

Genetic Interception and Structural Characterization of Thiopeptide Cyclization Precursors from *Bacillus cereus*

Albert A. Bowers, Christopher T. Walsh,* and Michael G. Acker

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Received May 25, 2010; E-mail: christopher_walsh@hms.harvard.edu

Abstract: The pyridine core of the thiocillins has long been postulated to arise from a late-stage tail-to-tail condensation of two dehydroalanines. Genetic disruption of *tcIM*, a proposed “Diels–Alderase”, allowed isolation of acyclic precursors to this pyridine ring. The isolated products possess the full cohort of post-translational modifications that are normally displayed by the thiocillins, including dehydrobutyrines, thiazoles, C-terminal decarboxylation, and the two previously unconfirmed dehydroalanines. Additionally, leader peptides have undergone extensive N-terminal degradation and the remaining leader peptide residues have been N-succinylated. These results identify TcIM and its homologues in other thiazolyl peptide producing strains as the enzymes responsible for the trans-annular heteroannulation at core of this class of molecules.

Thiazolyl peptides are ribosomally encoded, highly post-translationally modified peptide antibiotics with potent activity against gram-positive bacteria.¹ Common structural motifs include several typically nonribosomal moieties: dehydroalanines (Dhas), dehydrobutyrines (Dhbs), oxazolines/oxazoles, and thiazolines/thiazoles. However, the most broadly distinguishing attribute of thiazolyl peptides is the presence of a core nitrogenous heterocycle, which serves as the lynchpin of a highly constrained macrocycle. Three different macrocycle sizes have been discovered thus far: 26 atoms (thiocillins,^{2a–c} nosiheptide,^{2d–f} and thiostrepton^{2g–i}), 29 atoms (GE2270A^{2j,k}), and 35 atoms (berninamycin^{2l–n}). Among members of the 29-atom and 35-atom subclasses, only fully aromatized pyridines have been observed in the lynchpin position, while pyridines, dehydropiperidines, and piperidines have all been observed in members of the 26-atom subclass. Early feeding studies demonstrated that the constituents of these lynchpin heterocycles are derived from the amino acid serine in what has long been postulated as a tail-to-tail condensation proceeding via the corresponding dehydroalanine moieties.³ The work presented herein demonstrates that, in the case of the thiocillins, this biosynthetic hypothesis is operative and that a fully modified peptide bearing two Dhas at positions 1 and 10 is the acyclic precursor in thiocillin biosynthesis. Further, we identify TcIM, a 39 kDa protein with no discernible homology to proteins of known function, as responsible for the putative cycloaddition reaction that combines the two residues.

To date, only two enzymes have been purified to homogeneity and demonstrated to catalyze reactions, wherein the substrates are consistent with a [4 + 2] cycloaddition: lovastatin nonaketide synthase (LovB) and macrophomate synthase.⁴ Such putative “Diels–Alderases” have been suggested in the biosyntheses of a number of other natural products, but the question remains as to whether or not these transformations require enzyme catalysis. Nevertheless, the “hetero-Diels–Alderase” that would be necessary for construction of the thiocillin pyridine moiety is unprecedented in the catalog of enzymology. Total syntheses of amythiamycin,

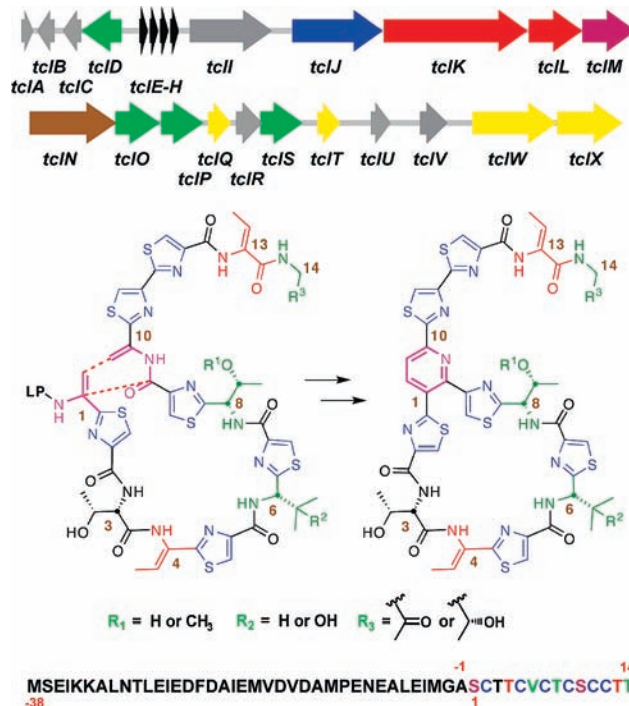


Figure 1. Annotated gene cluster responsible for production of thiocillins from a 52-residue prepeptide. Formation of the pyridine ring illustrated as the final step (LP = leader peptide).

thiostrepton, and GE2270A and T have made use of “biomimetic” heteroannulations; however, in all instances, the transformation has been used in early stage ring formations, free from constraint by an incipient macrocycle, and requires alternative catalysis such as heat or metals.⁵ Thus it is of interest to determine if that heterocyclization would be enzyme catalyzed.

Thiocillin biosynthesis in *B. cereus* ATCC 14579 proceeds via ribosomal translation of a 52 amino acid prepeptide, which contains both a 38-residue leader sequence and the 14 amino acids that end up in the mature natural product after a cascade of posttranslational modifications (Figure 1).⁶ In addition to the four copies of the prepeptide encoding gene (*tcIE-H*), the gene cluster contains two lantibiotic-type dehydratases (*tcIK* and *L*),^{7a,b} a McbBC-like cyclodehydratase and associated dehydrogenase (*tcIJ* and *N*, respectively),^{7c,d} several proposed resistance elements (*tcIQ*, *T*, *W*, and *X*), and a number of tailoring enzymes that effect stochastic changes at residues 6, 8, and 14 (*tcID*, *O*, *P*, and *S*). Combinations of modifications effected by the latter set of enzymes give rise to eight thiocillin variants all of which can be isolated from this strain of *B. cereus*: thiocillin I, II, III, and IV; Micrococcin P1 and P2; and YM-266183 and -266184 (compounds 1–8, respectively). Of the unassigned ORFs in the gene cluster, *tcIM* first garnered our attention due to the presence of

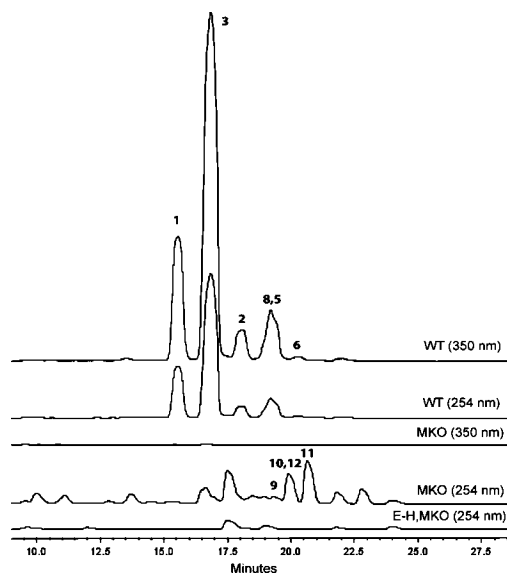


Figure 2. HPLC traces at 254 and 350 nm of wild type thiocillins, *MKO*, and *E-H,MKO* isolates.

homologues in several other recently reported gene clusters: *tsrL* (thiostrepton),^{8a,b} *nosO* (Nosiheptide),^{8c} and *tpdD* (GE2270A)^{8d} (see Supporting Information (SI) for sequences and alignment). Moreover, an M homologue was conspicuously absent from the genome of *Streptomyces* TP-A0584, which produces goadsporin, an acyclic thiazolyl variant.⁹ For these reasons, our initial efforts in characterization of a putative hetero-Diels–Alderase for *B. cereus* ATCC 14579 have focused on the 975 bp gene *tclM*. Due to the complexity of the substrates involved in thiocillin biosynthesis, we had previously made use of a double crossover knockout of the structural genes *tclE-H* to probe the assembly line *in vivo*.¹⁰ For initial investigation of *tclM*, we again relied on this strategy. The *tclM* knockout (*MKO*) was prepared via homologous recombination with a plasmid bearing sequence homology to areas adjacent to, but not containing, *tclM*. The same plasmid was employed in conjunction with our previously reported knockout of the structural genes to prepare a tandem E-H,M knockout (*E-H,MKO*), into which was next introduced a single copy of *tclE* via Campbell integration (*E_{in}MKO*). All three strains, *MKO*, *E-H,MKO*, and *E_{in}MKO* were grown and extracted in the same manner as that for wild-type. Analysis was performed by high-resolution ESI-LC/MS, with UV monitoring at 254 and 350 nm (Figure 2). Due to the

conjugated aromatic system at the core of the thiocillins, wild type compounds absorb across both wavelengths. In contrast, extracts from both the *MKO* and *E_{in}MKO* strains exhibited peaks that absorbed at 254 nm, but not at 350 nm. Four significant masses corresponded to the new peaks in the isolates: 1421.3447 ($C_{58}H_{69}N_{16}O_{15}S_6^+$ **9**), 1364.3232 ($C_{56}H_{66}N_{15}O_{14}S_6^+$ **10**), 1194.2541 ($C_{49}H_{56}N_{13}O_{11}S_6^+$ **11**), and 1196.1697 ($C_{49}H_{58}N_{13}O_{11}S_6^+$ **12**). These peaks and associated masses were absent from traces and mass spectra of both the wild type producer and the tandem *E-H,MKO*.

By combination of high-resolution MS/MS and 1D and 2D 1H and ^{13}C NMR spectra (Figure 3 and SI) we were able to assign the products from the *tclM* knockout cultures as the fully modified, linear, and successively proteolyzed 16–14 residues at the C-termini of the structural peptide; in instances where residues from the leader peptide remained, their N-termini had been succinylated. Thus, compound **9**, with mass 1421.3447 represents the fully modified 14-mer with a glycine and alanine from the leader peptide remaining at the N-terminus and the glycine nitrogen N-succinylated. Similarly, compound **10**, with mass 1364.3232 contains only the alanine residue of the leader peptide and is again N-succinylated, this time on alanine (Figure 2). We had previously observed N-succinylation in the *E-HKO* system, with prepeptide genes that encode lysines for incorporation into the macrocycle at positions 3, 4, and 8.^{10a} In those cultures, with a functional *TclM*, cyclization was effected satisfactorily, but the lysine distal nitrogen was succinylated, perhaps in self-protection, as may be the case here. Lastly, compounds **11** and **12**, with masses 1194.2541 and 1196.2697, appear to have been degraded back to the first residue, serine 1, of the 14-mer product peptide. In this case, cleavage of the Ala–Ser₁ amide bond would leave Dha₁ as a primary enamine, which rather than be captured by a presumed succinyl transferase has tautomerized and hydrolyzed to the methyl ketone (**11**). This ketone is further reduced to the secondary alcohol (**12**) in a fashion similar to that of the threonine at residue 14 of the natural product (see SI for structures and scheme).

To reconstitute thiocillin production, *tclM* was reintroduced via homologous recombination: peaks at 254 nm in the LC trace were again replaced by peaks at 350 nm, with masses corresponding to the fully cyclized and aromatized natural product. As an additional confirmation that the intermediate peaks were on track from the thiocillin gene cluster, a threonine-3 to alanine prepeptide mutant was introduced into the *E-H,MKO* and a corresponding –30 mass shift was observed in the MS-trace of the resultant compounds: 1391.3341 ($C_{57}H_{67}N_{16}O_{14}S_6^+$ **13**), 1334.3126 ($C_{55}H_{64}N_{15}O_{13}S_6^+$ **14**),

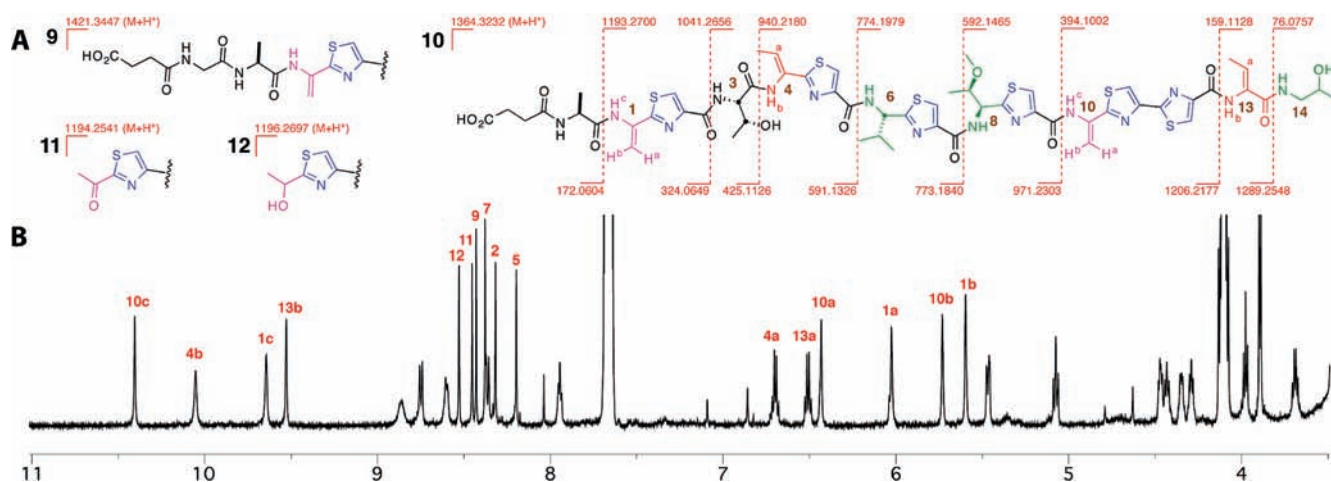


Figure 3. (A) Structure of linear thiocillin precursors isolated from fermentations of the *tclM* knockout strain and primary mass spectral fragmentation pattern of **10**. (B) 1H NMR of **10**, characteristic peaks labeled in red (600 MHz, DMSO- d_6 ; see SI for complete assignment).

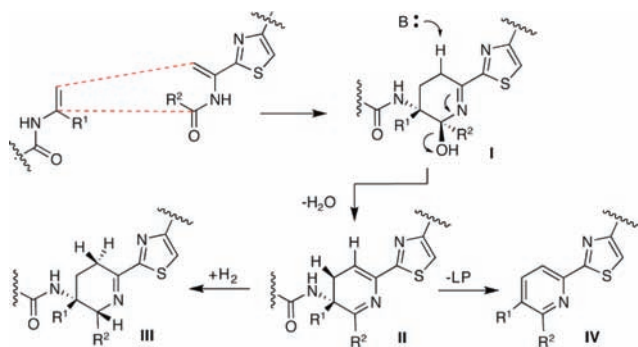


Figure 4. Possible routes for conversion of cycloadduct **II** to dehydropiperidine **III** or pyridine **IV**.

1164.2435 ($C_{48}H_{54}N_{13}O_{10}S_6^+$ **15**), and 1166.1591 ($C_{48}H_{56}N_{13}O_{10}S_6^+$ **16**) (see SI for structures).

Isolation of these derivatives from the *tcIM* knockout strain demonstrates that key post-translational modifications in thiocillin biosynthesis, thiazole formation, serine and threonine dehydration, and C-terminal oxidation/decarboxylation can, and perhaps must, precede central pyridine ring formation. Moreover, it strongly suggests that the pyridine ring is formed through a late stage cycloaddition (either stepwise or concerted) between the two dehydroalanines formed at serines 1 and 10. Although it cannot be conclusively ruled out that these modifications regularly succeed ring formation and that disruption of cyclization has allowed persistence long enough to permit nonspecific modification, this seems highly unlikely given the level of production and amount of peptidase-affected degradation. The conversion of 10 residues, 6 cysteines to planar thiazoles, 2 threonines, and 2 serines to sp^2 -containing Dhb and Dha residues may impose conformational constraints that allow Dha₁ and Dha₁₀ to come within bonding distance in the active site of TcIM. These results clearly implicate a specific class of enzyme, TcIM and its homologues, in pyridine/piperidine formation.

Moreover, the N-terminal peptide degradative trimming to -2 and -1 tails, as well as their *N*-succinylation, raises questions regarding the mode of cleavage of the leader peptide in biosynthesis of the mature natural products. On the one hand, the presence of both glycine and alanine cleavage products suggests action of nonspecific peptidases. However, the gly ala motif itself recalls the specific peptidases involved in maturation of the lantibiotics¹¹ and the cyanobactins.¹² *N*-Succinylation may be an effort on the part of the bacteria to detoxify uncyclized substrates, which could explain the presence of *tcIV*, an as yet uncharacterized acetyltransferase homologue, in the gene cluster. Alternatively, a role for such succinylation in cyclization, perhaps to enhance a final elimination step of the upstream peptide, cannot be ruled out.

The overall transformation of two serine residues to either a trisubstituted pyridine or dehydropiperidine in the final maturation step of the thiazolyl peptide antibiotics is a remarkable chemical outcome. If the mechanism depicted in Figure 4 is assumed, it seems clear that aromatization of intermediate **II** would be thermodynamically favored. It is unclear if, in the case of the dehydropiperidine cores, where again there are TcIM homologues, this intermediate is sustained and then reduced or if an alternate mechanism is in effect. Answers to these questions, as well as insight into the more precise timing and mechanism of cyclization to the core heterocycle, whether stepwise or concerted, await further work *in vivo* and *in vitro*.

Acknowledgment. We thank Michael Fischbach for his tremendous help and insight throughout the project. We also thank

Brian Ames for aid with bioinformatic analysis. This work was supported by NIH NIAID Grant Nos. AI057159 and GM20011 (C.T.W.) and the New England Regional Center of Excellence (NIAID 057159). Support for instrumentation was provided by the Taplin Funds for Discovery Program (Suzanne Walker, PI). A.A.B. is supported by the NIH National Cancer Institute (CA136283). M.G.A. is a fellow of the Helen Hay Whitney Foundation.

Supporting Information Available: Supporting figures and tables, experimental procedures, and spectral data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) (a) Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. *Chem. Rev.* **2005**, *105*, 685–714. (b) Arndt, H.-D.; Schoof, S.; Lu, J.-Y. *Angew. Chem., Int. Ed.* **2009**, *48*, 6770–6773.
- (2) Thiocillin: (a) Su, T. L. *Br. J. Exp. Pathol.* **1948**, *29*, 473–481. (b) Bycroft, B. W.; Gowland, M. S. *J. Chem. Soc., Chem. Commun.* **1978**, 256–258. (c) Fenet, B.; Pierre, F.; Cundliffe, E.; Ciufolini, M. A. *Tetrahedron Lett.* **2002**, *43*, 2367–2370. Nosiheptide: (d) Depaire, H.; Thomas, J. P.; Brun, A.; Lukacs, G. *Tetrahedron Lett.* **1977**, 1395–1396. (e) Pascard, C.; Ducruix, A.; Lunel, J.; Prange, T. *J. Am. Chem. Soc.* **1977**, *99*, 6418–6423. (f) Prange, T.; Ducruix, A.; Pascard, C.; Lunel, J. *Nature* **1977**, *265*, 189–190. Thiostrepton: (g) Anderson, B.; Hodgkin, D. C.; Viswamitra, M. A. *Nature* **1970**, *225*, 233–235. (h) Hensens, O. D.; Albers-Schonberg, G. *J. Antibiot.* **1983**, *36*, 832–845. (i) Trejo, W. H.; Dean, L. D.; Pluscec, J.; Meyers, E.; Brown, W. E. *J. Antibiot.* **1977**, *30*, 639–643. GE2270A: (j) Kettenring, J.; Colombo, L.; Ferrari, P.; Tavecchia, P.; Nebuloni, M.; Vekey, K.; Gallo, G. G.; Selva, E. *J. Antibiot.* **1991**, *44*, 702–715. (k) Selva, E.; Beretta, G.; Montanini, N.; Sandler, G. S.; Gastaldo, L.; Ferrari, P.; Lorenzetti, R.; Landini, P.; Ripamonti, F.; Goldstein, B. P.; Berti, M.; Montanaro, L.; Denaro, M. *J. Antibiot.* **1991**, *44*, 693–701. Berninamycin: (l) Abe, H.; Kushida, K.; Shiobara, Y.; Kodama, M. *Tetrahedron Lett.* **1988**, *29*, 1401–1404. (m) Liesch, J. M.; Millington, D. S.; Pandey, R. C.; Rinehart, K. L., Jr. *J. Am. Chem. Soc.* **1976**, *98*, 8237–8249. (n) Liesch, J. M.; Rinehart, K. L., Jr. *J. Am. Chem. Soc.* **1977**, *99*, 1645–1646.
- (3) (a) Mocek, U.; Beale, J. M.; Floss, H. G. *J. Antibiot.* **1989**, *42*, 1649–1652. (b) Mocek, U.; Knaggs, A. R.; Tsuchiya, R.; Nguyen, T.; Beale, J. M.; Floss, H. G. *J. Am. Chem. Soc.* **1993**, *115*, 7557–7568. (c) Mocek, U.; Zeng, Z.; O'Hagan, D.; Zhou, P.; Fan, L. D. G.; Beale, J. M.; Floss, H. G. *J. Am. Chem. Soc.* **1993**, *115*, 7992–8001.
- (4) (a) Kelly, W. L. *Org. Biomol. Chem.* **2008**, *6*, 4483–4493. (b) Auclair, K.; Sutherland, A.; Kennedy, J.; Witter, D. J.; Van den Heever, J. P.; Hutchinson, C. R.; Vederas, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 11519–11520. (c) Watanabe, K.; Oikawa, H.; Yagi, K.; Ohashi, S.; Mie, T.; Ichihara, A.; Honma, M. *J. Biochem.* **2000**, *127*, 467–473.
- (5) (a) Nicolaou, K. C.; Dethle, D. H.; Leung, G. Y.; Zou, B.; Chen, D. Y. *Chem. Asian J.* **2008**, *3*, 413–429. (b) Nicolaou, K. C.; Zak, M.; Rahimpour, S.; Estrada, A. A.; Lee, S. H.; O'Brate, A.; Giannakakou, P.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2005**, *127*, 15042–15044. (c) Nicolaou, K. C.; Zou, B.; Dethle, D. H.; Li, D. B.; Chen, D. Y. *Angew. Chem., Int. Ed.* **2006**, *45*, 7786–7792. (d) Delgado, O.; Muller, H. M.; Bach, T. *Chemistry* **2008**, *14*, 2322–2339. (e) Muller, H. M.; Delgado, O.; Bach, T. *Angew. Chem., Int. Ed.* **2007**, *46*, 4771–4774. (f) Lefranc, D.; Ciufolini, M. A. *Angew. Chem., Int. Ed.* **2009**, *48*, 4198–4201.
- (6) Brown, L. C.; Acker, M. G.; Clardy, J.; Walsh, C. T.; Fischbach, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2549–2553.
- (7) (a) Xie, L.; Miller, L. M.; Chatterjee, C.; Averin, O.; Kelleher, N. L.; van der Donk, W. A. *Science* **2004**, *303*, 679–681. (b) Willey, J. M.; van der Donk, W. A. *Annu. Rev. Microbiol.* **2007**, *61*, 477–501. (c) Hughes, R. A.; Moody, C. J. *Angew. Chem., Int. Ed.* **2007**, *46*, 7930–7954. (d) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274*, 1188–1193.
- (8) (a) Kelly, W. L.; Pan, L.; Li, C. *J. Am. Chem. Soc.* **2009**, *131*, 4327–4334. (b) Liao, R.; Duan, L.; Lei, C.; Pan, H.; Ding, Y.; Zhang, Q.; Chen, D.; Shen, B.; Yu, Y.; Liu, W. *Chem. Biol.* **2009**, *16*, 141–147. (c) Yu, Y.; Duan, L.; Zhang, Q.; Liao, R.; Ding, Y.; Pan, H.; Wendt-Pienkowski, E.; Tang, G.; Shen, B.; Liu, W. *ACS Chem. Biol.* **2009**, *4*, 855–864.
- (9) (a) Onaka, H.; Tabata, H.; Igarashi, Y.; Sato, Y.; Furumai, T. *J. Antibiot.* **2001**, *54*, 1036–1044. (b) Igarashi, Y.; Kan, Y.; Fujii, K.; Fujita, T.; Harada, K.; Naoki, H.; Tabata, H.; Onaka, H.; Furumai, T. *J. Antibiot.* **2001**, *54*, 1045–1053. (c) Onaka, H.; Nakaho, M.; Hayashi, K.; Igarashi, Y.; Furumai, T. *Microbiology* **2005**, *151* (Pt 12), 3923–3933. (d) Onaka, H. *Biosci. Biotechnol. Biochem.* **2009**, *73*, 2149–2155.
- (10) (a) Acker, M. G.; Bowers, A. A.; Walsh, C. T. *J. Am. Chem. Soc.* **2009**, *131*, 17563–17565. (b) Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc.* **2010**, *132*, 7319–7327.
- (11) Furgerson Ihnken, L. A.; Chatterjee, C.; van der Donk, W. *Biochemistry* **2008**, *47*, 7352–7363.
- (12) (a) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–735. (b) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343. (c) Lee, J.; McIntosh, J.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–2124. (d) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 4089–4091.

JA104524Q