Organic LETTERS

Letter

pubs.acs.org/OrgLett

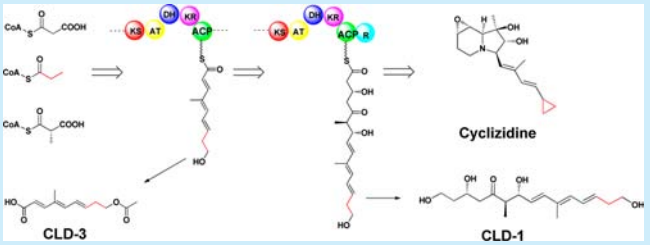
Identification of the Polyketide Biosynthetic Machinery for the Indolizidine Alkaloid Cyclizidine

Wei Huang, Seong Jong Kim, Joyce Liu, and Wenjun Zhang*

Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: The cyclizidine biosynthetic gene cluster was identified from *Streptomyces* NCIB 11649, which revealed the polyketide biosynthetic machinery for cyclizidine alkaloid biosynthesis. Both *in vivo* mutagenesis study and *in vitro* biochemical analysis provided insight into the timing and mechanism of the biosynthetic enzymes that produce cyclizidine-type indolizidine alkaloids.



Indolizidine alkaloids are natural products containing a fused five- and six-membered ring system with a tertiary amine and various substitutions (Figure 1). Over a hundred indolizidine

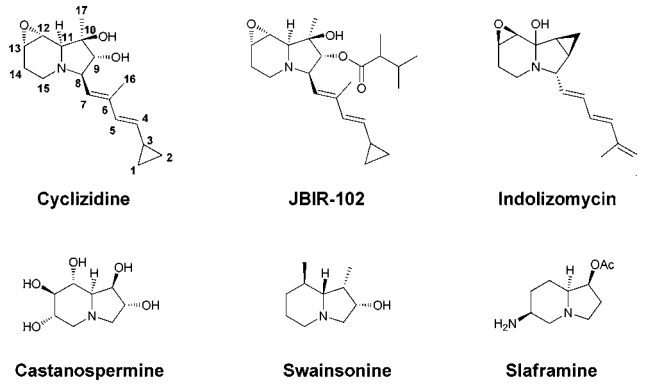


Figure 1. Molecular structures of indolizidine alkaloids.

alkaloids are known, with activities including phytotoxic, insecticidal, antibacterial, and fungicidal.[1] They have also shown promising activities for the study of the immune response, cancer metastasis, intracellular transport, and viral infectivity.[2] Indolizidine alkaloids have been isolated from a myriad of sources, including ants, frogs, plants, and fungi.[3] Surprisingly, only cyclizidine (CLD) and two congeners, JBIR-102 and indolizomycin, are known to be of bacterial origin.[4] In particular, CLD was produced by *Streptomyces* NCIB 11649, and structural analysis revealed that the indolizidine core of CLD is substituted with two hydroxyl moieties (C9 and C10), a C12/C13 epoxide, and a monosubstituted cyclopropyl *trans*-dienic subunit at C8 (Figure 1).[4c,5] CLD showed nonselective immunostimulatory properties and cytotoxic activities against human malignant pleural mesothelioma ACC-MESO-1 cells and cervical carcinoma HeLa cells with IC_{50} values of 32 and 16 μM, respectively.[4b] Furthermore, the acetate derivative of CLD demonstrated β-blocker activity and could thus be potentially useful for the management of cardiac arrhythmias.[4a,c]

The unique structures and potent biological activities of indolizidine alkaloids have rendered this family of compounds popular targets for synthesis.[5,6] Recently, the first total synthesis of CLD was reported; however, only the enantiomer was obtained in 26 steps with an overall yield of 2.7%.[5] The complexity in the molecular structure of CLD and the difficulty in its organic synthesis thus make the study of its biosynthetic pathway interesting. Although isotope-labeled precursor feeding studies have been performed to probe the possible biosynthetic origins of several indolizidines,[3c,7] no dedicated genes or enzymes have yet been identified for this family of alkaloids. Here, we report the identification of the biosynthetic gene cluster for CLD from *Streptomyces* through genome scanning, targeted gene deletion, and enzymatic studies.

For mining the CLD biosynthetic genes, the genome of *Streptomyces* NCIB 11649 was subjected to Illumina and PacBio sequencing, which resulted in ~8.1 M nonredundant bases after assembly of paired sequence reads. Because secondary metabolite biosynthetic genes are typically clustered in microbes, a local BLASTP analysis was performed using two enzyme probes for putative gene cluster identification. One probe was a polyketide synthase (PKS) on the basis of previous isotope labeling results,[7a] and the other probe was an aminotransferase that is likely involved in the incorporation of the nitrogen atom. Through our bioinformatics search, we located one putative gene cluster (GenBank accession no. KT327068), which spans 47 kb and consists of 15 open reading frames (ORFs), here designated as *cycA-O* (Figure 2, Table S1). Of these ORFs, six PKS and related proteins are encoded, which are organized into seven PKS modules containing catalytic domains including ketosynthase (KS), acyltransferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), enoylreductase (ER),

Received: September 17, 2015
Published: October 16, 2015

DOI: 10.1021/acs.orglett.5b02707
Org. Lett. 2015, 17, 5344−5347

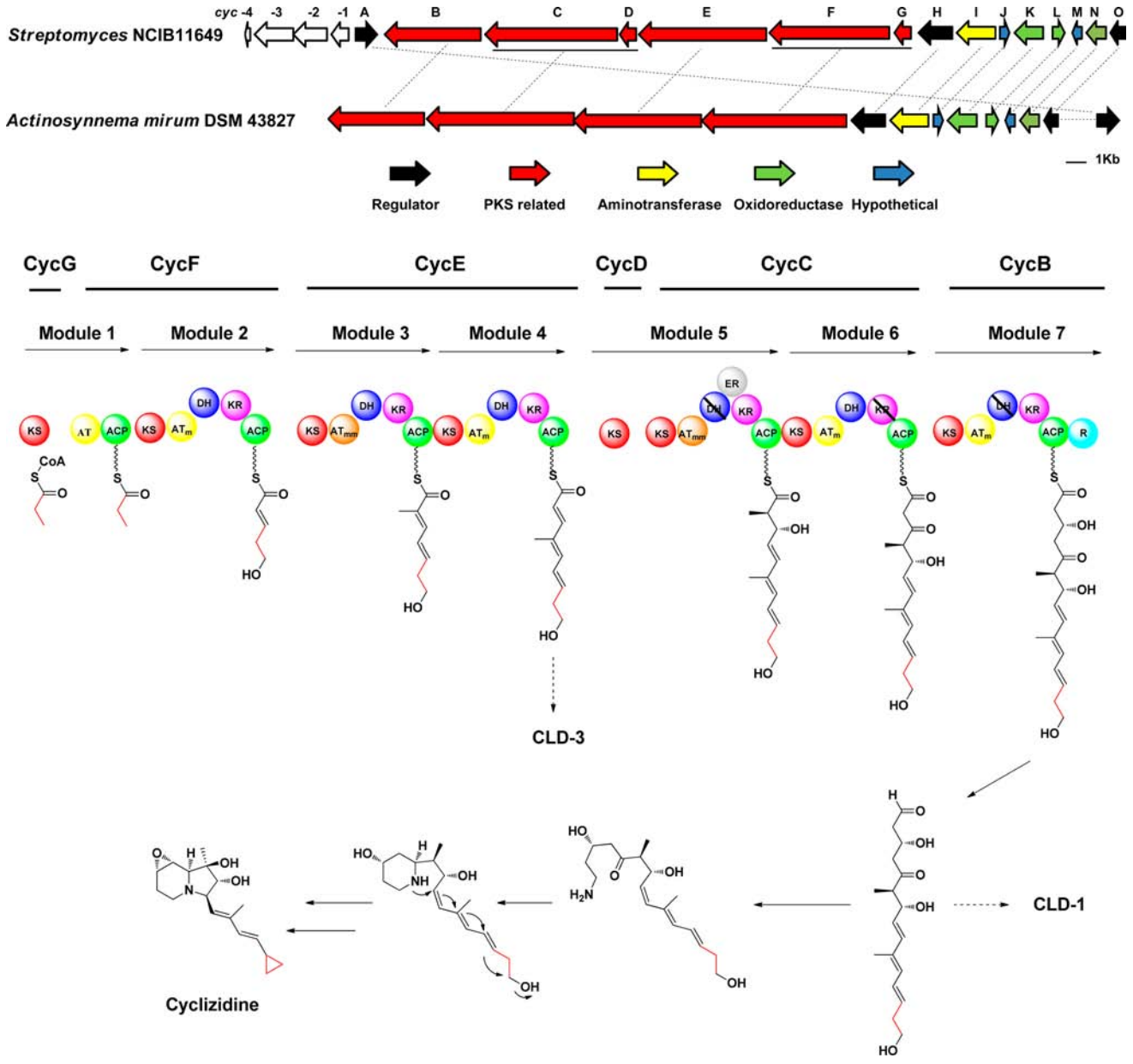


Figure 2. Map of the cyclizidine gene cluster and proposed biosynthetic pathway.

and terminal reductase (R). The total number of PKS modules is in agreement with the seven PKS building monomers of CLD (one starter unit and six extender units) predicted from precursor feeding experiments.[7a] As expected, an aminotransferase homologue (CycI) and several oxidoreductases including an acyl-CoA dehydrogenase homologue (CycK), a flavin reductase homologue (CycL), and a ribonucleotide reductase homologue (CycN) are also encoded in the gene cluster. Additional ORFs in the cluster encode three putative transcriptional regulators (CycA, H, and O) and two hypothetical proteins with unknown functions (CycJ and M). To confirm the involvement of this *cyc* cluster in CLD biosynthesis, we constructed a clean gene-knockout mutant Δ*H*Wcyc5−7 of *Streptomyces* NCIB 11649, in which the gene cluster was inactivated by in-frame deletion of a 7 kb DNA region from *cycE* to *cycF*, and the genotype of the resulting mutant was verified by PCR (Figure S1). The production of CLD was completely abolished in this mutant (Figure 3A), confirming that the identified *cyc* cluster is involved

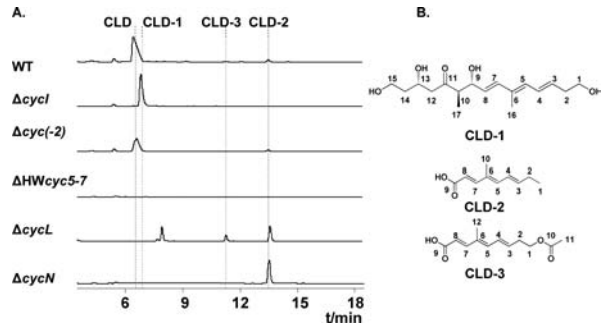


Figure 3. Genetic characterization of CLD gene cluster. (A) HPLC analysis (260 nm) of metabolites produced by wild-type and mutant *Streptomyces* NCIB 11649 strains.[9] (B) Structures of purified metabolites. The stereochemistry at C9, C10, and C13 of CLD-1 is mainly predicted from the bioinformatics analysis of CycKR5 (B1-type) and CycKR7 (B1-type) (Figure S22).[13b]

DOI: 10.1021/acs.orglett.5b02707
Org. Lett. 2015, 17, 5344−5347

in the biosynthesis of CLD. Interestingly, a homologous gene cluster was also found in the published genome of *Actinosynnema mirum* DSM 43827, although no CLD-type product has been reported from this strain. By comparing the sequences around the two clusters, the boundary of the *cyc* cluster was putatively identified (Figure 2).

To further explore the functions of the *cyc* ORFs, a set of clean gene-knockout experiments were carried out, followed by metabolic profiling. The knockout targets included CycI, the aminotransferase, CycN, the ribonucleotide reductase, CycL, the flavin reductase, and a putative amine oxidase encoded by *cyc(−2)*. All of the genes were deleted in-frame through double crossover leaving a scar of 6 bp (Figure S1). The deletion of *cyc(−2)* had no impact on the production of CLD (Figure 3A), which further supports the cluster boundary assignment. In contrast, the production of CLD was completely abolished in the Δ*cycI*, Δ*cycN*, and Δ*cycL* mutants (Figure 3A), demonstrating that these genes are essential for the biosynthesis of CLD. In addition, fermentation of the Δ*cycI* mutant accumulated a new major metabolite, CLD-1, that was not found in the wild-type or any other mutant cultures (Figure 3). CLD-1 was isolated at ∼2 mg/L after reversed-phase high performance liquid chromatography (HPLC) purification, and the molecular formulas of the resulting yellow powder were indicated to be $C_{17}H_{28}O_5$ based on high resolution mass spectrometry (HRMS) analysis, suggesting the production of a full-length polyketide lacking a nitrogen atom (Figure S18). Analysis of the ^{1}H and ^{13}C NMR spectra of purified CLD-1 showed the presence of a ketone, conjugated olefins, two methyl moieties, and four hydroxylated carbons, and further 2D NMR analysis confirmed their connectivity as shown in Figure 3B (Table S4 and Figures S2−S7). Accumulation of CLD-1 as a major product in Δ*cycI* suggested that the cyclopropyl ring formation, C10 hydroxylation, and C12/C13 epoxidation take place after the transamination. Although an intermediate with a C15 terminal aldehyde is the most likely substrate for CycI (Figure 2),[8] a C15 terminal alcohol was observed in CLD-1 which is proposed to be generated by the R domain of CycB or other reductases for possible detoxification.

No new metabolites were identified from the cultures of Δ*cycN* and Δ*cycL* by HPLC analysis; however, the production level of two minor metabolites, CLD-2 and CLD-3, was significantly altered compared to that of the wild-type (Figure 3A). In particular, the titers of CLD-2 increased ∼10-fold in the cultures of Δ*cycN* and Δ*cycL*; the titer of CLD-3 increased ∼10-fold in the culture of Δ*cycL*, but its production was completely abolished in the Δ*cycN* mutant. We then scaled up these mutant cultures and purified CLD-2 and CLD-3 with a yield of ∼1.5 and ∼1.2 mg/L, respectively. The subsequent HRMS and NMR analysis showed that both CLD-2 and CLD-3 are polyene acids resembling the conjugated olefin portion of CLD-1, with the exception that CLD-2 lacks the terminal hydroxyl group at C1 while CLD-3 is acetylated at the C1 hydroxyl (Figure 3B, Tables S5−S6, and Figures S8−S17). We propose that the acetylation observed in CLD-3 is likely catalyzed by an endogenous acetyltransferase that is not related to CLD biosynthesis. These results suggested that CycN could be involved in the installation of a hydroxyl group at C1 found in CLD-1 and CLD-3, but the function of CycL in the CLD biosynthesis remains unclear.

The structural elucidation of CLD-1, CLD-2, and CLD-3 allowed us to scrutinize the action of the PKS assembly line comprised of CycB, C, D, E, F, and G in providing the carbon backbone in CLD synthesis (Figure 2). Previous feeding studies demonstrated the incorporation of three propionate and four acetate units in generating the core scaffold of CLD,[7a] consistent with the predicted substrate preference of the AT domains for malonyl-CoA in modules 2, 4, 6, and 7, and for methylmalonyl-CoA in modules 3 and 5.[10] The loading module (module 1) probably takes propionyl-CoA, which is suggested by the production of CLD-2 as a shunt product in the wild-type culture. To confirm that the loading AT$_L$ domain recognizes and transfers the propionyl unit from CoA to ACP$_L$, we dissected the loading module of CycF and purified the AT$_L$ and ACP$_L$ domains as individual proteins from *E. coli* for biochemical analysis (Figure S21). The *in vitro* assays with AT$_L$, propionyl-CoA, and holo-ACP$_L$ demonstrated AT$_L$-promoted formation of propionyl-ACP$_L$ by LC-HRMS analysis (Figure 4), confirming the activity

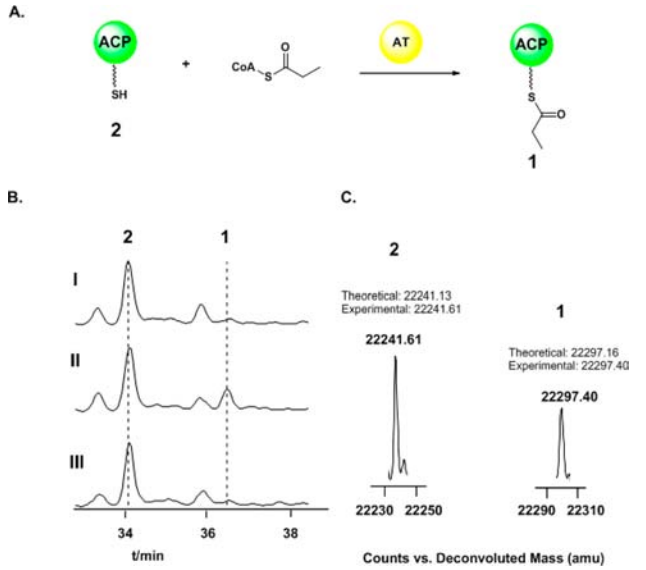


Figure 4. Biochemical characterization of CycF-AT$_L$ and CycF-ACP$_L$ (A) Schematic of the transacylation reaction. (B) LC/HRMS chromatograms showing AT$_L$-promoted formation of propionyl-ACP$_L$ (**1**) from holo-ACP$_L$ (**2**) (trace II). **1** was undetectable in controls without propionyl-CoA (trace I) or AT$_L$ (trace III). (C) Deconvoluted masses of **1** and **2** in HRMS analysis.

of AT$_L$ in the uptake of the propionate starter unit. Although the timing of the hydroxylation of the propionyl unit is unknown, we propose that the hydroxylation proceeds at a relatively early stage on the PKS assembly line based on the production of CLD-3. Previously, Leeper et al. proposed that sequential reactions catalyzed by an acyl-CoA dehydrogenase and an enoyl-CoA hydratase on propionyl-CoA could lead to the terminal hydroxylation,[11] and it is yet to be determined if CycK, an acyl-CoA dehydrogenase homologue encoded in the gene cluster, is involved in the proposed reaction on propionyl-CoA, or more likely, on propionyl-ACP$_L$. Alternatively, our gene knockout studies suggest that CycN could be related to the hydroxylation at C1. CycN is homologous to the ribonucleotide reductase (RNR) β subunit that contains a diferric iron center. RNRs are known to catalyze the formation of deoxyribonucleotides from ribonucleotides via a free radical mechanism,[12] and it is thus very likely that CycN is able to catalyze the terminal hydroxylation on a less reactive carbon center through a radical-based mechanism. Further biochemical analysis is needed to elucidate the exact roles of CycK and CycN in CLD biosynthesis.

The accumulation of CLD-1 in Δ*cycI* together with the previous feeding study using a doubly labeled $CH_3{}^{13}C^{18}O_2Na$

DOI: 10.1021/acs.orglett.5b02707
Org. Lett. 2015, 17, 5344−5347

precursor[11] indicate that the functions of several catalytic domains of the PKSs are skipped in synthesizing CLD-1 and CLD. These include the DH domain (along with the ER domain) in module 5 (CycDH5), the KR domain (along with the DH domain) in module 6 (CycKR6), and the DH domain in module 7 (CycDH7) (Figure 2). Further bioinformatics analysis showed that CycDH5 seems to be catalytically active with the conserved active site motif (HxxxGxxxxP), CycKR6 is likely inactive with the key catalytic residue tyrosine mutated to phenylalanine (the NADPH binding site being intact), and CycDH7 is probably inactive missing key active site residues (Figure S22).[13] It is notable that domain/module skipping is not uncommon during polyketide biosynthesis,[14] which makes the structure prediction of polyketides using the linear processive logic of modular biosynthesis often unreliable. Interestingly, the gene homologous to *cycC* in *A. mirum* encodes a PKS with domains organized into KS-AT-DH-KR-ACP-KS-AT-ACP, naturally omitting the ER domain in module 5 and the KR-DH didomain in module 6. In addition, the PKS machinery in *A. mirum* seems to have the two putative KS domains (encoded by *cycG* and *cycD* in *Streptomyces* NCIB 11649, respectively) fused to the *N*-terminus of the CycF and CycC homologues, respectively during the cluster evolution, and the functions of these extra KS domains remain unknown. The terminal R domain in the PKS assembly line finally catalyzes the reductive release of the nascent, full-length polyketide chain to form a terminal aldehyde, which presumably serves as the substrate for CycI. We propose that the six-membered ring of CLD is formed through further transamination, which is followed by the likely concerted five- and three-membered ring cyclizations that lead to the generation of both the indolizidine scaffold and the terminal cyclopropyl ring found in CLD (Figure 2).[11]

In summary, we have identified the biosynthetic gene cluster of cyclizidine by genome mining and targeted gene deletion, which sets the stage for scrutinizing the complex PKS machinery in cyclizidine scaffold assembly. The isolation and structural elucidation of several new metabolites shed light on the functions of some biosynthetic enzymes, and we further confirmed the selection and loading of a propionyl starter unit for polyketide assembly and cyclopronyl ring formation via the intermediacy of C1-hydroxylation. Similar chemical logic and enzymatic machinery are likely involved in JBIR-102 and indolizomycin biosynthesis, although indolizidine alkaloids lacking the extended olefinic appendage are probably generated differently.[3c,7b] Additional genetic and biochemical studies will provide more insight into the timing and mechanism of the biosynthetic enzymes in producing cyclizidine-type indolizidine alkaloids and promote the engineered biosynthesis of new indolizidine alkaloids for drug discovery.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.5b02707.

Experimental details; tables of plasmids; mutant stains; HRMS, UV, and NMR spectral data for CLD1−3 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: wjzhang@berkeley.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank J. Pelton (UC Berkeley) for helping with NMR spectroscopic analysis, J. Skerker (UC Berkeley) for helping with genome assembly, and X. Zhu (UC Berkeley) for helping with LC-MS analysis. This work was supported by the Pew Scholars Program and the National Institutes of Health (DP2AT009148).

■ REFERENCES

(1) (a) Gellert, E. *J. Nat. Prod.* **1982**, *45*, 50. (b) Michael, J. P. *Nat. Prod. Rep.* **2008**, *25*, 139. (c) Broquist, H. P. *Annu. Rev. Nutr.* **1985**, *5*, 391. (d) Lourenço, A. M.; Mports cts 2AT009148holars Programra, M. M. A., Studies in Natural Products Chemistry; Rahman, A. u., Ed.; Elsevier: 2002; *Vol.* 27, Part H, p 233.

(2) (a) Ahmed, S. P.; Nash, R. J.; Bridges, C. G.; Taylor, D. L.; Kang, M. S.; Porter, E. A.; Tyms, A. S. *Biochem. Biophys. Res. Commun.* **1995**, *208*, 267. (b) Costa, J.; Pinto Ricardo, C. P. *Plant Sci.* **1994**, *101*, 137. (c) Goss, P. E.; Baker, M. A.; Carver, J. P.; Dennis, J. W. *Clin. Cancer Res.* **1995**, *1*, 935. (d) Jacob, G. S. *Curr. Opin. Struct. Biol.* **1995**, *5*, 605. (e) Kang, M. S. *Glycobiology* **1996**, *6*, 209. (f) Nojima, H.; Kimura, I.; Chen, F. J.; Sugihara, Y.; Haruno, M.; Kato, A.; Asano, N. *J. Nat. Prod.* **1998**, *61*, 397. (g) Pili, R.; Chang, J.; Partis, R. A.; Mueller, R. A.; Chrest, F. J.; Passaniti, A. *Cancer Res.* **1995**, *55*, 2920. (h) Vlietinck, A. J.; De Bruyne, T.; Apers, S.; Pieters, L. A. *Planta Med.* **1998**, *64*, 97.

(3) (a) Clevenstine, E. C.; Walter, P.; Harris, T. M.; Broquist, H. P. *Biochemistry* **1979**, *18*, 3663. (b) Hohenschutz, L. D.; Bell, E. A.; Jewess, P. J.; Leworthy, D. P.; Pryce, R. J.; Arnold, E.; Clardy, J. *Phytochemistry* **1981**, *20*, 811. (c) Schneider, M. J.; Ungemach, F. S.; Broquist, H. P.; Harris, T. M. *J. Am. Chem. Soc.* **1982**, *104*, 6863.

(4) (a) Gomi, S.; Ikeda, D.; Nakamura, H.; Naganawa, H.; Yamashita, F.; Hotta, K.; Kondo, S.; Okami, Y.; Umezawa, H.; Iitaka, Y. *J. Antibiot.* **1984**, *37*, 1491. (b) Izumikawa, M.; Hosoya, T.; Takagi, M.; Shin-ya, K. *J. Antibiot.* **2012**, *65*, 41. (c) Freer, A. A.; Gardner, D.; Greatbanks, D.; Poyser, J. P.; Sim, G. A. *J. Chem. Soc., Chem. Commun.* **1982**, 1160.

(5) Hanessian, S.; Soma, U.; Dorich, S.; Deschenes-Simard, B. *Org. Lett.* **2011**, *13*, 1048.

(6) (a) Abels, F.; Lindemann, C.; Koch, E.; Schneider, C. *Org. Lett.* **2012**, *14*, 5972. (b) Bowen, E. G.; Wardrop, D. J. *Org. Lett.* **2010**, *12*, 5330. (c) Dhand, V.; Draper, J. A.; Moore, J.; Britton, R. *Org. Lett.* **2013**, *15*, 1914.

(7) (a) Leeper, F. J.; Padmanabhan, P.; Kirby, G. W.; Sheldrake, G. N. *J. Chem. Soc., Chem. Commun.* **1987**, 505. (b) Clevenstine, E. C.; Broquist, H. P.; Harris, T. M. *Biochemistry* **1979**, *18*, 3658.

(8) Pugh, C. E.; Harwood, J. L.; John, R. A. *J. Biol. Chem.* **1992**, *267*, 1584.

(9) The peak (~8 min) in the trace of Δ*cycL* was a mixture of two metabolites which we were unable to separate. The NMR data on the mixture suggested that one of the metabolites was likely an unacetylated CLD-3, and the other was likely an unknown truncated polyketide shunt product.

(10) (a) Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009**, *458*, 181. (b) Khayatt, B. I.; Overmars, L.; Siezen, R. J.; Francke, C. *PLoS One* **2013**, *8*, e62136.

(11) Leeper, F. J.; Shaw, S. E.; Satish, P. *Can. J. Chem.* **1994**, *72*, 131.

(12) Nordlund, P.; Reichard, P. *Annu. Rev. Biochem.* **2006**, *75*, 681.

(13) (a) Aparicio, J. F.; Molnar, I.; Schwecke, T.; Konig, A.; Haydock, S. F.; Khaw, L. E.; Staunton, J.; Leadlay, P. F. *Gene* **1996**, *169*, 9. (b) Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T. *Structure (Oxford, U. K.)* **2010**, *18*, 913.

(14) (a) Beck, B. J.; Yoon, Y. J.; Reynolds, K. A.; Sherman, D. H. *Chem. Biol.* **2002**, *9*, 575. (b) Rowe, C. J.; Bohm, I. U.; Thomas, I. P.; Wilkinson, B.; Rudd, B. A.; Foster, G.; Blackaby, A. P.; Sidebottom, P. J.; Roddis, Y.; Buss, A. D.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2001**, *8*, 475. (c) Tang, G. L.; Cheng, Y. Q.; Shen, B. *J. Nat. Prod.* **2006**, *69*, 387.

DOI: 10.1021/acs.orglett.5b02707
Org. Lett. 2015, 17, 5344−5347