Structure and Biosynthesis of Amychelin, an Unusual Mixed-Ligand Siderophore from *Amycolatopsis sp.* AA4

Mohammad R. Seyedsayamdost,[†] Matthew F. Traxler,[‡] Shao-Liang Zheng,[§] Roberto Kolter,[‡] and Jon Clardy*,[†]

[†]Department of Biological Chemistry and Molecular Pharmacology and [‡]Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, United States
[§]Center for Crystallographic Studies, Harvard University, Cambridge, Massachusetts 02138, United States

S *Supporting Information*

ABSTRACT: Actinobacteria generate a large number of structurally diverse small molecules with potential therapeutic value. Genomic analyses of this productive group of bacteria show that their genetic potential to manufacture small molecules exceeds their observed ability by roughly an order of magnitude, and this revelation has prompted a number of studies to identify members of the unknown majority. As a potential window into this cryptic secondary metabolome, pairwise assays for developmental interactions within a set of 20 sequenced actinomycetes were carried out. These assays revealed that *Amycolatopsis sp.* AA4, a so-called "rare" actinomycete, produces a novel siderophore, amychelin, which alters the developmental processes of several neighboring streptomycetes. Using this phenotype as an assay, we isolated amychelin and solved its structure by NMR and MS methods coupled with an X-ray crystallographic analysis of its Fe-complex. The iron binding affinity of amychelin was determined using EDTA competition assays, and a biosynthetic cluster was identified and annotated to provide a tentative biosynthetic scheme for amychelin.

Bacteria sense and respond to the world around them with small molecules, and many of these molecules carry messages between microbial neighbors.[1] Eavesdropping on these chemical conversations represents an attractive approach to discovering new, naturally occurring molecules while simultaneously unveiling their biological function(s).[2] However, the molecular cacophony of most natural systems—the 10 000 bacterial species in a gram of soil, for example[3]—frustrates efforts to detect any single conversation. In an attempt to recreate simplified but plausible interaction models, we began a systematic screen of binary interactions within a panel of 20 soil-dwelling Actinobacteria, whose genomes have largely been sequenced at the Broad Institute. This report focuses on the structural elucidation and biosynthesis of amychelin, an unusual siderophore made by *Amycolatopsis sp.* AA4[4,5] that arrested development of several co-cultured streptomycetes, including the model strain *Streptomyces coelicolor*.

Using the developmental arrest, that is, the lack of white aerial hyphae formation in *S. coelicolor* (Figure 1A), as a phenotypic assay, we carried out activity-guided fractionation to isolate the responsible molecule produced by *Amycolatopsis sp.* AA4.

UV—visible spectra of the purified compound, which we named amychelin, suggested that it contained a hydroxybenzoyl group, a feature commonly found in microbial siderophores.[6] A chrome azurol S (CAS) assay[7] showed that iron-chelating activity co-eluted with the ability to alter development. However, further purification led to the loss of both activities, an effect attributed to decomposition of the active molecule. Purification under buffered neutral pH, mildly acidic conditions, or with different solvents and resins failed to prevent decomposition. Following the precedents from other research groups,[8] we added a 10-fold excess of $GaBr_3$ (or $FeCl_3$), which led to the appearance of a metal-to-ligand charge transfer band with $\lambda_{max} = 330$ nm/$\varepsilon = 3400$ M^{-1} cm^{-1} (or $\lambda_{max} = 435$ nm/$\varepsilon = 3000$ M^{-1} cm^{-1}), characteristic of metal complexation (Figure S1). Both Fe- and Ga-complexes of amychelin were purified to homogeneity and were stable during the course of purification and, in the case of Ga-amychelin, acquisition of NMR spectra.

The structure of amychelin was solved using 1D/2D NMR and high-resolution (HR)-ESI—MS analysis (Figures S2—S7, Table S1). ^{1}H NMR and gHSQC spectra revealed five amide protons at 8.65, 8.47, 7.87, 7.56, and 7.33 ppm along with a formyl singlet at 7.96 ppm (^{13}C NMR, 154 ppm). Analysis of gCOSY and TOCSY cross peaks gave seven spin systems consisting of an *ortho*-substituted phenol group, a modified Ser (lacking an amide hydrogen), three additional Ser residues, and two modified ornithines (Orn) (Figure 1B). On the basis of long-range [1]H—[13]C and ROESY interactions, we assigned the modified Ser to an *N*-terminal 2-hydroxybenzoyl-oxazoline group and the ornithines as N^δ-hydroxy-N^δ-formylornithine (*N*-OH-*N*-formyl-Orn) and cyclic N^δ-hydroxyornithine (*N*-OH-Orn), consistent with a molecular formula of $C_{30}H_{39}O_{14}N_8Ga$ ([M+H]$^+$ calc 805.1921, expt 805.1910) obtained by HR-ESI-MS. The novel structure thus obtained is shown in Figure 1C. The ability of amychelin (**1**) to chelate iron lies in three residues at the ends of the peptide chain: a capped 2-hydroxybenzoyl-oxazoline moiety at the *N*-terminus and an *N*-OH-*N*-formyl-Orn followed by a cyclic *N*-OH-Orn at the *C*-terminus. A tri-Ser linker joins the two ends. While the chelating groups are well-precedented,[6] amychelin is only the second siderophore containing this combination.[9]

The three-dimensional structure of amychelin was first addressed through Marfey derivatization[10] of the amino acid fragments
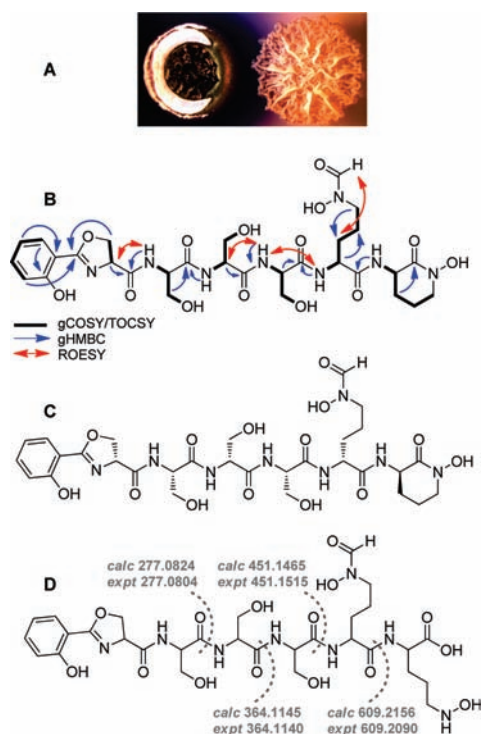
Figure 1. Activity and structure of amychelin. (A) *Amycolatopsis sp.* AA4 (right) inhibits development in *S. coelicolor* (left), which normally forms white aerial hyphae under these conditions. (B) Relevant spin systems, gHMBC and ROESY correlations. (C) Structure of amychelin (**1**). The stereochemistry is derived from the X-ray crystal structure (see below). (D) Structure of amychelin decomposition product **2** and the deduced MS/MS fragmentation data.

Figure 2. (A) X-ray crystal structure of amychelin solved at 0.84 Å resolution. (B) Magnified view of Fe binding by amychelin. The orientation has been modified with respect to (A) to highlight the distorted octahedron within the complex. (C) EDTA competition assay with Fe-amychelin yields pFe^{III} = 30.0 ± 1.6 from three independent iron binding assays.[16] A typical assay result is shown. The blue line describes a fit to Eq. S1. See text and Figure S14 for details.

produced by hydrolysis, which clearly indicated two D-Ser and two L-Ser residues. Interpretation of the data for the Orn residues was complicated by their chemical modifications, and we elected to resolve the complete structure of amychelin through single-crystal X-ray diffraction analysis. Crystals were prepared by layering hexane on a CH_2Cl_2 solution of the iron complex of amychelin and allowing the two to mix slowly. The X-ray-derived structure is shown in Figure 2A and represents one of the relatively few discrete siderophore structures with bound iron (Table S2).[11] The Fe-amychelin complex consists of a large loop in which four of the six amino acid-derived residues have a D-configuration (see Figure 1C). Three residues, one *N*-terminal and two *C*-terminal, form the iron-binding pocket, while the other three, all of which are serines, form a loop with solvent-exposed hydroxyl groups. The alternating D/L-Ser epimers allow the side chains of the linker to radiate out in the plane of the complex and likely contribute to both favorable solvation energy and possibly siderophore receptor recognition. The cyclic D-*N*-OH-Orn is in a half-chair conformation, and the D-*N*-OH-*N*-formyl-Orn conformation is twisted to allow side-on chelation of Fe (Figure 2A), while the opposite face of Fe remains open, perhaps to provide access to the bound Fe by downstream siderophore receptors. The structure reveals the unusual right-hand propeller or Δ configuration at the Fe center.[6a,12] In the complex, Fe is bound in a distorted octahedron, similar to structures of Fe-[11] and Ga-bound siderophores solved by NMR and computational methods.[13] The equatorial ligands, as shown in Figure 2B, are out of plane and deviate from the optimal 90° angle. The upper axial bond is bent toward the hydroxamate ligands (Figure 2B).
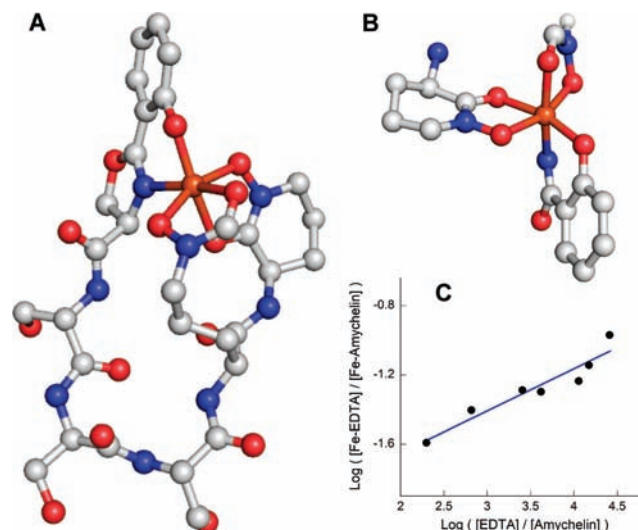
With the structure of amychelin in hand, the structure of its primary degradation product (Figure 1D, **2**) was deduced using 1D/2D NMR analysis and MS fragmentation studies (Figures S8–S13, Table S3). NMR revealed that **2** resulted from the hydrolysis of the *C*-terminal hydroxamide bond, in line with results from HR-ESI-MS analysis (**2**, $[M+H]^+$ calc 757.3005, expt 757.3015). MS fragmentation confirmed the linear sequence with major fragments consistent with the losses of the *C*-terminal *N*-OH-Orn (b_5 ion, $[M]^+$ calc 609.2156, expt 609.2092), both modified Orn residues (b_4 ion, $[M]^+$ calc 451.1465, expt 451.1515), the *C*-terminal Ser-Orn-Orn fragment (b_3 ion, $[M]^+$ calc 364.1145, expt 364.1140), and the Ser-Ser-Orn-Orn fragment (b_2 ion, $[M]^+$ calc 277.0824, expt 277.0804) (Figure 1D). Hydrolysis of the cyclic *N*-OH-Orn in **2** rationalizes the loss of its ability to chelate iron.

To quantitate the affinity of amychelin for iron, we carried out competition titrations with EDTA, as previously reported.[14] In this method, varying concentrations of EDTA are incubated with a constant concentration of Fe-amychelin at pH 7.4, and the distribution of Fe between EDTA and amychelin after equilibration is determined using UV–vis spectroscopy. A plot of the log of free [EDTA]/[amychelin] vs [Fe-EDTA]/[Fe-amychelin] can be fit to Eq. S1 to determine a difference in pFe^{III} between EDTA and amychelin.[14] The stability constants for EDTA are known (pFe^{III} = 23.42);[15] therefore, this analysis provides a pFe^{III} for amychelin.[16] Three independent competition titrations with various ranges of [EDTA] were carried out, and a representative data set is shown in Figure 2C (see also Figure S14). The data yield ΔpFe^{III} = +6.6 ± 1.6, indicating pFe^{III} = 30.0 ± 1.6 for amychelin. Interestingly, *S. coelicolor* is only known to produce hydroxamate-type siderophores with significantly lower pFe^{III} values than amychelin (typical pFe^{III} ≈ 22–26),[14,15,17,18] suggesting a model for how amychelin may arrest development (see below). The greater pFe^{III} of amychelin reflects in part the contribution of the 2-hydroxybenzoyl-oxazoline ligand, consistent with pFe^{III} ranges reported for other known siderophores.
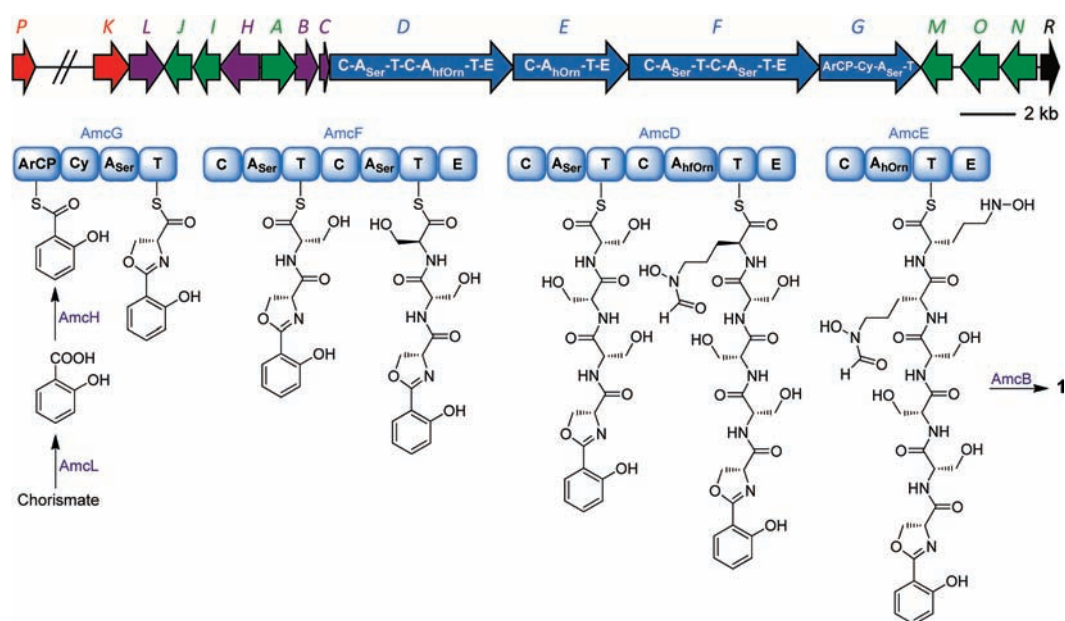
Figure 3. (Top) Annotated *amc* gene cluster for amychelin detailing genes *amcA−amcR*. Genes are color-coded as follows: NRPS genes, blue; genes involved in amino acid or amychelin transport, green; genes involved in initiation or chain termination, including the mbtH-like *amcC*, purple; and amino acid modification genes, red. *AmcR* (black) is a tetR-type regulator. (Bottom) Biosynthetic model for amychelin: ArCP, aryl carrier protein; Cy, cyclization domain; A, adenylation domain; T, peptidyl carrier protein or thiolation domain; C, condensation domain; E, epimerization domain.

We next attempted to find a candidate biosynthetic cluster for amychelin in the draft genome of *Amycolatopsis sp.* AA4. The 2-hydroxybenzoyl-oxazoline group, found in amychelin and other siderophores, is typically introduced into the biosynthetic pathway by a hydroxybenzoyl-AMP ligase.[19] A homology search using the amino acid sequence of this enzyme from *Mycobacterium smegmatis* retrieved a single gene (SSMG_02542), which was clustered with several non-ribosomal peptide synthetase (NRPS) genes in the *Amycolatopsis sp.* AA4 draft genome. Preliminary knockout mutagenesis indicates that this cluster is responsible for amychelin production (see Supporting Information). We have assigned this the *amc* gene cluster and defined its boundaries on the basis of short intergenic sequences within the cluster. It is 35.2 kb in size and consists of 16 coding sequences (SSMG_02531 to SSMG_02546). Annotation of the *amc* cluster is shown in Table S4, and the results are summarized in Figure 3. The cluster consists of four NRPS genes (blue, *amcDEFG*), amino acid tailoring genes (red, *amcKP*), genes involved in chain initiation and termination (purple, *amcBCHL*), and amino acid or amychelin transporters (green, *amcAIJMON*). AmcK is likely a flavin-dependent hydroxylase with 61% identity to the Orn hydroxylase implicated in both erythrochelin[20] and coelichelin[18] biosynthesis. The cluster does not encode a formyltransferase. A homology search using the formyltransferase from *S. coelicolor* used in coelichelin biosynthesis[18] retrieved a single hit with 74% identity. We tentatively assign this gene *amcP* and propose that it is involved in formylation of Orn (or *N*-OH-Orn). This assignment is analogous to erythrochelin biosynthesis, where the *N*-acetyltransferase used to generate N^δ-OH-N^δ-acetyl-Orn was found to reside outside of the erythrochelin biosynthetic cluster.[20a] Thus, AmcK and AmcP likely generate the *N*-OH-Orn and *N*-OH-*N*-formyl-Orn employed in amychelin biosynthesis.

AmcL is homologous to salicylate synthase from *M. smegmatis* (54% identity) and likely generates salicylate from chorismate (Figure 3).[21] AmcH is similar to EntE of *Bacillus subtilis* (58% identity) and assigned as a hydroxybenzoyl-AMP ligase, which likely adenylates salicylate and loads it onto the aryl carrier protein (ArCP) domain in AmcG (Figure 3).[19] The domain organization of the four NRPS genes is shown in Figure 3, and the specificities of the adenylation (A) domains are summarized in Table S5. By comparison with known A-domains,[22] the active-site residues of the A-domain in AmcG predict activation of Cys, though only Ser is incorporated. The cyclization domain (Cy) then catalyzes oxazoline formation, and the growing chain is transferred to AmcF (Figure 3). Both A-domains in AmcF are predicted to activate Ser (Table S5). Following incorporation of the second Ser residue, the epimerization (E) domain generates its D-stereoisomer.[19b] The A-domains in AmcD are predicted to activate Ser and *N*-OH-*N*-formyl-Orn, consistent with the structure of amychelin (Table S5). Subsequent to condensation of both residues, the E-domain in AmcD generates D-*N*-OH-*N*-formyl-Orn. The A-domain in the final NRPS, AmcE, is homologous to Gln-activating A-domains. This NRPS also lacks a C-terminal thioesterase domain, suggesting that it requires a specific hydrolase for peptide chain release. We propose that after condensation of *N*-OH-Orn and epimerization of its α-H in AmcE, the transacylase, AmcB, releases the peptide by generating a terminal cyclic *N*-OH-Orn. AmcB belongs to the α/β hydrolase superfamily known to carry out such reactions,[23] and this chain release mechanism resembles coelichelin biosynthesis, where a free-standing esterase, also belonging to the α/β hydrolase superfamily, catalyzes hydrolysis of the final thiolation (T)-domain-bound thioester.[18] In this case, the transacylase, AmcB, likely uses the hydroxylamine of *N*-OH-Orn as a nucleophile, rather than water, which could represent a general route for the biosynthesis of cyclic C-terminal *N*-OH-Orn residues found in other siderophores.[6] Our bioinformatic analysis does not identify an E-domain in AmcG to generate the *N*-terminal D-Ser. It is possible that the system makes use of the E-domain in

11436

dx.doi.org/10.1021/ja203577e |*J. Am. Chem. Soc.* 2011, 133, 11434–11437

AmcF. Such domain and module reuse has been implicated in the biosynthesis of coelichelin,[18] fuscachelin,[24] yersiniabactin,[25] and in other systems.[26] Alternatively, the Cy-domain in AmcG may be D-Ser selective, and the racemase may be provided by another gene product outside of the amychelin cluster. *In vitro* experiments will be necessary to distinguish between these options and to test the predictions made by the model.

In summary, we present the structure, iron-binding properties, and a putative biosynthetic model for amychelin. This report is not the first to link siderophores and iron with streptomycete development, but in the other reports, siderophores *promoted* streptomycete development rather than inhibiting it.[27] The usual explanation for this previously observed stimulatory effect is "siderophore piracy", the ability of bacteria to use siderophores that they do not themselves biosynthesize.[28] In accord with this model, bacterial genomes typically contain multiple receptors for siderophores for which there are no corresponding biosynthetic clusters. Amychelin's unusual structure, its greater than usual (for streptomycetes) iron-binding affinity, and its ability to shut down development in neighboring streptomycetes collectively point to a model in which *Amycolatopsis sp.* AA4 produces amychelin to frustrate siderophore piracy, thereby monopolizing scarce iron resources for itself. This model is currently being examined further in our laboratories.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Experimental details and characterization data for **1** and **2**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
jon_clardy@hms.harvard.edu

■ ACKNOWLEDGMENT

■ REFERENCES

(1) (a) Shank, E. A.; Kolter, R. *Curr. Opin. Microbiol.* **2009**, *12*, 205. (b) Camilli, A.; Bassler, B. L. *Science* **2006**, *311*, 1113. (c) Yim, G.; Wang, H. H.; Davies, J. *Philos. Trans. R. Soc. London B: Biol. Sci.* **2007**, *362*, 1195. (d) Yim, G.; Wang, H. H.; Davies, J. *Int. J. Med. Microbiol.* **2006**, *296*, 163.

(2) (a) Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 466. (b) Piel, J. *Nat. Prod. Rep.* **2009**, *26*, 338.

(3) Curtis, T. P.; Sloan, W. T.; Scannell, J. W. *Proc. Natl. Acad. Sci. U. S.A.* **2002**, *99*, 10494.

(4) D'Costa, V. M.; McGrann, K. M.; Hughes, D. W.; Wright, G. D. *Science* **2006**, *311*, 374.

(5) Initial analysis identified this strain as a streptomycete, but the 16S sequence confirms that it belongs to the genus *Amycolatopsis*.

(6) (a) Hider, R. C.; Kong, X. *Nat. Prod. Rep.* **2010**, *27*, 637. (b) Sandy, M.; Butler, A. *Chem. Rev.* **2009**, *109*, 4580. (c) Miethke, M.; Marahiel, M. A. *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 413.

(7) Schwyn, B.; Neilands, J. B. *Anal. Biochem.* **1987**, *160*, 47.

(8) (a) Mohn, G.; Koehl, P.; Budzikiewicz, H.; Lefevre, J. F. *Biochemistry* **1994**, *33*, 2843. (b) Sharman, G. J.; Williams, D. H.; Ewing, D. F.; Ratledge, C. *Chem. Biol.* **1995**, *2*, 553.

(9) Sontag, B.; Gerlitz, M.; Paululat, T.; Rasser, H.-F.; Grün-Wollny, I.; Hansske, F. G. *J. Antibiot.* **2006**, *59*, 659.

(10) Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591.

(11) (a) Zalkin, A.; Forrester, J. D.; Templeton, D. H. *J. Am. Chem. Soc.* **1966**, *88*, 1810. (b) Norrestam, R.; Stensland, B.; Brändén, C.-I. *J. Mol. Biol.* **1975**, *99*, 501. (c) van der Helm, D.; Poling, M. *J. Am. Chem. Soc.* **1976**, *98*, 82. (d) Hossain, M. B.; Eng-Wilmot, D. L.; Loghry, R. A.; van der Helm, D. *J. Am. Chem. Soc.* **1980**, *102*, 5766. (e) Miller, M. C.; Parkin, S.; Fetherston, J. D.; Perry, R. D.; Demoll, E. *J. Inorg. Biochem.* **2006**, *100*, 1495.

(12) Stack, T. D. P.; Karpishin, T. B.; Raymond, K. N. *J. Am. Chem. Soc.* **1992**, *114*, 1512.

(13) (a) Atkinson, R. A.; Salah El Din, A. L. M.; Kieffer, B.; Lefevre, J.-F.; Abdallah, M. A. *Biochemistry* **1998**, *37*, 15965. (b) Wasielewski, E.; Atkinson, R. A.; Abdallah, M. A.; Kieffer, B. *Biochemistry* **2002**, *41*, 12488. (c) Fadeev, E. A.; Luo, M.; Groves, J. T. *J. Am. Chem. Soc.* **2004**, *126*, 12065.

(14) (a) Abergel, R. J.; Zawadzka, A. M.; Raymond, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 2124. (b) Harris, W. R.; Carrano, C. J.; Raymond, K. N. *J. Am. Chem. Soc.* **1979**, *101*, 2722. (c) Harris, W. R.; Carrano, C. J.; Raymond, K. N. *J. Am. Chem. Soc.* **1979**, *101*, 2213.

(15) Smith, R. M.; Martell, A. E. *Critical Stability Constants*; Plenum: New York, 1977; Vols. 1—4.

(16) pFe^{III} is a measure of the affinity of a siderophore for Fe and is defined as the negative logarithm of free [Fe] in the presence of 10^{-6} M total Fe and 10^{-5} M total ligand.

(17) (a) Patel, P.; Song, L.; Challis, G. L. *Biochemistry* **2010**, *49*, 8033. (b) Barry, S. M.; Challis, G. L. *Curr. Opin. Chem. Biol.* **2009**, *13*, 205.

(18) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nat. Chem. Biol.* **2005**, *1*, 265.

(19) (a) Quadri, L. E. N.; Sello, J.; Keating, T. A.; Weinreb, P. H.; Walsh, C. T. *Chem. Biol.* **1998**, *5*, 631. (b) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468.

(20) (a) Lazos, O.; Tosin, M.; Slusarczyk, A. L.; Boakes, S.; Cortes, J.; Sidebottom, P. J.; Leadlay, P. F. *Chem. Biol.* **2010**, *17*, 160. (b) Robbel, L.; Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. *FEBS J.* **2010**, *277*, 663.

(21) Nagachar, N.; Ratledge, C. *FEMS Microbiol. Lett.* **2010**, *308*, 159.

(22) (a) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493. (b) Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009**, *458*, 181. (c) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211.

(23) Carr, P. D.; Ollis, D. L. *Protein Pept. Lett.* **2009**, *16*, 1137.

(24) Dimise, E. J.; Widboom, P. F.; Bruner, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15311.

(25) (a) Suo, Z.; Tseng, C. C.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S. A.* **2001**, *98*, 99. (b) Mootz, H. D.; Schwarzer, D.; Marahiel, M. A. *Chembiochem* **2002**, *3*, 490.

(26) (a) Shaw-Reid, C. A.; Kellerher, N. L.; Losey, H. C.; Gehring, A. M.; Berg, C.; Walsh, C. T. *Chem. Biol.* **1999**, *6*, 385. (b) Blodgett, J. A.; Oh, D. C.; Cao, S.; Currie, C. R.; Kolter, R.; Clardy, J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11692. (c) Gaitatzis, N.; Kunze, B.; Müller, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11136.

(27) (a) Ueda, K.; Kawai, S.; Ogawa, H.-O.; Kiyama, A.; Kubota, T.; Kawanobe, H.; Beppu, T. *J. Antibiot.* **2000**, *53*, 979. (b) Yamanaka, K.; Oikawa, H.; Ogawa, H.-O.; Hosono, K.; Shinmachi, F.; Takano, H.; Sakuda, S.; Beppu, T.; Ueda, K. *Microbiology* **2005**, *151*, 2899.

(28) Schubert, S.; Fischer, D.; Heesemann, J. *J. Bacteriol.* **1999**, *181*, 6387.