JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Caerulomycins and Collismycins Share a Common Paradigm for 2,2′-Bipyridine Biosynthesis via an Unusual Hybrid Polyketide–Peptide Assembly Logic

Xudong Qu,* Bo Pang, Zhicong Zhang, Ming Chen, Zhuhua Wu, Qunfei Zhao, Qinglin Zhang, Yinyan Wang, Yun Liu, and Wen Liu*

State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China

Supporting Information

ABSTRACT: Caerulomycins (CAEs) and collismycins (COLs), which mainly differ in sulfur decoration, are two groups of structurally similar natural products containing a 2,2′-bipyridine (2,2′-BP) core, derivatives of which have been widely used in chemistry. The biosynthetic pathways of CAEs and COLs remain elusive. In this work, cloning of the CAE biosynthetic gene cluster allowed us to mine a highly conserved gene cluster encoding COL biosynthesis in a *Streptomyces* strain that was previously unknown as a 2,2′-BP producer. In vitro and in vivo investigations into the biosynthesis revealed that CAEs and COLs share a common paradigm featuring an atypical hybrid polyketide synthase/nonribosomal peptide synthetase system that programs the 2,2′-BP formation. This likely involves an unusual intramolecular cyclization/rearrangement sequence, and a difference in processing of the sulfhydryl group derived from the same precursor cysteine drives the biosynthetic route toward CAEs or COLs.

2,2′-Bipyridine (2,2′-BP) has long been appreciated as an attractive core structure of a large class of chelating compounds that are able to form stable complexes with size- or charge-variable metal ions.[1] The application of these heterocyclic ligands is extremely widespread in the areas of coordination chemistry, asymmetric chemistry, materials chemistry, and analytical chemistry, therefore continuously motivating synthetic efforts in methodology development toward efficient functionalization of the 2,2′-BP core for structural diversity. 2,2′-BP also serves as the key molecular scaffold of a number of natural products,[2] including caerulomycins (CAEs), collismycins (COLs), camptothecin, orelline, and streptonigrin, with a remarkable variety of biological activities. Among them, CAEs and COLs are two groups of antibiotics that are structurally similar to each other,[2a−f] characterized by a BP core containing a di- or trisubstituted ring A conjugated with an unmodified ring B (Figure 1). Their major difference is presented at the C5 position of ring A, with (for COLs) or without (for CAEs) decoration of a sulfur-containing group. In contrast to the well-developed synthetic approaches,[3] how nature builds this BP moiety remains poorly understood.

Isotope-labeling experiments were previously carried out on CAEs[4] and showed that the two heterocycles of 2,2′-BP differ in biosynthetic origin. Ring B most likely derives from lysine via a route involving picolinic acid (PA), the carboxylate of which may contribute to C2 of ring A. In contrast, ring A biosynthesis can involve a two-carbon unit, acetate, to build C3 and C4; however, the origins of the atoms C5, C6, and particularly N1 are highly hypothetical: (1) carbon labeling was found by feeding glycerol, but this failed to account for N1; and (2) to address all these atoms, the utilization of serine was suggested yet contradictory to the inefficient incorporation. To elucidate the biosynthetic mechanism of 2,2′-BP, we here set out to clone and characterize the CAE biosynthetic gene cluster from *Actinoalloteichus cyanogriseus* NRRL B-2194 (originally classified as *Streptomyces*). During the analytic process, we identified and subsequently confirmed an overall similar gene cluster that encodes the biosynthesis of the sulfur-containing COL analogues in the genome-sequenced strain *Streptomyces roseosporus* NRRL 11379, which was previously unknown as a 2,2′-BP producer. Comparative analysis of the available *cae* and *col* gene clusters allowed us to trace further the biosynthetic origin of 2,2′-BP in vitro and in vivo, characterization of which revealed the BP-forming generality and provided insights into the sulfhydryl-processing specificity in biosynthesis of these BP-containing natural products.

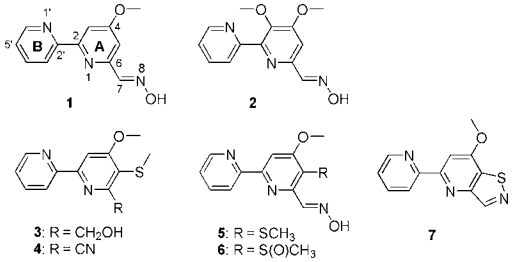

Figure 1. Structures of selected 2,2′-BP natural products. CAEs include the members CAE-A (**1**) and CAE-C (**2**). COLs include SF2738C (**3**, or COL-C), SF2738D (**4**), COL-A (**5**), pyrisulfoxin A (**6**), and SF2738F (**7**).

Taking advantage of the biosynthetic relevance of ring B of 2,2′-BP to PA, we started by exploring the genetic basis of CAE production through cloning of the lysine aminotransferase gene [see the Supporting Information (SI)], the counterpart of

Received: February 19, 2012
Published: May 17, 2012

dx.doi.org/10.1021/ja3016457 | J. Am. Chem. Soc. 2012, 134, 9038–9041

which (*nikC*, 57% identity) participates in PA formation in the biosynthesis of the pyridine-containing antibiotic nikkomycin.[5a] This led to the identification of a cluster of 25 genes from *A. cyanogriseus* (Figure 2A and Figure S3 and Table S3 in the SI).

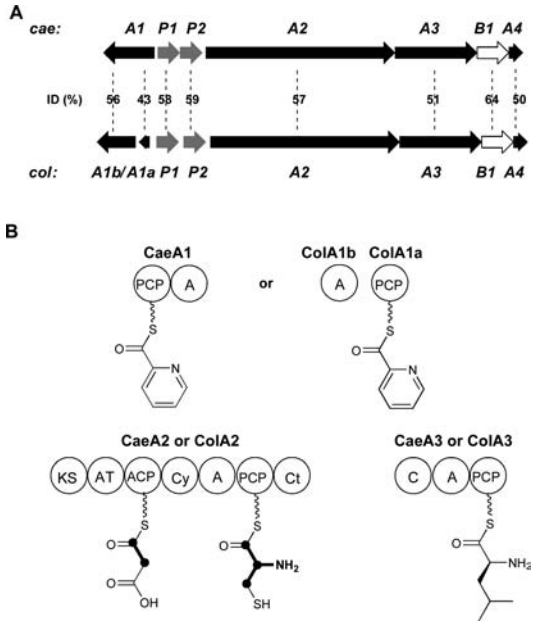

Figure 2. Biosynthetic genes and proteins. (A) Gene organization of the highly conserved PKS/NRPS-encoding cassettes from the *cae* and *col* clusters. The deduced functions (summarized in Tables S3 and S4) are labeled, NRPS or PKS/NRPS shown in black, PA formation in gray, and oxidation in white. ID denotes sequence identity. (B) Domain organization of the hybrid PKS/NRPS system. For the associated building blocks, the parts from the isotope-labeled precursors are highlighted in bold.

Within the cluster, the target lysine aminotransferase gene, *caeP1*, is closely linked with its PA-forming partner *caeP2* [a flavin adenine dinucleotide (FAD)-dependent oxidase gene similar to *nikD* (66% identity)] and flanks the genes putatively encoding a set of nonribosomal peptide synthetases (NRPSs) and a polyketide synthase (PKS)/NRPS hybrid (CaeA1−A4) as well as those involved in oxidoreduction, methyl transfer, amino transfer, hydrolysis, regulation, self-resistance, and others. To correlate the genetic locus with CAE biosynthesis, we inactivated the target gene *caeP1*. Clearly, the resulting mutant strain completely lost the ability to produce CAE (Figure 3A, II). This result validated the indispensability of CaeP1 and supported the proposal that ring B of 2,2′-BP originates from lysine, conversion of which into PA is necessary before the incorporation. The catalysis may be similar to that in nikkomycin biosynthesis, requiring the oxidative amino transfer activity of CaeP1 and subsequent CaeP2-catalyzed four-electron oxidation.[5b] Additionally, the hybrid PKS/NRPS system-encoding genes surround the PA-forming genes within the *cae* cluster, suggesting that biosynthetically CAEs are a group of polyketide−peptide hybrid natural products.

During functional assignment of the CAE biosynthetic genes, we unexpectedly found two identical gene clusters (Figure S3 and Table S4), overall similar to the *cae* cluster, individually in the genome-available *S. roseosporus* strains numbered as NRRL 11379 and NRRL 15998, which were previously known as the daptomycin antibiotic producers.[6] In particular, the PKS/NRPS-encoding cassette (denoted as *caeA1*, *P1*, *P2*, *A2*, *A3*, *B1*, and *A4* in the *cae* cluster) is highly conserved (43−64% identity) in sequence and gene organization (Figure 2A), with the only exception that the didomains of the NRPS gene *caeA1* are present as two discrete genes, namely, *colA1a* and *colA1b*. This finding strongly implied that the *S. roseosporus* strains have the potential to produce polyketide−peptide hybrid, CAE-like 2,2′-BP natural products. To validate the hypothesis, we chose the NRRL 11379 strain for fermentation and product examination. Indeed, HPLC analysis revealed a group of products with the patterns of UV absorbance nearly identical to those of CAEs (Figure 3A, IV). Next, five relevant compounds **3**−**7** (Figure 1) were selectively purified and then subjected to high-resolution electrospray ionization mass spectrometry and ^{1}H NMR spectral analyses (see the SI), which showed that they are COL members of the 2,2′-BP family that differ from CAEs mainly in the sulfur-containing side chain at C5 of ring A. Finally, in-frame deletion of *colA2*, the largest PKS/NRPS gene homologous to *caeA2* in CAE biosynthesis, was performed in *S. roseosporus* and resulted in the complete abolishment of COL

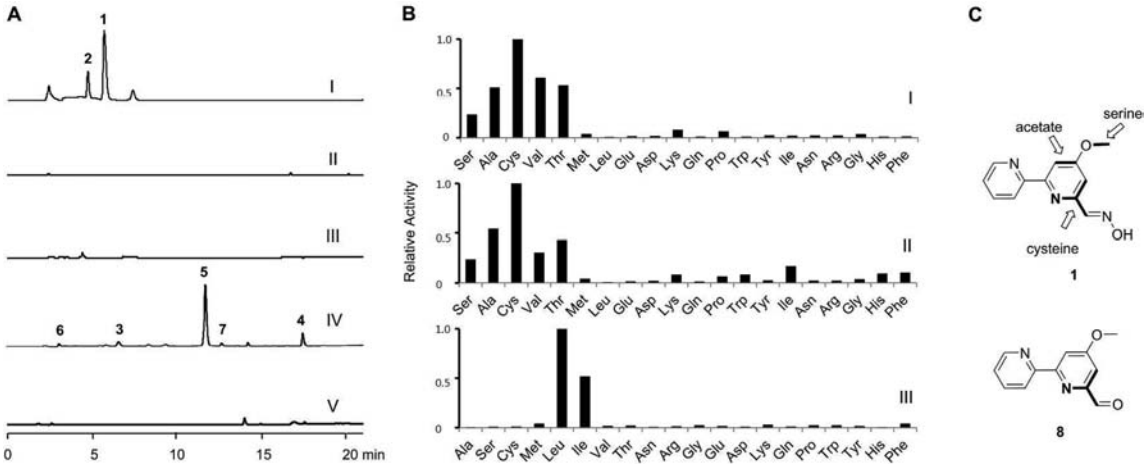

Figure 3. In vivo and in vitro investigations. (A) HPLC analysis of the production. For CAEs: (I) wild-type *A. cyanogriseus* strain; (II) mutant strain QL2001 (Δ*caeP1*); (III) mutant strain QL2002 (Δ*caeA3*). For COLs: (IV) wild-type *S. roseosporus* strain; (V) mutant strain QL2003 (Δ*colA2*). (B) Assay of A-domain activity by individual use of 20 amino acids: (I) CaeA2-Cy-A-PCP-Ct; (II) ColA2-A; (III) CaeA3-A. (C) Isotope-labeling patterns (in bold) shown on compound **1** and its derivative aldehyde **8**.

dx.doi.org/10.1021/ja3016457 | *J. Am. Chem. Soc.* 2012, 134, 9038−9041

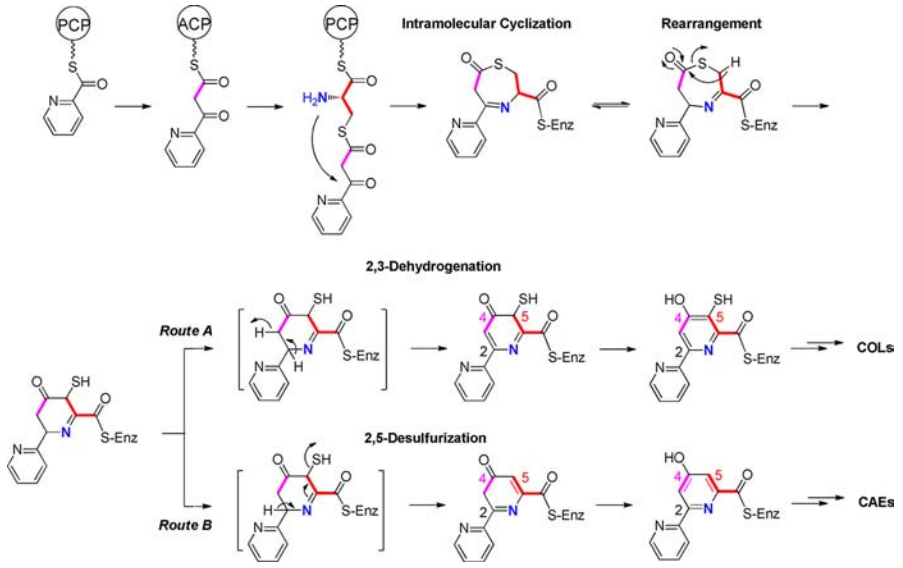

Figure 4. Proposed biosynthetic pathway of 2,2′-BP in CAE and COL biosynthesis. The colors indicate the parts derived from the isotope-labeled precursors, including [1,2-$^{13}C_2$]acetate (pink) and L-[1,2,3-$^{13}C_3$,^{15}N]cysteine (red). The nitrogen atom from the latter is shown in blue.

production (Figure 3A, V), clearly confirming that this identified cluster governs COL biosynthesis.

Benefiting from two available gene clusters, we then focused on the analysis of the hybrid PKS/NRPS system that comprises the proteins potentially responsible for scaffold formation of 2,2′-BP in both CAEs and COLs. CaeA1−A3 and ColA1−A3 have a head-to-tail sequence homology and an identical domain organization (Figure 2B), supporting the idea that CAEs and COLs share a common paradigm for assembling the skeleton in a PKS/NRPS catalytic logic. CaeA1 is a bifunctional enzyme containing a peptidyl carrier protein (PCP, as the discrete protein ColA1a in COL biosynthesis) and an adenylation (A) domain (ColA1b), presumably for activating PA as a starter unit to initiate construction of ring B first. This is consistent with the substrate-specificity-coffering codes of the A domain, CaeA1-A/ColA1b, being nearly identical to those of SanJ,[5c] which has been shown to incorporate PA into nikkomycin biosynthesis. CaeA2/ColA2 is an atypical PKS/NRPS hybrid protein, domains of which are arranged in the following order: ketosynthase (KS), acyltransferase (AT), acyl carrier protein (ACP), condensation/cyclization (Cy) domain, A domain, PCP, and terminal C (Ct) domain. The substrate specificity was accordingly predicted, indicating that CaeA2-AT/ColA2-AT is specific for malonyl-CoA and that CaeA2-A/ColA2-A prefers to activate cysteine. An additional NRPS protein, CaeA3/ColA3, is composed of the C-A-PCP domains, in which the A domain has the key coffering codes similar to those specifically for selecting threonine. On the basis of these results, we reasoned that CaeA2/ColA2 most likely employs cysteine as the extender unit for constructing the uncharacterized atoms C5, C6, and N1 of the conjugated ring A of 2,2′-BP according to known hybrid PKS/NRPS assembly manners.[7]

To provide the experimental evidence, we overexpressed the truncated proteins CaeA2-A, CaeA2-Cy-A-PCP-Ct, and ColA2-A in and purified them from *Escherichia coli* (Figure S5) and then examined the substrate specificity upon the reaction of the A domain by using the well-developed [^{32}P]ATP-labeled PPi exchange assay method.[8] As expected, the highest activity was readily observed in the CaeA2-Cy-A-PCP-Ct or ColA2-A-catalyzed reaction when cysteine served as the substrate (Figure 3B); the activity of the single A domain CaeA-2 failed to be detected (data not shown). This finding suggested that cysteine is a common precursor in the biosyntheses of CAEs and COLs; however, two interesting questions arose here regarding the diversity between CAEs and COLs and the cysteine-involved level: (1) it is readily understood that cysteine can be built into the sulfur-containing COLs, but the incorporation of cysteine into the sulfur-lacking CAEs needs further proof; and (2) if so, whether cysteine alone is sufficient for the formation of ring A remains elusive because of the presence of the NRPS CaeA3 in CAE biosynthesis.

We thus put the attention back to the atom construction of ring A of 2,2′-BP in CAE biosynthesis. Isotope-labeling experiments were carried out by individually feeding [1,2-$^{13}C_2$]-acetate, L-[1,2,3-$^{13}C_3$,^{15}N]serine, and L-[1,2,3-$^{13}C_3$,^{15}N]cysteine into the wild-type CAE-producing strain *A. cyanogriseus* and then spectrally analyzing the major product CAE-A (**1**). The feedings of acetate and serine produced labeling patterns similar to those found previously (Figure 3C and the SI),[4] consistent with the PKS activity of CaeA2 and its AT domain selectivity of malonyl-CoA to build C3 and C4 and excluding the participation of serine in ring A formation. Remarkably, the feeding of cysteine resulted in high ^{13}C-enrichment (13 ± 1%) at C5 and C6 in ring A as well as at C7 (Table S5 and Figure S10) that resides in the C6-substituted side chain of **1**. The ^{15}N NMR spectrum of **1** revealed an enriched signal at 284 ppm (Figure S12), which falls into the range of the pyridine nitrogen (usually shown at 200−300 ppm). To exclude further the possibility that the labeling takes place at N8 of the oximyl side chain, we performed chemical degradation of **1** and generated analogue **8** in an aldehyde form by removing oximyl. **8** was subjected to NMR spectral analysis, which indeed showed that C5, C6, C7, and only nitrogen atom N1 were efficiently labeled and spectrally well-coupled (for C−C coupling, C5−C6, J_{CC} = 60 Hz; C6−C7, J_{CC} = 60 Hz; and C5−C7, J_{CC}= 7 Hz; for C−N coupling, N1−C6, J_{CN} = 2 Hz; N1−C5, J_{CN} = 3 Hz; and N1−C7, J_{CN} = 14 Hz) (Figure S13). These findings unambiguously validated that cysteine is completely incorporated by the NRPS activity of CaeA2 into ring A of 2,2′-BP to set up the remaining atoms N1, C5, and C6 as well as the outside C7 (Figure 3C).

dx.doi.org/10.1021/ja3016457 | J. Am. Chem. Soc. 2012, 134, 9038−9041

Taking all of the above evidence into account, we consequently propose the biosynthetic mechanism of the 2,2′-BP core of CAEs and COLs and account for their difference in sulfur decoration (Figure 4). CaeA1/ColA1 may function as a loading module to initiate the skeleton assembly by activating the starter unit PA for providing ring B. The hybrid PKS/NRPS protein CaeA2/ColA2 can sequentially condense the building blocks malonyl-CoA and cysteine to afford a full length linear intermediate for ring A formation. In view of the fact that the amino group of cysteine is eventually conjugated with the carbonyl of the PA moiety to form a C2−N1 linkage, nucleophilic addition for chain elongation likely occurs between the sulfhydryl of cysteine and the carbonyl resulting from malonyl-CoA via a two-carbon extension ahead, forming a thioester bond. This intermediate may undergo an intramolecular cyclization to give a sulfur-containing seven-membered heterocycle, subsequent rearrangement of which could rerelease the SH group and provide a six-membered, potentially common intermediate in CAE and COL biosynthesis. At this point, further tailoring may be diverse in the pattern for processing the SH group: (1) 2,3-dehydrogenation results in the sulfur-containing ring A for generating the COL members (route A), and (2) 2,5-desulfurization affords the sulfur-lacking ring A that eventually leads to production of the CAE members (route B).

Functional association of CaeA2/ColA2 with CaeA1/ColA1 is apparently enough to construct the hybrid polyketide−peptide backbone of 2,2′-BP, leaving a query regarding the role of the additional NRPS CaeA3/ColA3 in CAE and COL biosynthesis. The participation of CaeA3 in CAE biosynthesis was subsequently confirmed, as the corresponding gene mutant strain failed to produce CAEs (Figure 3A, III). Following the procedure described above, we expressed and purified the A domain CaeA3-A for substrate determination. The PPi exchange assay clearly indicated that CaeA3-A highly selected leucine (Figure 3B), correcting the initial assignment as threonine based solely upon sequence analysis. An additional peptidyl extension catalyzed by CaeA3/ColA3 can be speculated to occur on a hybrid polyketide−peptide intermediate for appending the side chain at C6 of 2,2′-BP. The process may involve cleavage and further modifications to render the variable form (e.g., the oximyl group rarely found in current known natural products). This is consistent with the fact that a number of highly conserved tailoring genes were found in both the *cae* and *col* clusters (Tables S3 and S4).

In conclusion, we have uncovered a common paradigm featuring an unusual hybrid PKS/NRPS system for biosynthesis of the 2,2′-BP core structure that is shared by sulfur-containing COLs and sulfur-lacking CAEs. We discovered this pathway by first cloning the *cae* cluster from *A. cyanogriseus* and then validating the generality by genome mining and ultimate confirmation of COL production in *S. roseosporus*, which was previously known to produce daptomycin antibiotics but unknown as a 2,2′-BP natural product producer. In vitro and in vivo determination of the biosynthetic origins supported the conclusion that lysine-derived PA, malonyl-CoA, and particularly cysteine are employed by this PKS/NRPS system, likely via an unprecedented cyclization/rearrangement sequence, to construct the pyridine heterocycles of 2,2′-BP. This core may then be diverted toward CAEs or COLs through different SH-processing pathways. During the revision of this manuscript, Salas and co-workers published a COL biosynthetic gene cluster from *Streptomyces* sp. CS40, which is highly homologous to that mined in *S. roseosporus*.[9] Comparison of our hypothesis regarding the 2,2′-BP-forming chemistry to theirs will facilitate the ongoing experimental validation to address the interesting enzymatic mechanism.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Experimental details and characterization data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

xudongqu@gmail.com; wliu@mail.sioc.ac.cn

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by grants from NNSF (30900021, 20832009 and 20921091), SRF for 39th ROCS-SEM, the 973 Program (2010CB833200 and 2012CB721100), and CAS (KJCX2-YW-201) of China.

■ REFERENCES

(1) (a) Kaes, C.; Katz, A.; Hosseini, M. W. *Chem. Rev.* **2000**, *100*, 3553. (b) Hapke, M.; Brandt, L.; Lutzen, A. *Chem. Soc. Rev.* **2008**, *37*, 2782.

(2) (a) Divekar, P. V.; Read, G.; Vining, L. C. *Can. J. Chem.* **1967**, *45*, 1215. (b) McInnes, A. G.; Smith, D. G.; Wright, J. L. C.; Vining, L. C. *Can. J. Chem.* **1977**, *55*, 4159. (c) Fu, P.; Wang, S.; Hong, K.; Li, X.; Liu, P.; Wang, Y.; Zhu, W. *J. Nat. Prod.* **2011**, *74*, 1751. (d) Shindo, K.; Yamagishi, Y.; Okada, Y.; Kawai, H. *J. Antibiot.* **1994**, *47*, 1072. (e) Gomi, S.; Amano, S.; Sato, E.; Miyadoh, S.; Kodama, Y. *J. Antibiot.* **1994**, *47*, 1385. (f) Tsuge, N.; Furihata, K.; Shin-Ya, K.; Hayakawa, Y.; Seto, H. *J. Antibiot.* **1999**, *52*, 505. (g) Wall, M. E.; Wani, M. C.; Cook, C. E.; Palmer, K. H.; McPhail, A. T.; Sim, G. A. *J. Am. Chem. Soc.* **1966**, *88*, 3888. (h) Antkowiak, W. Z.; Gessner, W. P. *Tetrahedron Lett.* **1979**, *20*, 1931. (i) Rao, K. V.; Biemann, K.; Woodward, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2532.

(3) (a) Mongin, F.; Trécourt, F.; Gervais, B.; Mongin, O.; Quéguiner, G. *J. Org. Chem.* **2002**, *67*, 3272. (b) Sammakia, T.; Stangeland, E. L.; Whitcomb, M. C. *Org. Lett.* **2002**, *4*, 2385. (c) Duan, X.-F.; Ma, Z.-Q.; Zhang, F.; Zhang, Z.-B. *J. Org. Chem.* **2008**, *74*, 939. (d) Dash, J.; Reissig, H.-U. *Chem.—Eur. J* **2009**, *15*, 6811. (e) Bobrov, D. N.; Tyvorskii, V. I. *Tetrahedron* **2010**, *66*, 5432. (f) Zhang, F.; Duan, X.-F. *Org. Lett.* **2011**, *13*, 6102.

(4) (a) McInnes, A. G.; Smith, D. G.; Walter, J. A.; Vining, L. C.; Wright, J. L. C. *Can. J. Chem.* **1979**, *57*, 3200. (b) Vining, L. C.; McInnes, A. G.; McCulloch, A. W.; Smith, D. G.; Walter, J. A. *Can. J. Chem.* **1988**, *66*, 191.

(5) (a) Bruntner, C.; Bormann, C. *Eur. J. Biochem.* **1998**, *254*, 347. (b) Bruckner, R. C.; Zhao, G.; Venci, D.; Jorns, M. S. *Biochemistry* **2004**, *43*, 9160. (c) Niu, G.; Liu, G.; Tian, Y.; Tan, H. *Metab. Eng.* **2006**, *8*, 183.

(6) Miao, V.; Coëffet-LeGal, M.-F.; Brian, P.; Brost, R.; Penn, J.; Whiting, A.; Martin, S.; Ford, R.; Parr, I.; Bouchard, M.; Silva, C. J.; Wrigley, S. K.; Baltz, R. H. *Microbiology* **2005**, *151*, 1507.

(7) Walsh, C. T.; Fischbach, M. A. *J. Am. Chem. Soc.* **2010**, *132*, 2469.

(8) Linne, L.; Walsh, C. T. *Methods Enzymol.* **2004**, *388*, 293.

(9) Garcia, I.; Vior, N. M.; Gonzalez-Sabin, J.; Rohr, J.; Moris, F.; Mendez, C.; Salas, J. A. *Chem. Biol.* **2012**, *19*, 399.