at the nanomole scale. 1. Trisoxazole macrolides and thiazole-containing cyclic peptides from the nudibranch *Hexabranchus sanguineus*, *J. Nat. Prod. 72*, 732–738.

[21] McDonald, L. A., Barbieri, L. R., Carter, G. T., Kruppa, G., Feng, X., Lotvin, J. A., and Siegel, M. M. (2003) FTMS structure elucidation of natural products: Application to muraymycin antibiotics using ESI Multi-CHEF SORI-CID FTMS^n, the top-down/bottom-up approach, and HPLC ESI capillary-skimmer CID FTMS, *Anal. Chem. 75*, 2730–2739.

[22] Laird, D. W., LaBarbera, D. V., Feng, X., Bugni, T. S., Harper, M. K., and Ireland, C. M. (2007) Halogenated cyclic peptides isolated from the sponge *Corticium* sp., *J. Nat. Prod. 70*, 741–746.

[23] Maruyama, C., Toyoda, J., Kato, Y., Izumikawa, M., Takagi, M., Shin-ya, K., Katano, H., Utagawa, T., and Hamano, Y. (2012) A stand-alone adenylation domain forms amide bonds in streptothricin biosynthesis. *Nat. Chem. Biol. 8*, 791–797.

[24] Gevers, W., Kleinkauf, H., and Lipmann, F. (1969) Peptidyl transfers in gramicidin S bisoynthesis from enzyme-bound thioester intermediates, *Proc. Natl. Acad. Sci. USA 63*, 1335-1342.

[25] Kleinkauf, H., Gevers, W., and Lipmann, F. (1969) Interrelation between activation and polymerization in gramicidin S biosynthesis, *Proc. Natl. Acad. Sci. USA 62*, 226–233.

[26] Simpson, T. J., and Holker, J. S. E. (1975) The biosynthesis of a pyrone metabolite of *aspergillus melleus* an application of long-range ^{13}C-^{13}C coupling constants, *Tetrahedron letters 16*, 4693–4696.

[27] Lee, M. S., Repeta, D. J., Nakanishi, K., and Zagorski, M. G. (1986) Biosynthetic origins and assignments of carbon 13 NMR peaks of brevetoxin B, *J. Am. Chem. Soc. 108*, 7855–7856.
[28] Julien, B., Tian, Z. Q., Reid, R., and Reeves, C. D. (2006) Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis, *Chem. Biol. 13*, 1277–1286.
[29] Teufel, R., Miyanaga, A., Michaudel, Q., Stull, F., Louie, G., Noel, J. P., Baran, P. S., Palfey, B., and Moore, B. S. (2013) Flavin-mediated dual oxidation controls an enzymatic Favorskii-type rearrangement, *Nature 503*, 552–556.
[30] Piel, J., Hoang, K., and Moore, B. (2000) Natural metabolic diversity encoded by the enterocin biosynthesis gene cluster, *J. Am. Chem. Soc. 122*, 5415–5416.
[31] Cortina, N. S., Krug, D., Plaza, A., Revermann, O., and Müller, R. (2012) Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome, *Angew. Chem. Int. Ed. Engl. 51*, 811–816.
[32] Tang, M. C., Fu, C. Y., and Tang, G. L. (2012) Characterization of SfmD as a Heme peroxidase that catalyzes the regioselective hydroxylation of 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in saframycin A biosynthesis, *J. Biol. Chem. 287*, 5112–5121.
[33] Sugano, Y. (2009) DyP-type peroxidases comprise a novel heme peroxidase family, *Cell. Mol. Life Sci. 66*, 1387–1403.
[34] Connor, K. L., Colabroy, K. L., and Gerratana, B. (2011) A heme peroxidase with a functional role as an L-tyrosine hydroxylase in the biosynthesis of anthramycin, *Biochemistry 50*, 8926–8936.
[35] Morris, F. (1950) Non-enzymatic oxidation of tyrosine and dopa, *Proc. Natl. Acad. Sci. USA 36*, 606–611.
[36] Savitskaya, J., Protzko, R. J., Li, F. Z., Arkin, A. P., and Dueber, J. E. (2019) Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production, *Sci. Rep. 9*, 5815.
[37] Valderrama, B., Ayala, M., and Vazquez-Duhalt, R. (2002) Suicide inactivation of peroxidases and the challenge of engineering more robust enzymes, *Chem. Biol. 9*, 555–565.
[38] Gademann, K. (2005) Mechanistic studies on the tyrosinase-catalyzed formation of the anachelin chromophore, *ChemBioChem 6*, 913–919.
[39] Zurcher, S., Wackerlin, D., Bethuel, Y., Malisova, B., Textor, M., Tosatti, S., and Gademann, K. (2006) Biomimetic surface modifications based on the cyanobacterial iron chelator anachelin, *J. Am. Chem. Soc. 128*, 1064–1065.

SUPPORTING INFORMATION

Structural and biosynthetic analysis of the fabrubactins, unusual siderophores from *Agrobacterium fabrum* strain C58

Vladimir Vinnik[1,5], Fan Zhang[2], Hyunjun Park[1,3], Taylor B. Cook[4], Kurt Throckmorton[1], Brian F. Pfleger[4], Tim S. Bugni[2], Michael G. Thomas[1*]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

[2]Pharmaceutical Sciences Division, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA

[3]Current address: CATALOG, Boston, Massachusetts 02129, USA

[4]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

[5]Current address: Department of Chemistry, University of Utah, Salt Lake City, Utah, 84112

*Correspondence: michael.thomas@wisc.edu (M.G.T)

TABLE OF CONTENTS:

Experimental Procedures:

Figures:

Tables:

EXPERIMENTAL METHODS

Plasmid and bacterial strain construction. The NRPS, PKS, and acyl-CoA ligase domains were cloned into the overexpression vector pET28b (EMD chemicals) and the MLP was cloned into the vector pACYCDuet-1 (Novagen) using standard PCR-based methods as previously described[1]. The primers used for PCR amplification (Table S6) were designed so that overexpression from pET28b produced a protein with a hexa-histidine tag at the N-terminus except for FbnP which had a C-terminal tag. The strain *A. fabrum* C58 Δ*fbnN* was constructed with a pK18mobsacB plasmid using a sucrose counter-selection. Complementation of the Δ*fbnN* strain was done with pBBR1MCS-2-*fbnN*[2, 3].

Overproduction and purification of NPRS and PKS components. All proteins were overproduced in *Escherichia coli* BL21(DE3)*ybdZ*::*aac(3)-IV* and purified using nickel chelate chromatography, as previously described[1, 4]. We co-produced the NRPS components with the MLP FbnK. Some constructs were extended to include other domains flanking the A domains: the A-T pair of FbnD, the entire protein FbnH, the condensation (C)-A1-T1 domains of FbnO module one, the cyclization (Cy)-A2-oxidation (Ox)-T2 domains of FbnO module two, the full protein FbnP, and the C-A2-T2 of FbnP module two. The PKSs, encoded by *fbnF* and *fbnN*, were overproduced as an entire module.

ATP-[^{32}P]PP$_i$ exchange assays to determine A domain specificity. Amino acid-dependent ATP-[^{32}P]PP$_i$ exchange assays were done as previously described[1, 4]. The purified proteins FbnO and FbnH were assayed for ability to active the 20 proteinogenic amino acids and FbnH was also assayed with L-DOPA. FbnD was assayed with Gly, D-Ala, L-Ala, β-Ala, and β-Ala-Gly. FbnP proteins were assayed with four pools of five amino acids (Pool 1: SATPG, Pool 2: VLIMC, Pool 3: NDEQF, Pool 4: YWHKR) and then with the individual amino acids from Pool 1. FbnG was assayed with octanoate, nonanoate, decanoate, and 3-hydroxydecanoate.

ATP-[^{32}P]PP$_i$ exchange assays to determine A domain specificity of FbnD-*S*-β-Ala. To obtain holo-FbnD, FbnD was overproduced and purified from BL21(DE3) cells that overproduced Sfp (strain BAP1). FbnD-*S*-β-Ala was formed in a reaction containing 75 mM Tris-HCl (pH 7.5 at 25 ºC), 10 mM $MgCl_2$, 1 mM TCEP, 3.5 mM ATP, 2 μM β-Ala and 5 μM holo-FbnD. This reaction was incubated for 1 hour at 25 ºC. Ten-microliters of this reaction was added to standard ATP-[^{32}P]PP$_i$ exchange assays described above where the reactions had 1 mM of β-Ala, L-Ala, or just water added to the reaction. Repeating the assays with FbnD overproduced and purified from BAP1 or BL21(DE3) strains and then treated with Sfp and CoA to form holo-T domains as previously described[5] did not have any impact on the observed results.

Radioactive assays of FbnF and FbnN with malonyl-CoA or (2*R,S*)-methylmalonyl-CoA. Recognition of either malonyl-CoA or (2*R*,*S*)-methylmalonyl-CoA by the AT domains of FbnF and FbnN was tested as previously described[6].

L-DOPA Dioxygenase Assay. Supernatants from cultures of *E. coli* BL21(DE3) overexpressing *fbnLM* were analyzed with a colorimetric and fluorometric L-DOPA dioxygenase (DOD) assay. Pre-cultures were prepared by growing *E. coli* BL21(DE3) containing pTEV-*fbnLfbnM* or pTEV5 overnight at 37 °C. Fresh culture tubes containing 5 mL LB with carbenicillin (100 ug mL^{-1}) were inoculated to OD_{600nm}=0.01. Cultures were incubated at 30 °C while shaking. Once the OD_{600} reached ~0.5, FbnLM overexpression was induced with 100 µM IPTG. The cultures grew for another 5 hr, and at the end of incubation cells were removed from the cultures by centrifugation and the supernatants were sterilized by filtration.

The 5X DOD reaction buffer was prepared by mixing 5 mL 10X DOD buffer, 3 mL water, 0.088 g L-ascorbic acid, and 50 µL 100 mM $FeCl_2$. The pH was adjusted to 6.0 using 5 M NaOH, and water was added to a final volume of 9.5 mL. 10X DOD buffer contains 500 mM morpholinoethanesulfonic acid (MES) monohydrate, 1 M NaCl, 1 M tricine, and 100 mM glycine. The final 5X DOD reaction master mix was prepared by mixing 237.5 µL of the 5X DOD reaction buffer with 12.5 µL of 2.5 mg mL^{-1} 6xHis-MBP-L-DOPA dioxygenase in 50% (v/v) glycerol with storage buffer. The storage buffer contains 250 mM NaCl and 10 mM Tric-Cl.

Every sample was assayed by mixing 80 µL with 20 µL 5X DOD reaction master mix and incubating at 30 °C for up to 16 hr. Reactions were monitored by measuring absorbance at 420 nm and 470 nm and fluorescence (excitation: 470 nm, emission: 515 nm). A standard curve was prepared by adding 1000, 500, 250, 100, 50, and 0 µM L-DOPA to supernatant from *E. coli* BL21(DE3) containing pTEV5.

***In vitro* L-DOPA formation assays.** Fifty-microliter reactions contained 100 mM Tris-HCl (pH 7.5 at 25 ºC), 100 mM NaCl, 20 mM H_2O_2, 500 μM L-Tyr, 2 mM dithiothreitol, 5 μg FbnL/M. Reactions were run for 30 min and then stopped by the addition of 15 μL of 20% (w/v) trichloroacetic acid. *In vitro* assays were analyzed by HPLC with a Shimadzu HPLC system (Shimadzu Co., Columbia, MD, USA) equipped with a fluorescence detector. HPLC separations were performed using a Luna C18 column (250 x 4.6 mm; 5 μm particle size). The mobile phase was 2.4% acetonitrile and 0.08% formic acid in water, the flow rate was 1.0 mL min^{-1}, and the injection volume was 2 μL. L-DOPA and L-tyrosine were detected using the fluorescence detector (excitation: 281 nm, emission: 314 nm).

Purification of siderophores FBN A and B. Wild-type *A. fabrum* C58 was grown in yeast extract mannitol (YEM) media[7] (mannitol 10 g, yeast extract 1 g, K_2HPO_4 0.5 g, $CaCl_2 \cdot H_2O$ 0.2 g, NaCl 0.2 g, and $MgSO_4 \cdot 7H_2O$ 0.2 g L^{-1}), with 100 μM dipyridyl (Sigma) for 2 days at 28 °C shaking at 200 RPM. Turbid culture (1.5 mL), twice washed with tris minimal media adapted for Agrobacteria[8] (NaCl 5.8 g, KCl 3.7 g, $CaCl_2 \cdot 2H_2O$ 0.15 g, $MgCl_2 \cdot 6H_2O$ 0.1 g, NH_4Cl 1.1 g, Na_2SO_4 0.142 g, KH_2PO_4 0.272 g, glucose 4 g, $MnSO_4$ 1 μM, and tris base 12.1 g L^{-1}), was used to inoculate 1 L tris minimal media with 20 μM dipyridyl. Eight 1 L cultures were incubated in 2.8 L glass baffled flasks at 28 °C while shaking at 200 RPM for 6-8 days, depending on presence of siderophore as determined by CAS assay[9]. Cells were pelleted by centrifugation and the supernatant was filtered by vacuum through a 0.2 μm filter (Millipore).

Filtered supernatant was combined with washed HP-20 resin (Sigma) (~70 g L^{-1} of supernatant) and shaken for 4 hr at 25 °C. The HP-20 was then retained by filtration using Miracloth (Millipore), washed with H_2O and extracted with acetone while stirring for 1 hr. The acetone was filtered by gravity through a milk filter and HP-20 was washed with additional acetone. Acetone was removed *in vacuo* until a small volume remained, then 15 g of washed HP-20SS were added, contents were mixed, and the remaining acetone was removed. HP-20SS was further dried overnight under high vacuum. Organic compounds bound to the HP-20SS were eluted by column chromatography with a step gradient containing 10, 25, 50, 75, and 100% MeOH in H_2O.

The fractions were analyzed for the compound of interest by analytical RP HPLC-MS (5 min isocratic development at 100% H_2O, 50 min gradient at 0.5 mL min^{-1} 10/90% to 33/67% ACN-H_2O containing 0.05% trifluoroacetic acid, and 5 min isocratic development at 100% ACN) using a Phenomenex Luna C18 column (250 x 10 mm, 5 µm) on a Shimadzu Prominence HPLC system and a Micromass LCT mass spectrometer. The elution of metabolites was monitored at 254 nm and 286 nm. The 50% fraction contained several related compounds including the putative compound of interest (t_R 46.5 min) which was further purified by semi-preparative RP HPLC fractionation using a Phenomenex Luna C18 column (250 x 10 mm, 5 µM) and the aforementioned separation conditions (flow rate was adjusted to 4 mL min^{-1}) to yield FBN A and FBN B (5.0 mg). These processes were repeated for the supernatant from a culture of a mutant strain containing a deletion in the gene *fbnN* and the comparable fractions were analyzed by RP HPLC-MS.

General Experimental Procedures for Structure elucidation. Optical rotations were measured on a Perkin–Elmer 241 Polarimeter. UV spectra were recorded on an Aminco/OLIS UV-Vis Spectrophotometer. IR spectra were measured with a Bruker Equinox 55/S FT–IR Spectrophotometer. 1D and 2D NMR spectra were obtained in DMSO with a Bruker Avance 600 MHz spectrometer with $^{1}H\{^{13}C/^{15}N\}$ cryoprobe and 500 MHz spectrometer with $^{13}C/^{15}N\{^{1}H\}$ cryoprobe. HR-ESI-MS and ESI-MS^2 data were acquired with a Bruker MaXis 4G QTOF mass spectrometer. FT-ICR MS and FT-ICR MS^2 data were acquired with a Bruker SolariX XR^{TM} FT-ICR mass spectrometer, which is able to reveal the fine structure in isotopic patterns that are specific to the exact molecular formulae of the detected compounds. ESIMS of the Fe^{3+} complex of FBN A was acquired with a Waters Micromass LCT TOF mass spectrometer.

FBN A and B: colorless oil; $[\alpha]^{25}_{D}$ −20 (c 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 244 (3.29) nm, 289 (3.11) nm, 304 (2.75) nm; IR (ATR) u_{max} 3416, 2927, 2856, 1701, 1635, 1613, 1566, 1476, 1459, 1407, 1372, 1354, 1303, 1267, 1228, 1195, 1134, 1088, 1039, 1021, 1007 cm^{-1}; ^{1}H and ^{13}C NMR (See Table S1, S2); HRESIMS $[M]^+$ *m/z* 1008.4348 (calculated for $C_{44}H_{66}N_9O_{16}S^+$, 1008.4343).

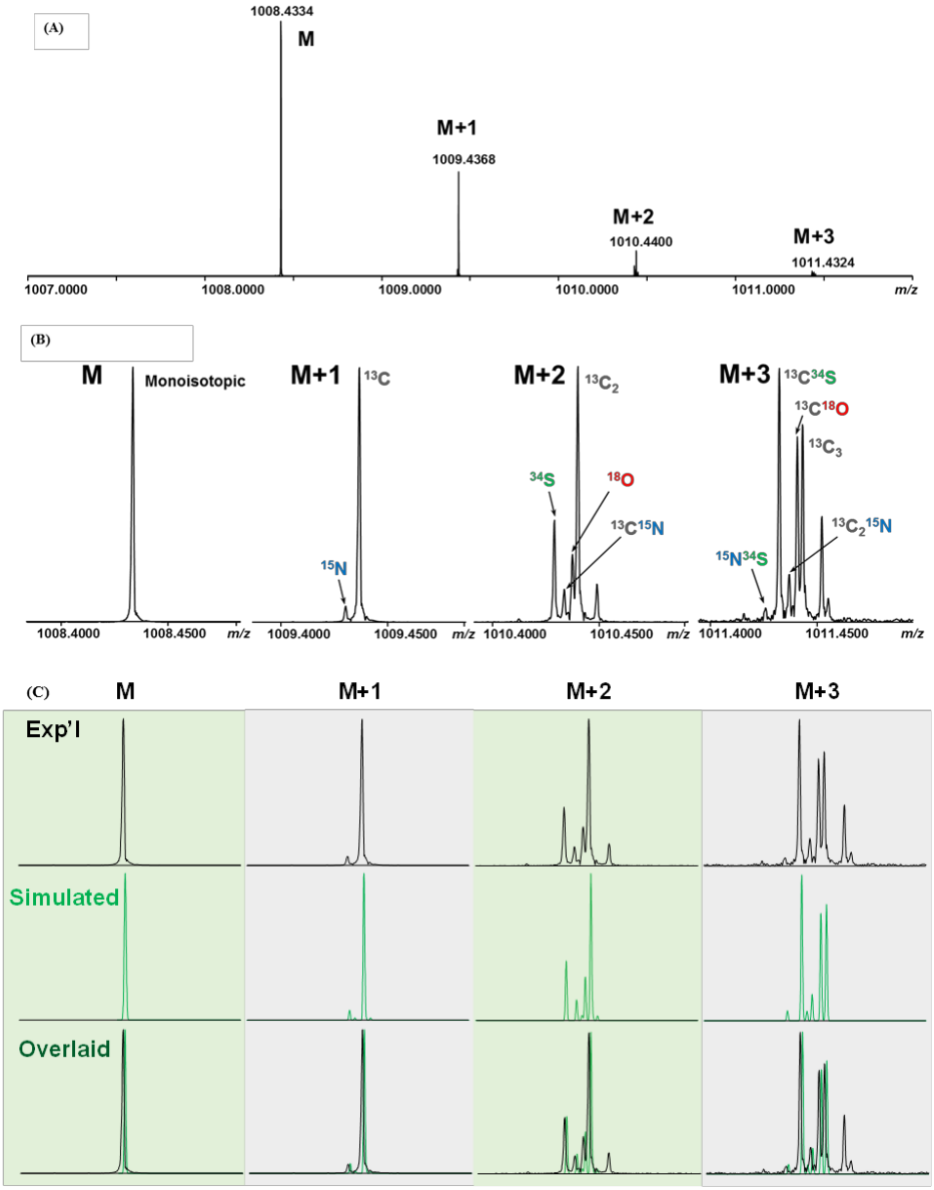


Figure S1 FT-ICR MS of FBN A. (A) The ^{13}C isotopic envelope of the singly charged peptide and the expanded segments. (B) The isotopic fine structure of the M+1, M+2, and M+3 isotopic patterns. (C) Comparison between experimental and theoretical isotopic fine structure distributions.

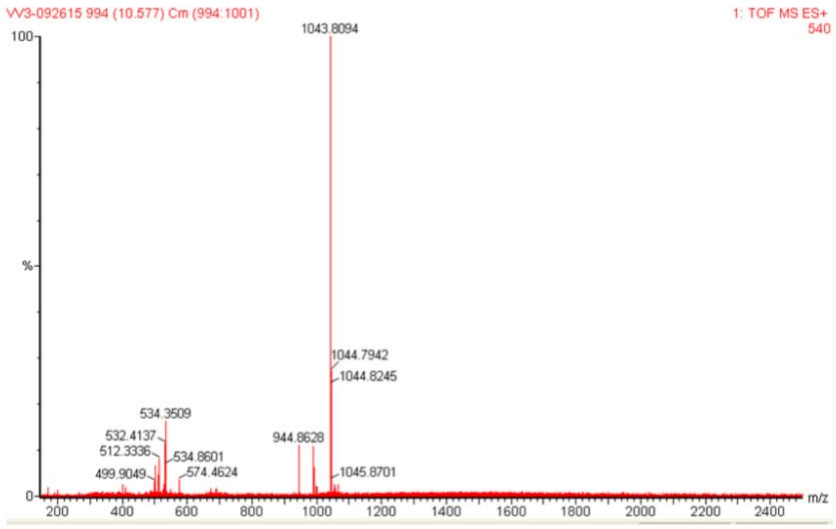
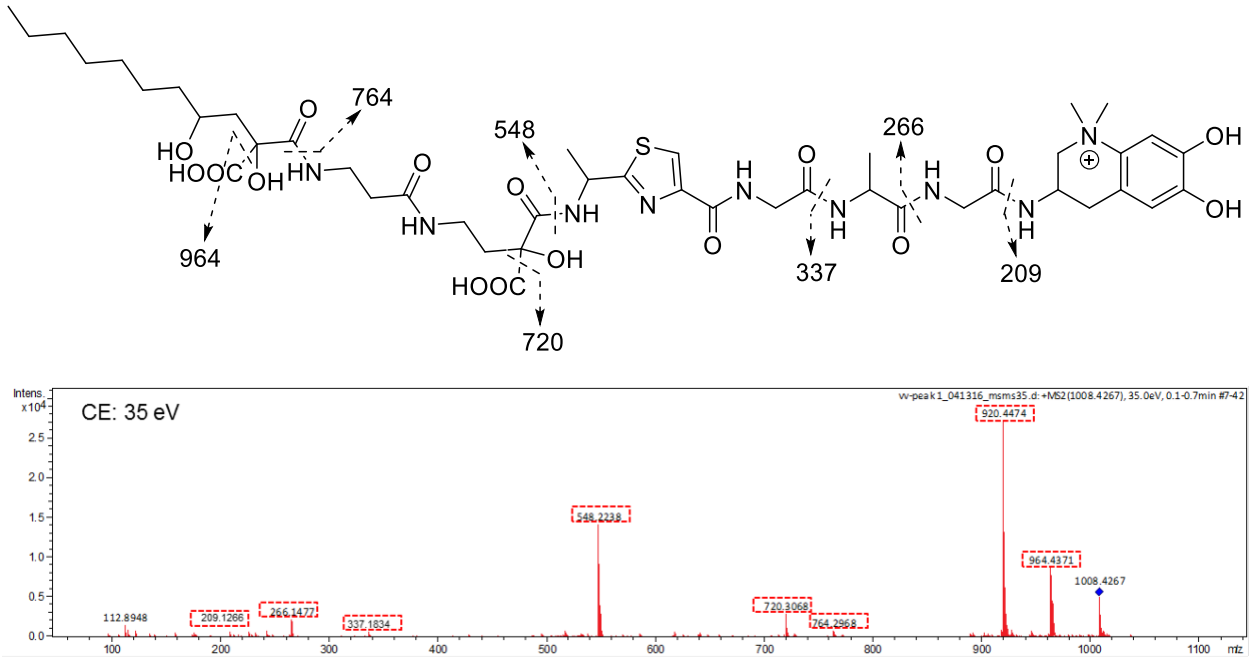


Figure S2 ESI-MS of FBN iron complex formation. ESI-MS showing *m*/z shift of the FBNs when mixed with $FeCl_3$ to produce a 1:1 complex as $M+Fe^{3+}-3H-H_2O]^+$.

A



B

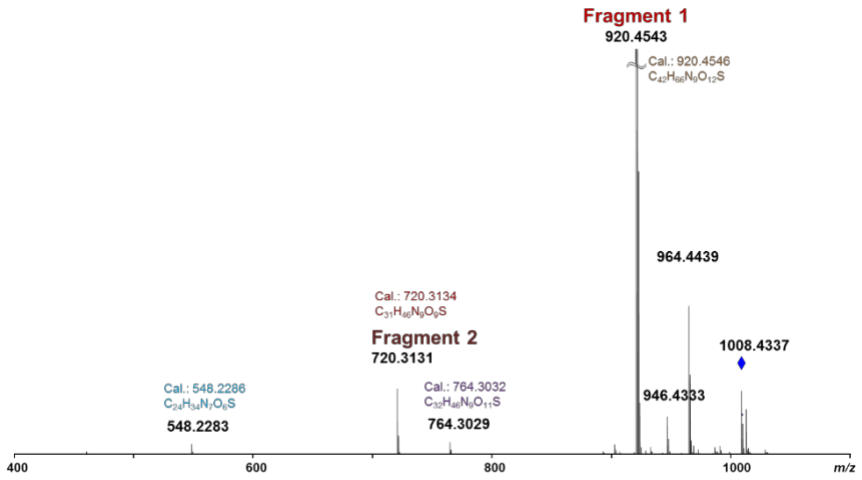


Figure S3 Characteristic fragments in the ESI-MS^2 spectrum of FBN A and FBN B. A) Fragmentation of FBN A and B observed in the ESI-MS^2 analysis. B) Fragments 1 and 2 indicate those for which an exact molecular formula was assigned based on fine structure isotopic patterns (data not shown).

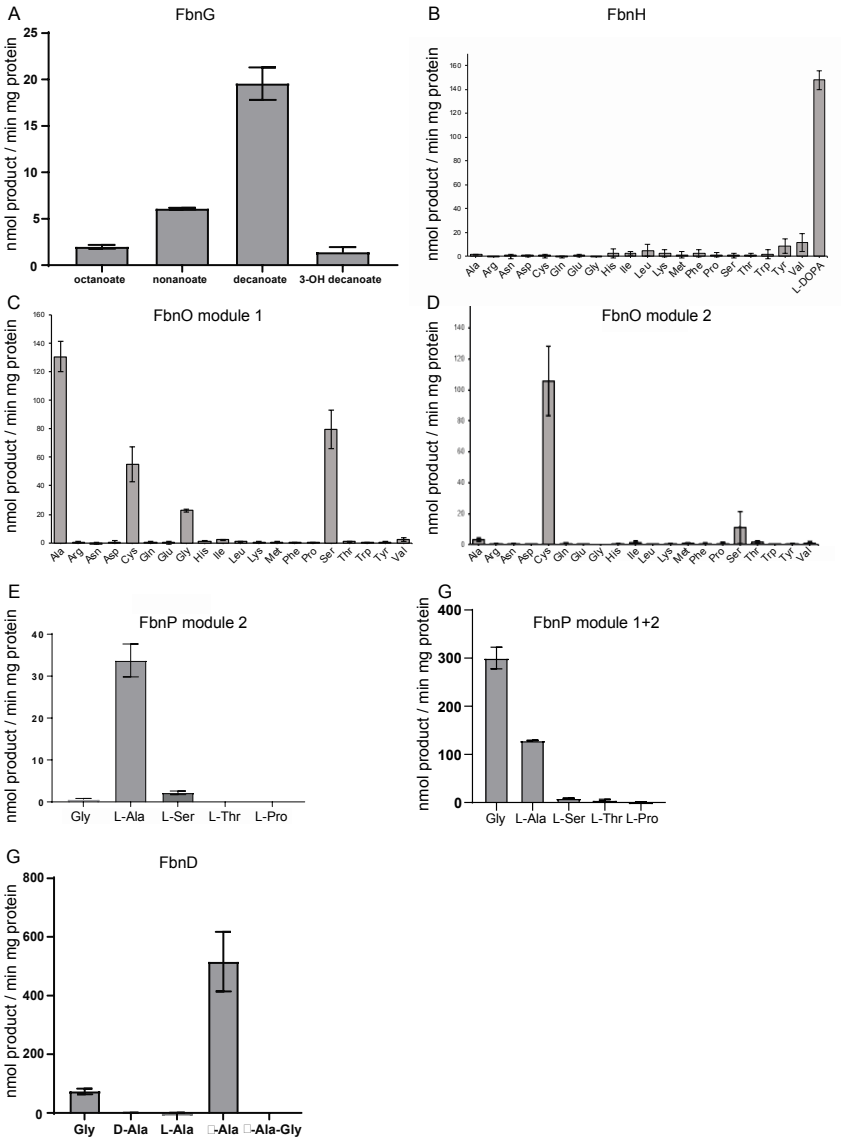


Figure S4 Substrate specificity of NRPS A domains and Acyl-CoA ligase domain of FbnG. Proteins were assayed for their ability to activate select fatty acids (A) and select amino acids (B – G) as determined by ATP/PP_i exchange assay.

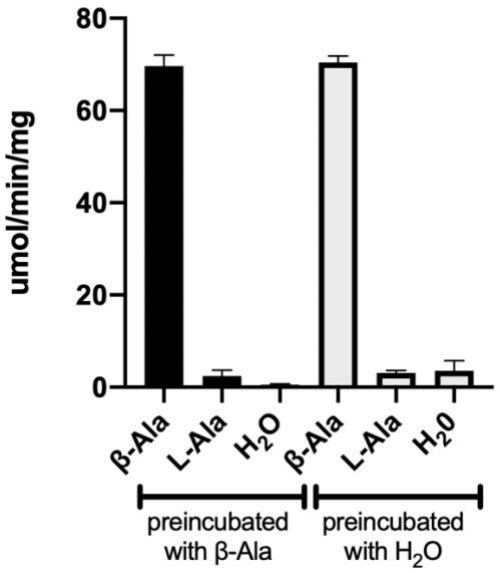


Figure S5 Evaluating FbnD activation of β-Ala and L-Ala after FbnD-*S*-β-Ala formation. ATP/PPi exchange assays monitoring β-Ala and L-Ala activation by FbnD preincubated with substoichiometric amounts of β-Ala or no β-Ala prior to the start of the assay.

Figure S6 Substrate specificities of the AT domains of FbnF and FbnN. The left side of each panel shows the SDS-PAGE/Coomassie blue analysis of the reaction mixtures; the right side of each panel shows the corresponding phosphorimage. Plus or minus signs indicate the presence or absence, respectively, of the radioactive substrate. FbnF (A) and FbnN (B) are shown to be radiolabeled only in the presence of malonyl-CoA. M-CoA, [^{14}C-C2]-malonyl-CoA; MM-CoA, [^{14}C-C2]-(*R,S*)-methylmalonyl-CoA.

A. B.

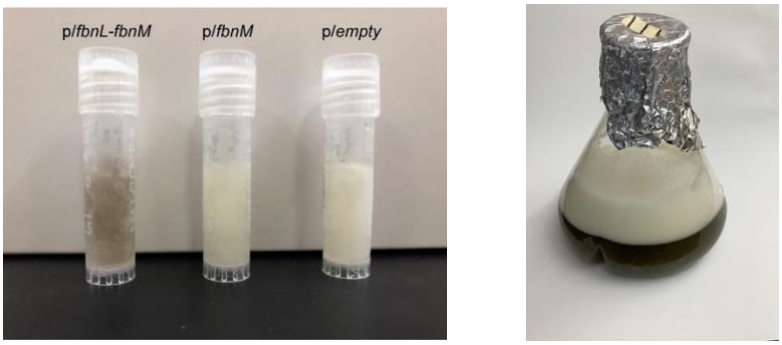


Figure S7 Pigment production by overproduced FbnL/M in *E. coli.* A) Pigment production was observed only when *fbnL* and *fbnM* are expressed together. B) Representative 1L *fbnLM* overexpression culture used to purify FbnL/M.

Table S1. ^{1}H and ^{13}C NMR data for FBN A (600 MHz, ^{1}H; 125 MHz, ^{13}C, DMSO-d_6).

	Position	^{13}C	^{1}H	^{1}H-^{1}H COSY	HMBC	^{13}C-^{13}C COSY
Dmaq	2	64.4, CH_2	3.76, m; 3.47, m	2b, 3; 2a, 3		3
	3	39.2, CH	4.45, m	2, 4, 13		2, 4
	4	30.9, CH_2	2.82, m	3	3, 5, 10	3, 5
	5	118.5, C				4, 6, 10
	6	115.4, CH	6.50, m		4, 7, 8, 10	5, 7
	7	146.8, C				6
	8	145.4, C				9
	9	107.9, CH	7.30, m		4, 5, 7, 8, 10	8, 10
	10	132.8, C				5, 9
	11	57.9, CH_3	3.49, s		2, 10, 12	
	12	56.9, CH_3	3.57, s		2, 10, 11	
	13		7.98, d (7.0)	3	3, 14	
	7-OH		9.88, br s			
	8-OH		9.41, br s			
Gly1	14	169.1, C				15
	15	42.1, CH_2	3.76, m; 3.69, m	16	14, 16	14
	16		8.43, m	15	17	
Ala	17	172.8, C				18
	18	49.1, CH	4.19, m	19, 20	17, 19	17, 19
	19	17.5, CH_3	1.26, d (6.7)	18	17, 18	18
	20		8.44, m		18, 19, 21	
Gly2	21	169.4, C				22
	22	41.9, CH_2	4.05, dd (5.3, 16); 3.88, dd (5.0, 16)	23	21, 24	21
	23		8.43, m	22	24	
Ala-Thz	24	160.9, C				25
	25	148.8, C				24, 29
	27	174.5, C				30
	29	124.0, CH	7.88, s		25, 27	25
	30	47.2, CH	5.17, dq (6.8, 7.6)	31, 32	27, 31	27, 31
	31	20.3, CH_3	1.56, d (6.8)	30	27, 30	30
	32		8.57, d (7.6)	30	33	
Ahpa	33	169.8, C				
	34	78.0, C				35
	35	35.2, CH_2	2.07, m	36	34, 36, 37	34
	36	34.2, CH_2	3.13, m; 3.04, m	35, 38		
	37	171.9, C				
	38		7.87, m	36	36, 38	
β-Ala	39	170.3, C				40
	40	34.8, CH_2	2.28, m	41	39, 41	39

	41	35.4, CH_2	3.31, m	40, 42	39, 40	
	42		8.15, m	41	41, 43	
Cdua	43	169.8, C				
	44	77.6, C				45
	45	40.7, CH_2	2.32, m	46	43, 44, 47, 54	44
	46	78.1, CH	4.60, m	45, 47	48	47
	47	34.5, CH_2	1.69, m; 1.60, m	46, 48	48, 49	46
	48	24.7, CH_2	1.34, m	47	49	49
	49	28.8, CH_2	1.29, m		50, 51	48
	50	28.7, CH_2	1.29, m		49, 51	51
	51	31.3, CH_2	1.24, m		50, 52	50
	52	22.1, CH_2	1.28, m	51	51, 53	53
	53	14.0, CH_3	0.86, t (6.6)	52	51, 52	52
	54	173.9, C				

Table S2. ^{1}H and ^{13}C NMR data for FBN B (600 MHz, ^{1}H; 125 MHz, ^{13}C, DMSO-d_6).

	Position	^{13}C	^{1}H	^{1}H-^{1}H COSY	HMBC	^{13}C-^{13}C COSY
Dmaq	2	64.4, CH_2	3.76, m; 3.47, m	2b, 3; 2a, 3		3
	3	39.2, CH	4.45, m	2, 4, 13		2, 4
	4	30.9, CH_2	2.82, m	3	3, 5, 10	3, 5
	5	118.5, C				4, 6, 10
	6	115.4, CH	6.50, m		4, 7, 8, 10	5, 7
	7	146.8, C				6
	8	145.4, C				9
	9	107.9, CH	7.30, m		4, 5, 7, 8, 10	8, 10
	10	132.8, C				5, 9
	11	57.9, CH_3	3.49, s		2, 10, 12	
	12	56.9, CH_3	3.57, s		2, 10, 11	
	13		7.98, d (7.0)	3	3, 14	
	7-OH		9.88, br s			
	8-OH		9.41, br s			
Gly1	14	169.1, C				15
	15	42.1, CH_2	3.76, m; 3.69, m	16	14, 16	14
	16		8.43, m	15	17	
Ala	17	172.8, C				18
	18	49.1, CH	4.19, m	19, 20	17, 19	17, 19
	19	17.5, CH_3	1.26, d (6.7)	18	17, 18	18
	20		8.44, m		18, 19, 21	
Gly2	21	169.4, C				22
	22	41.9, CH_2	4.05, dd (5.3, 16); 3.88, dd (5.0, 16)	23	21, 24	21
	23		8.43, m	22	24	
Ala-Thz	24	160.9, C				25
	25	148.8, C				24, 29
	27	174.5, C				30
	29	124.0, CH	7.88, s		25, 27	25
	30	47.2, CH	5.17, dq (6.8, 7.6)	31, 32	27, 31	27, 31
	31	20.3, CH_3	1.56, d (6.8)	30	27, 30	30
	32		8.57, d (7.6)	30	33	
Ahpa	33	169.8, C				
	34	78.0, C				35
	35	35.2, CH_2	2.07, m	36	34, 36, 37	34
	36	34.2, CH_2	3.13, m; 3.04, m	35, 38		
	37	171.9, C				
	38		7.87, m	36	36, 38	
Ala	39	170.3, C				
	40	50.2, CH	3.91, m	41		41
	41	16.8, CH_3	1.26, m	40, 42		40
	42		8.15, m	41		
Cdua	43	169.8, C				
	44	77.6, C				45

45	40.7, CH_2	2.32, m	46	43, 44, 47, 54	44
46	78.1, CH	4.60, m	45, 47	48	47
47	34.5, CH_2	1.69, m; 1.60, m	46, 48	48, 49	46
48	24.7, CH_2	1.34, m	47	49	49
49	28.8, CH_2	1.29, m		50, 51	48
50	28.7, CH_2	1.29, m		49, 51	51
51	31.3, CH_2	1.24, m		50, 52	50
52	22.1, CH_2	1.28, m	51	51, 53	53
53	14.0, CH_3	0.86, t (6.6)	52	51, 52	52
54	173.9, C				

Table S3. Predicted functions of ORFs in the *fbn* gene cluster.

GenBank Gene Notation[a]	Fbn gene	Protein ID	Putative Function
panD	*fbnA*	NP_356949.1	Aspartate α-decarboxylase
Atu3668	*fbnB*	NP_356948.1	IscR family transcriptional regulator
Atu3669	*fbnC*	NP_356947.2	MATE family multidrug efflux pump
Atu3670	*fbnD*	NP_356946.2	NRPS (C-$A_{\beta\text{-}Ala}$-T)
Atu3671	*fbnE*	NP_356945.1	Flavin-dependent monooxygenase
Atu3672	*fbnF*	NP_356944.1	PKS (KS-AT_{mal}-KR-T)
Atu3673	*fbnG*	NP_356943.2	PKS (FAAL-T)
Atu3675	*fbnH*	NP_356942.1	NRPS (C-A_{DOPA}-T-Red)
Atu3676	*fbnI*	NP_356941.1	SAM-dependent methyltransferase
Atu3677	*fbnJ*	NP_356940.1	Glutamate-1-semialdehyde aminotransferase
Atu3678	*fbnK*	NP_356939.2	MbtH-like protein
Atu3679	*fbnL*	NP_356938.1	Heme peroxidase-associated protein
Atu3680	*fbnM*	NP_356937.2	Heme peroxidase
Atu3681	*fbnN*	NP_356936.2	PKS (KS-AT_{mal}-KR-T)
Atu3682	*fbnO*	NP_356935.1	NRPS (C-A_{Ala}-T-Cy-ACy_{s}-T)
Atu3683	*fbnP*	NP_356934.2	NRPS (C-A_{Gly}-T-C-A_{Ala}-T-E)
Atu3684	*fbnQ*	NP_356933.1	NRPS (C-T)
Atu3685	*fbnR*	NP_356932.2	Type II Thioesterase
Atu3686	*fbnS*	NP_356931.1	Ferric iron reductase protein
Atu3687	*fbnT*	NP_356930.2	TonB-dependent receptor
Atu3688	*fbnU*	NP_356929.2	ABC transporter permease
Atu3689	*fbnV*	NP_356928.1	ABC transporter permease
Atu3690	*fbnW*	NP_356927.2	ABC transporter permease
Atu3691	*fbnX*	NP_356926.1	ABC transporter ATPase
Atu3692	*fbnY*	NP_356925.2	FecE-like sigma factor
Atu3693	*fbnZ*	NP_356924.1	FecR-like two component sensor kinase

[a]GenBank gene notation was assigned by Wood *et al.*[10].

Table S4. Kinetic parameters of FbnO-1 for activation of select amino acid substrates.

Amino acid	K_m (mM)[a]	V_{max} (nmol product/min/mg protein)	V_{max}/K_m
L-Alanine	0.38 ± 0.03	121.57 ± 3.57	318
L-Cysteine	5.98 ± 0.92	91.96 ± 8.39	15
L-Glycine	22.20 ± 3.96	120.78 ± 12.49	5
L-Serine	5.99 ± 0.46	137.79 ± 4.05	23

[a] The values are from the Michaelis-Menten curve, fit to the results of three independent assays with the standard errors from the nonlinear regression

Table S5. Proposed Dmaq-containing siderophore BGC locus tags shown in Figure 5.

Species	Locus Tag 5' end gene	Locus Tag 3' end gene
A. fabrum strain C58	Atu3670 (FbnD)	Atu3685 (FbnR)
A. fabrum 12D13[a]	Ga0399242_02_681708_685094	Ga0399242_02_722712_723458
A. fabrum 1D132[a]	Ga0398321_02_718220_721606	Ga0398321_02_759229_759975
A. fabrum ATCC 31749[a]	Ga0347210_02_691421_694807	Ga0347210_02_732431_733177
A. fabrum ARqua1[a]	HI842_18770	HI842_18705
A. fabrum Arqua[a]	EXN51_RS18715	EXN51_RS18790
Rhizobium sp. PDC82[a]	Ga0115500_2830	Ga0115500_2816
A. tumefaciens 15955[a]	Ga0398022_03_301123_304509	Ga0398022_03_262752_263495
A. tumefaciens 15-174[a]	Ga0340186_27_47785_51171	Ga0340186_27_9417_10160
A. tumefaciens CFBP5877[a]	Ga0399375_02_860168_863551	Ga0399375_02_901209_901955
A. tumefaciens CFBP5499[a]	Ga0398586_02_859197_862580	Ga0398586_02_901129_901905
A. tumefaciens LBA4404[a]	EN41_04020	EN41_03945
A. tumefaciens F4[a]	Ga0077515_1646	Ga0077515_16420
Rhizobium sp. NFIX01[a]	Ga0059025_03355	Ga0059025_03341
Rhizobium sp. NFIX02[a]	Ga0059026_03353	Ga0059026_03339
Rhizobium sp. UGM030330-04[a]	N434DRAFT_04040	N434DRAFT_04026
A. rhizogenes BGGC3284	EXN68_RS23280	EXN68_RS23355
M. pamukkalensis PK2	A6302_RS02085	A6302_RS02005
R. rhizosphaerae RD15	BTR14_RS05860	BTR14_RS05795
S. kummerowiae CCBAU 25048	GR128_RS09690	GR128_RS09765
Agrobacterium sp. a22-2	HFC70_RS13395	HFC70_RS13395
R. oryziradicis N19	BJF95_15780	BJF95_15845
A. vitis AV25/95[b]	G6L27_RS24135	G6L27_RS24070
A. vitis SZ1[b]	BBL07_RS13815	BBL07_RS1745
A. vitis AB3[b]	BBI12_RS15030	BBI12_RS15100
A. vitis AB3[b]	BBI04_RS14510	BBI04_RS14580
A. tumefaciens 15-172[b]	Ga0338076_45_294809_295555	Ga0338076_45_257950_264471
A. tumefaciens 1D1609[b]	Ga0272068_4194	Ga0272068_4207
R. rhizosphaerae MH17[b]	Ga0308746_141154	Ga0308746_141141

Aquamicrobium sp. LC103	XW59_RS10960	XW59_RS10885
P. aquimaris KCTC23043	AKU94_RS09275	AKU94_RS09190
P. sp. S12M18	EOK75_RS08355	EOK75_RS08275
Anabaena cylindrica PCC7122	Anacy_5467	Anacy_5452
Acaryochloris marina MBIC11017	AM1_A0226	AM1_A0246
Nostoc sp. '*Lobaria pulmonaria'*	NLP_RS01445	NLP_RS01560
Nostoc sp. PCC7120	All2650	All2635
Chrocosphaera subtropica ATCC51142	CCE_3042	CCE_3082
*Gloeocapsa*sp. PCC7428	Glo7482_5008	Glo7482_4891
Nodularia spumigena CCY9414	N9414_20950	N9414_21025
Phormidesmis priestleyi Ana	HLUCCA11_05230	HLUCCA11_05320
Planctomycetaceae isolate SP176	CMJ77_14245	CMJ77_14350

[a] Strains included in the Group I *A. tumefaciens* complex in Fig. 5

[b] Strains included in the Group I that are similar to *A. vitis* AV25/95 in Fig. 5

Table S6. Primers used for plasmid construction.

Primer	Sequence (5' → 3')
3670 F NdeI	AGGATTTGCGCATATGATGTCGAATGCC
3670 R HindIII	AGCGGAAAGCTTCAACTGGAACAGG
3672 F NdeI	CCGCGCGGCAGCCATATGAACGTAAAACAACAGGCC
3672 R HindIII	GAGTGCGGCCGCAAGCTTTCAACGACCGGCCTCGACGTC
3673 F NdeI	GGAGTTTTTCAGCATATGCAGATCATACCG
3673 R SacI	TTCATCGGCGATGAGCTCCGATCATCGATT
3675 F NdeI	CCGCGCGGCAGCCATATGAATTACATCGAGCCAGTAGAG
3675 R HindIII	GAGTGCGGCCGCAAGCTTTCATCGTTCACCCTGCTTTCG
3678 Duet F	ATAAGGAGATATACCATGAGTTCCCAGACACCGGCTGAAG
3678 Duet R	TTATGCGGCCGCAAGCTTCTATTCTTGCGCGCTGGTGATG
3681 F NdeI	CCGCGCGGCAGCCATATGTCCGACTATACCAGCGCT
3681 R HindIII	GAGTGCGGCCGCAAGCTTTCATCGTGCTTCTGCTCCTCC
3682-1 F NdeI	CCGCGCGGCAGCCATATGATGAACGAATTCTACGCCAAAC
3682-1 R HindIII	GAGTGCGGCCGCAAGCTTTCAAAGCCGGTTTGCCAGTCCAGC
3682-2 F AseI	GCCGCGCGGCAGCATTAATCTGGCAAACCGGCTTGAGAAT
3682-2 R XhoI	GGTGGTGGTGGTGGTGCTCGAGCTACTGGTCCTTTTGGCCGG
3683-1 F NcoI	GACCCCATGGACAAGACGGAGCATGCGGC
3683-1 R SalI	GGTCATGTCGACCAGAGCATCGACAGC
3683-2 F NcoI	AAAACCATGGGGCAATGGAACGAGACGG
3683-2 R HindIII	CCCAAGCTTCACAGTGCTTTCCACGTG

Table S7. Strains used in this study.

Strain	Purpose or characteristics	Source
E. coli DH5α	*fhuA2, Δ(argF-lacZ)U169, phoA, glnV44, Φ80, Δ(lacZ)M15, gyrA96, recA1, relA1, endA1, thi-1, hsdR17*	Laboratory strain
E. coli BL21(DE3)*ybdZ*::*aac(3)-IV*	BL21 (DE3) with *ybdZ*-disrupting apramycin resistance cassette	[4]
E. coli S17-1	*recA thi pro hsdR*$^{-}$ *M*$^{+}$ RP4-2 Tc^{r} :: Mu Km :: Tn7	Laboratory strain
A. fabrum str.C58	Wild-type	Laboratory strain
A. fabrum str.C58 *ΔfbnN*	C58 with markerless deletion of *fbnN*	This study

Table S8. Plasmids used in this study.

Plasmid	Purpose	Source
pACYC duet-1	T7 coexpression vector	Novagen
pACYC duet-*fbnK*	Coexpression of *fbnK*	This study
pET28b	T7 expression vector	Novagen
pET28b-*fbnD*	Overexpression of *fbnD*	This study
pET28b-*fbnF*	Overexpression of *fbnF*	This study
pET28b-*fbnG*	Overexpression of *fbnG*	This study
pET28b-*fbnH*	Overexpression of *fbnH*	This study
pET28b-*fbnN*	Overexpression of *fbnN*	This study
pET28b-*fbnO-1*	Overexpression of *fbnO* module 1-coding	This study
pET28b-*fbnO-2*	Overexpression of *fbnD* module 2-coding	This study
pET28b-*fbnP*	Overexpression of *fbnP*	This study
pET28b-*fbnP-2*	Overexpression of *fbnP* module 2-coding	This study
pBBR1MCS-2	Broad host range gene expression	[2]
pBBR1MCS-2-*fbnN*	Complementation of *fbnN* deletion strain	This study
pK18mobsacB	Host chromosome integration	ATCC87097
pK18mobsacB-*fbnN*	Markerless deletion of *fbnN*	This study
pANT841	high-copy vector for cloning	[11]
pTEV841-*fbnLfbnM*	Overexpression of *fbnL* and *fbnM*	This study

References

[1] McMahon, M. D., Rush, J. S., and Thomas, M. G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*, *J. Bacteriol. 194*, 2809–2818.

[2] Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., and Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes, *Gene 166*, 175–176.

[3] Mersereau, M., Pazour, G. J., and Das, A. (1990) Efficient transformation of *Agrobacterium tumefaciens* by electroporation, *Gene 90*, 149–151.

[4] Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases, *Biochemistry 49*, 8815–8817.

[5] Schomer, R. A., and Thomas, M. G. (2017) Characterization of the functional variance in MbtH-like protein interactions with a nonribosomal peptide synthetase, *Biochemistry 56*, 5380–5390.

[6] Chan, Y. A., and Thomas, M. G. (2010) Recognition of (2S)-aminomalonyl-acyl carrier protein (ACP) and (2R)-hydroxymalonyl-ACP by acyltransferases in zwittermicin A biosynthesis, *Biochemistry 49*, 3667–3677.

[7] Vincent, J. M. (1970) *A manual for the practical study of root nodule bacteria*, Blackwell Scientific Publications, Oxford, UK.

[8] Leong, S. A., and Neilands, J. B. (1982) Siderophore production by phytopathogenic microbial species, *Arch. Biochem. Biophys. 218*, 351–359.

[9] Alexander, D. B., and Zuberer, D. A. (1991) Use of the chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria, *Biol. Fertil. Soils 12*, 39–45.

[10] Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E., Almeida, N. F., Woo, L., Chen, Y., Paulsen, I. T., Eisen, J. A., Karp, P. D., Bovee, D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M. J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z. Y., Dolan, M., Chumley, F., Tingey, S. V., Tomb, J. F., Gordon, M. P., Olson, M. V., and Nester, E. W. (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58, *Science 294*, 2317–2323.

[11] DeSanti, C. L., and Strohl, W. R. (2003) Characterization of the *Streptomyces* sp. strain C5 snp locus and development of snp-derived expression vectors, *Appl. Environ. Microbiol. 69*, 1647–1654.