ORGANIC LETTERS

2012
Vol. 14, No. 11
2666–2669

Identification of Caerulomycin A Gene Cluster Implicates a Tailoring Amidohydrolase

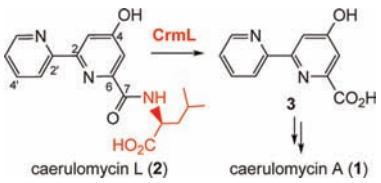
Yiguang Zhu,[†,§] Peng Fu,[‡,§] Qinheng Lin,[†] Guangtao Zhang,[†] Haibo Zhang,[†] Sumei Li,[†] Jianhua Ju,[†] Weiming Zhu,[*,‡] and Changsheng Zhang[*,†]

CAS Key Laboratory of Marine Bio-resources Sustainable Utilization, RNAM Center for Marine Microbiology, Guangdong Key Laboratory of Marine Materia Medica, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 164 West Xingang Road, Guangzhou 510301, China, and Key Laboratory of Marine Drugs, Ministry of Education of China, School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China

czhang2006@gmail.com; weimingzhu@ouc.edu.cn

Received March 7, 2012

ABSTRACT



The biosynthetic gene cluster for caerulomycin A (1) was cloned and characterized from the marine actinomycete *Actinoalloteichus cyanogriseus* WH1-2216-6, which revealed an unusual hybrid polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) system. The *crmL* disruption mutant accumulated caerulomycin L (2) with an extended L-leucine at C-7, implicating an amidohydrolase activity for CrmL. The leucine-removing activity was confirmed for crude CrmL enzymes. Heterologous expression of the 1 gene cluster led to 1 production in *Streptomyces coelicolor*.

Caerulomycins were first discovered from *Streptomyces caeruleus*[1] and constituted a unique group of 2,2′-bipyridyl ring-containing natural products, also represented by collismycins, pyrisulfoxins, streptonigrin, and orelline.[2] Recently, caerulomycin A (**1**, Figure 1B) and its numerous novel analogues were isolated from a marine actinomycete, *Actinoalloteichus cyanogriseus* WH1-2216-6.[3] Caerulomycins were found to have strong antifungal, antiamoebic, and antitumor activities, and mild antibacterial activities.[3,4] Caerulomycin A (**1**) was also demonstrated to have novel bioactivity with remarkable promise in immunosuppression.[5] The unique structure of caerulomycins and their significant bioactivities attracted many synthetic efforts, with accomplishments of the total synthesis of caerulomycins A-C and E.[6] The biosynthetic studies of **1** were limited to earlier feeding experiments with labeled putative precursors,[7] revealing the origin of the first pyridine ring from lysine

[†] South China Sea Institute of Oceanology.
[‡] Ocean University of China.
[§] Both authors contributed equally to this work.

(1) (a) Funk, A.; Divekar, P. V. *Can. J. Microbiol.* **1959**, *5*, 317–321. (b) McInnes, A. G.; Smith, D. G.; Wright, J. L. C.; Vining, L. C. *Can. J. Chem.* **1977**, *55*, 4159–4165. (c) McInnes, A. G.; Smith, D. G.; Walter, J. A.; Wright, J. L. C.; Vining, L. C.; Arsenault, G. P. *Can. J. Chem.* **1978**, *56*, 1836–1842. (d) Divekar, P. V.; Read, G.; Vining, L. C. *Can. J. Chem.* **1967**, *45*, 1215–1223.

(2) (a) Shindo, K.; Yamagishi, Y.; Okada, Y.; Kawai, H. *J. Antibiot.* **1994**, *47*, 1072–1074. (b) Gomi, S.; Amano, S.; Sato, E.; Miyadoh, S.; Kodama, Y. *J. Antibiot.* **1994**, *47*, 1385–1394. (c) Tsuge, N.; Furihata, K.; Shin-Ya, K.; Hayakawa, Y.; Seto, H. *J. Antibiot.* **1999**, *52*, 505–507. (d) Antkowiak, W. Z.; Gessner, W. P. *Tetrahedron Lett.* **1979**, *20*, 1931–1934. (e) Rao, K. V.; Biemann, K.; Woodward, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2532–2533.

(3) (a) Fu, P.; Wang, S.; Hong, K.; Li, X.; Liu, P.; Wang, Y.; Zhu, W. *J. Nat. Prod.* **2011**, *74*, 1751–1756. (b) Fu, P.; Liu, P.; Li, X.; Wang, Y.; Wang, S.; Hong, K.; Zhu, W. *Org. Lett.* **2011**, *13*, 5948–5951.

(4) Chatterjee, D. K.; Raether, W.; Iyer, N.; Ganguli, B. N. *Z. Parasitenkd.* **1984**, *70*, 569–573.

(5) Singla, A. K.; Agrewala, J. N.; Vohra, R. M.; Jolly, R. S. United States Patent US 20070078167, 2007.

(6) (a) Trecourt, F.; Gervais, B.; Mallet, M.; Queguiner, G. *J. Org. Chem.* **1996**, *61*, 1673–1676. (b) Mongin, F.; Trecourt, F.; Gervais, B.; Mongin, O.; Queguiner, G. *J. Org. Chem.* **2002**, *67*, 3272–3276. (c) Bobrov, D. N.; Tyvorskii, V. I. *Tetrahedron* **2010**, *66*, 5432–5434.

(7) (a) McInnes, A. G.; Smith, D. G.; Walter, J. A.; Vining, L. C.; Wright, J. L. C. *Can. J. Chem.* **1979**, *57*, 3200–3204. (b) Vining, L. C.; McInnes, A. G.; McCulloch, A. W.; Smith, D. G.; Walter, J. A. *Can. J. Chem.* **1988**, *66*, 191–194.

10.1021/ol300589r
Published on Web 05/16/2012

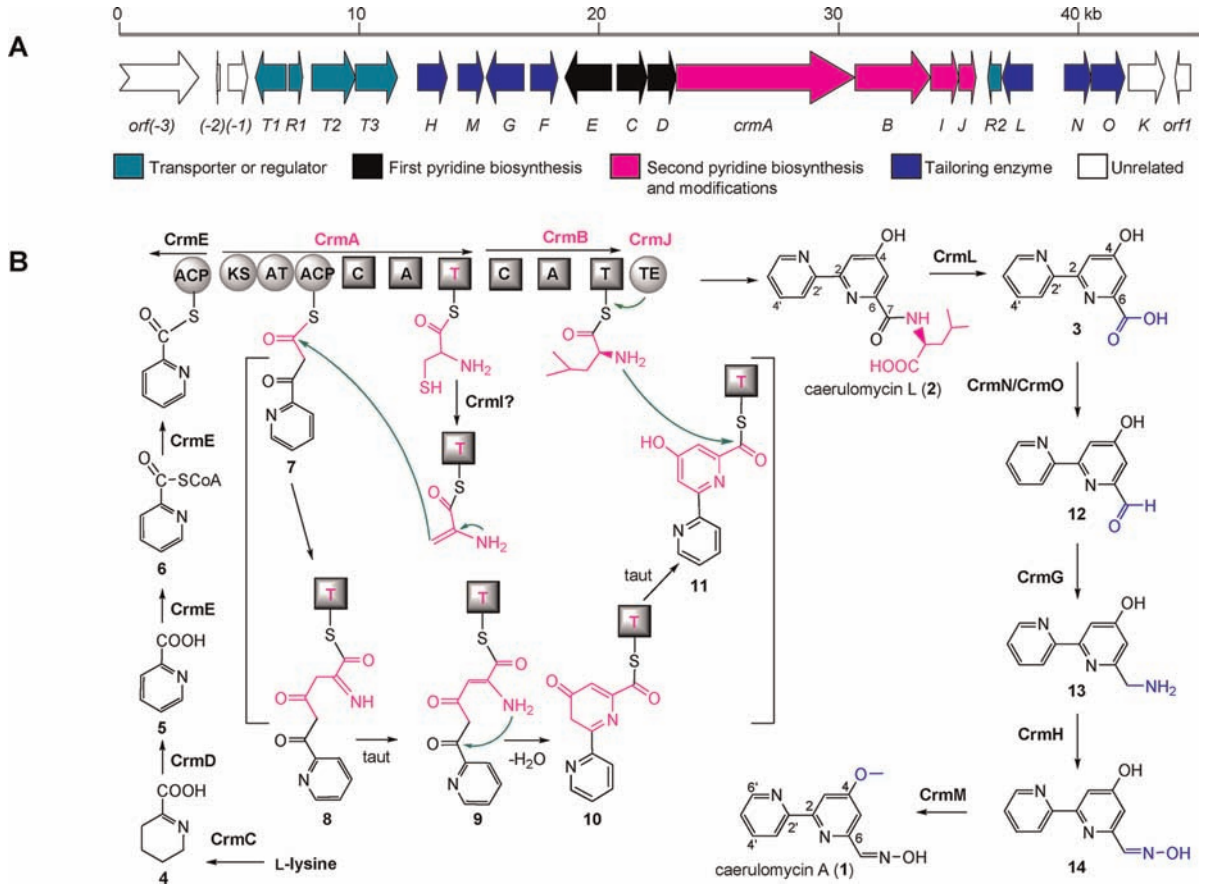


Figure 1. Biosynthesis of **1** in *A. cyanogriseus* WH1-2216-6. (A) Genetic organization of the **1** biosynthetic gene cluster. (B) Proposed biosynthetic pathway for **1**. The pathway starts from L-lysine, and the bracketed steps are hypothetical. Abbreviations: ACP, acyl-carrier protein; KS, ketosynthase; AT, acyltransferase; C, condensation; A, adenylation; T, thiolation; TE, thioestrase; taut, tautomerization.

via picolinic acid as an intermediate.[7] However, the origin and mechanism for the second pyridine ring formation remained elusive. Herein we reported the cloning and characterization of the **1** biosynthetic gene cluster from *A. cyanogriseus* WH1-2216-6 and the heterologous expression of the **1** gene cluster in *Streptomyces coelicor*.

The initiation of the first pyridine ring formation in **1** was proposed to be catalyzed by L-lysine 2-aminotransferase, similar to NikC in the nikkomycin pathway and VisA in the virginiamycin S pathway.[8] Using a pair of degenerate primers nikC-F and nikC-R (Table S1, Supporting Information), which were designed to target on conserved regions of NikC and its analogues (Figure S1, Supporting Information), we successfully identified seven positive cosmids from 1920 clones of a SuperCos1-based genomic library of *A. cyanogriseus* WH1-2216-6 (Table S2, Supporting Information). After restriction analyses and end-sequencing (Figure S2, Supporting Information), the cosmid pCSG2016 was chosen for entire sequencing by a shot-gun approach, to give a 44,640 bp of contiguous DNA sequence (GenBank accession no. JF419316). Bioinformatic analysis revealed the presence of 24 orfs (Figure 1A), 20 of which were proposed to be involved in **1** biosynthesis (Table S3, Supporting Information).

Three genes, *crmCDE*, are proposed to be responsible for the formation of the first pyridine ring in **1**. CrmC, an aminotransferase like NikC,[9] may convert L-lysine to the corresponding α-keto derivative, which spontaneously cyclizes and dehydrates to yield piperidine-2-carboxylic acid. Next, the flavoenzyme CrmD, likely catalyzes a 4-electron oxidation of piperidine-2-carboxylic acid to picolinic acid, by analogy to its closest homologue NikD, a biochemically and structurally well studied amino acid oxidase.[10] To verify the involvement of *crmD* in **1** biosynthesis, the *crmD* gene was disrupted by utilizing a genetic manipulation system previously established for *A cyanogriseus* WH1-2216-6 (Figure S3, Supporting Information).[11]

(8) (a) Bruntner, C.; Bormann, C. *Eur. J. Biochem.* **1998**, *254*, 347–355. (b) Namwat, W.; Kinoshita, H.; Nihira, T. *J. Bacteriol.* **2002**, *184*, 4811–4818.

(9) Bruntner, C.; Lauer, B.; Schwarz, W.; Mohrle, V.; Bormann, C. *Mol. Gen. Genet.* **1999**, *262*, 102–114.

(10) (a) Bruckner, R. C.; Zhao, G.; Venci, D.; Jorns, M. S. *Biochemistry* **2004**, *43*, 9160–9167. (b) Carrell, C. J.; Bruckner, R. C.; Venci, D.; Zhao, G.; Jorns, M. S.; Mathews, F. S. *Structure* **2007**, *15*, 928–941.

(11) Lin, Q.; Zhang, G.; Li, S.; Zhang, H.; Ju, J.; Zhu, W.; Zhang, C. *Acta Mirobiol. Sin.* **2011**, *51*, 1032–1041.

The resulting Δ*crmD* mutant CRM02 abolished **1** production, which could be chemically rescued by feeding picolinic acid (Figure S4, Supporting Information), confirming CrmD as a picolinic acid synthase. Subsequently, CrmE activates picolinic acid to its CoA form, in an ATP-dependent manner similar to its biochemically verified homologues, NikE and SanJ.[12] Notably, careful bioinformatic analysis revealed an ACP domain embedding at the N-terminus of CrmE (Figure S5, Supporting Information).

Four genes, *crmABIJ*, are probably involved in forming and modifying the second pyridine ring. CrmA is predicted as a hybrid polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) protein, containing typical domains of ketosynthase (KS), acyltransferase (AT), acyl-carrier protein (ACP), condensation (C), adenylation (A), and thiolation (T). CrmB contains C-A-T domains typical for a NRPS protein. Bioinformatic analyses of characteristic features of active site motifs predicted specificities of AT domain of CrmA for malonyl-CoA (QQGHSLGRFHTHV), A domain of CrmA for cysteine (DLFNMSLIWK) and A domain of CrmB for threonine (DYWSFGLVVK).[13] CrmI exhibits the highest similarity to a putative acyl-CoA dehydrogenase from *S. roseosporus* NRRL 11379 and CrmJ is a predicted thioesterase (TE) involved in nonribosomal peptide biosynthesis with the highest identity to the thioesterase GrsT from *Moorea producta*.[14] Most intriguingly, there are together two NRPS modules in CrmA and CrmB, theoretically capable of adding two amino acids during **1** biosynthesis, which is highly contradicted to the chemical structure of **1**. Our subsequent gene knockout experiments revealed that all four mutants, including CRM06 (Δ*crmA*), CRM09 (Δ*crmB*), CRM10 (Δ*crmI*), and CRM12 (Δ*crmJ*) lost the ability to produce **1** (Figure 2A−C, Figure S6AB, Supporting Information), proving their essentiality for **1** biosynthesis.

Seven genes, *crmFGHMLNO*, are putatively involved in post-PKS/NRPS modifications. CrmN and CrmO resemble GirC- and GriD-like dehydrogenase components, respectively. They may function together to reduce a carboxyl group into its aldehyde derivative, as demonstrated for carboxylic acid reductase pairs GriC and GriD in grixazone biosynthesis.[15] CrmG is predicted to be a diaminobutyrate−pyruvate aminotransferase, which may be responsible for incorporation of the amino group at C-7 in **1**. CrmF shows the highest similarity to a putative FAD-binding mono-oxygenase Oant_4393 from *Ochrobactrum anthropi*.[16] CrmH distantly resembles the *N*-hydroxylase VlmH (30% identity) in valanimycin biosynthesis.[17] Inactivation of *crmF* did not affect **1** production (Figure S6C, Supporting Information), while the Δ*crmH* mutant failed to produce **1** (Figure S6D, Supporting Information), indicating that CrmF is not essential for **1** biosynthesis and CrmH is possibly functional to add the *N*-hydroxyl group in **1**. CrmM shows above 40% identity to a number of *O*-methyltransferases involved in natural product biosynthesis, such as AzicL for azicemicin[18] and CalO6 for calicheamicin.[19] It may methylate the hydroxyl group at C-4 in **1**.

CrmL is predicted to be a metallo-dependent amidohydrolase (Figure S7, Supporting Information), displaying the highest identity (58%) to SSGG_05706 from *S. roseosporus* NRRL 15998. Interestingly, the Δ*crmL* mutant CRM04 could still produce **1**, but also accumulated another metabolite (Figure 2D), which was isolated and identified as a new compound caerulomycin L (**2**, Figure 1B) by ^{1}H, ^{13}C NMR and HRESIMS (Figure S8, Table S4, Supporting Information). Caerulomycin L (**2**) was assigned a molecular formula of $C_{17}H_{19}N_3O_4$ on the basis of HRESIMS, requiring 10 degrees of unsaturation. The ^{1}H NMR spectrum also displayed the characteristic signals of a 2-disubstituted pyridine ring system at δ 8.66, 8.48, 7.92, and 7.43 and a 2,4,6-trisubstituted pyridine ring system at δ 7.92 (s) and 7.46 (s), indicating that **2** is a 4,6-disubstituted 2,2′-bipyridine alkaloid. HMBC correlations from H-4′ to C-2′/C-6′, H-6′ to C-2′/C-5′, H-3 to C-5/C-4/C-2, and from H-5 to C-6/C-7 further identified a 4-hydroxy-2,2′-bipyridine-6-carboxy structure. The alkyl signals at $\delta_{H/C}$ 0.88/22.9 (CH_3), 0.91/23.7 (CH_3), 1.68/25.3 (CH_2), 1.68/42.6 (CH_2) and 4.31/52.9 (CH), and a carboxyl signal at δ_C 175.7 in **2** suggested the presence of a leucine fragment that was further supported by ^{1}H−^{1}H COSY (Figure S9, Table S4, Supporting Information). The acid hydrolysis of **2** yielded leucine and 4-hydroxy-2,2′-bipyridine-6-carboxylic acid (**3**, Figure 1B) which was identified by ^{1}H, ^{13}C NMR and HRESIMS (Figure S10, Table S4, Supporting Information). The absolute configuration of leucine was determined as L- by Marfey's method (Figure S11, Supporting Information).[20]

The incomplete *orf(-3)* gene, *orf(-2)*, and *orf(-1)* genes encode functions not relevant to **1** biosynthesis, and the *crmK* and *orf1* genes both encode flavoenzymes (Table S3, Supporting Information). The Δ*crmK* mutant CRM11 still produced **1**, excluding its essential role for **1** biosynthesis (Figure S6E, Supporting Information). There are three putative transporter genes *crmT1−T3* and two putative regulator genes *crmR1-R2*, they may be responsible for transportation, resistance, and biosynthetic regulation of **1**.

To confirm that we have cloned the entire **1** gene cluster, the recombinant cosmid pCSG2307 harboring our whole sequenced DNA region (Table S2, Supporting Information) was introduced into *Streptomyces coelicolor* YF11,[21] successfully leading to the heterologus production of **1**

(12) (a) Niu, G. Q.; Liu, G.; Tian, Y. Q.; Tan, H. R. *Metab. Eng.* **2006**, *8*, 183–195. (b) Moon, M.; Van Lanen, S. G. *Biopolymers* **2010**, *93*, 791–801.

(13) Ansari, M. Z.; Yadav, G.; Gokhale, R. S.; Mohanty, D. *Nucleic Acids Res.* **2004**, *32*, W405–413.

(14) Jones, A. C.; Monroe, E. A.; Podell, S.; Hess, W. R.; Klages, S.; Esquenazi, E.; Niessen, S.; Hoover, H.; Rothmann, M.; Lasken, R. S.; Yates, J. R., 3rd; Reinhardt, R.; Kube, M.; Burkart, M. D.; Allen, E. E.; Dorrestein, P. C.; Gerwick, W. H.; Gerwick, L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 8815–8820.

(15) Suzuki, H.; Ohnishi, Y.; Horinouchi, S. *J. Antibiot.* **2007**, *60*, 380–387.

(16) Chain, P. S.; Lang, D. M.; Comerci, D. J.; Malfatti, S. A.; Vergez, L. M.; Shin, M.; Ugalde, R. A.; Garcia, E.; Tolmasky, M. E. *J. Bacteriol.* **2011**, *193*, 4274–4275.

(17) Parry, R. J.; Li, W.; Cooper, H. N. *J. Bacteriol.* **1997**, *179*, 409–416.

(18) Ogasawara, Y.; Liu, H. W. *J. Am. Chem. Soc.* **2009**, *131*, 18066–18068.

(19) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S. *Science* **2002**, *297*, 1173–1176.

(20) Marfey, P. *Carlsberg. Res. Commun.* **1984**, *49*, 591–596.

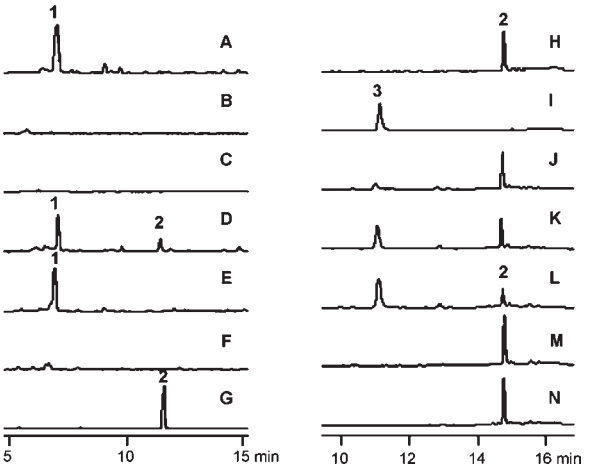


Figure 2. HPLC analyses of metabolite profiles and crude enzyme assays. (A) *A. cyanogriseus* WH1-2216-6 wild type; (B) Δ*crmA* mutant CRM06; (C) Δ*crmB* mutant CRM09; (D) Δ*crmL* mutant CRM04; (E) *S. coelicolor* YF11/pCSG2307; (F) YF11/pSET152; (G) YF11/pCSG2308; (H) and (I) **2** and **3** standards, (J–L) Incubation of **2** with crude extracts of YF11/pCSG2220 for 1 h (J), 4 h (K) and 8 h (L); control assays by incubation of **2** with YF11/pPWW50A (M) or bioled crude extracts of YF11/pCSG2220 (N) for 8 h. UV detection at 254 nm. The HPLC mobile phases were acetonitrile for traces A–G and methanol for traces H–N.

(Figure 2E), which was absent in the control strain (Figure 2F). Unlike the Δ*crmL* mutant CRM04, *S. coelicolor* YF11/pCSG2308, in which the *crmL* gene was disrupted (Table S2, Supporting Information), exclusively produced **2** but not **1** (Figure 2G). These observations indicated that CrmL specifically removed L-leucine from **2** and the existence of another enzyme outside of **1** gene cluster in *A. cyanogriseus*, performing a similar function as CrmL.

No soluble CrmL could be detected in *E. coli*, and only weak expression of *crmL* was achieved in *S. coelicolor* YF11/pCSG2220 (Figure S12, Supporting Information). Fortunately, **2** was indeed converted to **3** in *in vitro* assays using cell free extracts of YF11/pCSG2220 and the conversion ratio was increased with longer incubation time (Figure 2H-L, Figure S13, Supporting Information), while **2** remained unchanged in control assays either using cell free extracts of YF11/pPWW50A (Figure 2M) or using boiled crude extracts of YF11/pCSG2220 (Figure 2N). This experiment demonstrated that CrmL was responsible for removing the terminal L-leucine of **2** to provide **3**.

Combining our experimental data and bioinformatic analysis, we proposed a biosynthetic pathway for **1** (Figure 1B). Picolinic acid (**5**) was formed by CrmC and CrmD from L-lysine via piperidine-2-carboxylic acid (**4**). CrmE was probably a dual functional enzyme capable of both activating **5** to **6**, and loading **6** to its N-terminal ACP, by analogy to its homologue NikE, similar dual function of which was recently demonstrated.[12b] The resulting product then entered into the PKS module in CrmA to form the putative intermediate **7**. Next, the A domain in CrmA might activate cysteine based on its predicted specificity. It is highly hypothetical that CrmI might convert CrmA T-domain-tethered cysteine to dehydroalanine (Dha) via an oxidative β-elimination. The conversion of cysteine into Dha was precedent in post-translational modification of proteins.[22] Subsequently, the enamine may serve as a nucleophile to attack **7** (Figure 1B), resulting in an intermediate **8**, which would sequentially undergo tautomerization (**8**→**9**), intramolecular cyclization (**9**→**10**), and a second tautomerization to afford **11** (Figure 1B). After condensation of **11** with L-leucine tethered to the CrmB T-domain, caerulomycin L (**2**) might be released from CrmB by the thioesterase CrmJ, and is converted to **3** by CrmL. The dehydrogenation by CrmN/CrmO would generate **12**, and the transamination of **12** by CrmG yielded **13**. CrmH probably catalyzed the *N*-hydroxylation of **13** to provide **14**, mechanistically analogous to P450 enzyme NocL in nocardicin A biosynthesis.[23] Finally, the **1** biosynthesis would be completed by the CrmM-catalyzed *O*-methylation of **14**.

In summary, we have cloned and characterized the biosynthetic gene cluster for **1** from *A. cyanogriseus* WH1-2216-6, the identity of which was confirmed by heterologous expression. Functional elucidation of the amidohydrolase CrmL indicated a process of extension/removal of a redundant amino acid in **1** biosynthesis, similar to the protection/deprotection strategy recently demonstrated in the biosynthesis of macrolactam vicenistatin and the antibiotic xenocoumacin in which similar peptidases tailored a hybrid PKS/NRPS assembly line.[24] The collismycin biosynthetic gene cluster was also recently elucidated to share a similar organization as the **1** gene cluster (Table S5, Supporting Information),[25] suggesting a common mechanism for 2,2′-bipyridyl ring formation.

Acknowledgment. Financial support was provided by MOST (2010CB833800), NSFC (31070045, 31125001), and CAS (KZCX2-YW-JC202, 08SL111002, KSCX2-EW-G-12). We thank Prof. Meifeng Tao (Shanghai Jiao Tong University) for providing *S. coelicolor* YF11. We are grateful to the analytical facilities in SCSIO.

Supporting Information Available. Experimental details and 1D and 2D NMR spectra of **2** and **3**. This material is available free of charge via the Internet at http://pubs.acs.org. The authors declare no competing financial interest.

(21) Zhou, H.; Wang, Y.; Yu, Y.; Bai, T.; Chen, L.; Liu, P.; Guo, H.; Zhu, C.; Tao, M.; Deng, Z. *Curr. Microbiol.* **2012**, *64*, 185–190.

(22) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernandez-Gonzalez, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. *Chem. Sci.* **2011**, *2*, 1666–1676.

(23) Kelly, W. L.; Townsend, C. A. *J. Am. Chem. Soc.* **2002**, *124*, 8186–8187.

(24) (a) Shinohara, Y.; Kudo, F.; Eguchi, T. *J. Am. Chem. Soc.* **2011**, *133*, 18134–18137. (b) Reimer, D.; Pos, K. M.; Thines, M.; Grun, P.; Bode, H. B. *Nat. Chem. Biol.* **2011**, *7*, 888–890.

(25) Garcia, I.; Vior, N. M.; Brana, A. F.; Sabin, J. G.; Rohr, J.; Moris, F. *Chem. Biol.* **2012**, *19*, 399–413.

The authors declare no competing financial interest.