Cylindrocyclophane Biosynthesis Involves Functionalization of an Unactivated Carbon Center

Hitomi Nakamura, Hilary A. Hamer, Gopal Sirasani, and Emily P. Balskus*

Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

S *Supporting Information*

ABSTRACT: The cylindrocyclophanes are a family of natural products that share a remarkable paracyclophane carbon scaffold. Using genome sequencing and bioinformatic analyses, we have discovered a biosynthetic gene cluster involved in the assembly of cylindrocyclophane F. Through a combination of *in vitro* enzyme characterization and feeding studies, we confirm the connection between this gene cluster and cylindrocyclophane production, elucidate the chemical events involved in initiating and terminating an unusual type I polyketide synthase assembly line, and discover that macrocycle assembly involves functionalization of an unactivated carbon center.

Figure 1. The cylindrocyclophane family of natural products and previous biosynthetic studies. (A) Structures of cylindrocyclophanes isolated from *C. licheniforme* strains ATCC 29204 and 29412. (B) Feeding studies performed by Bobzin and Moore revealed the polyketide origin of the cylindrocyclophanes. (C) Biosynthetic hypothesis put forth by Bobzin and Moore.[4]

Living organisms such as bacteria and plants produce a wide variety of secondary metabolites, many of which have complex chemical structures. Natural products with unusual molecular architecture are of interest to both chemical biologists and synthetic chemists. Studying the biosynthesis of remarkable chemical functionality often reveals unprecedented enzymatic reactivity that may be further applied in combinatorial biosynthesis and metabolic engineering.[1] Unique natural product structures also inspire the development of new methodology and strategies for chemical synthesis.[2] The cylindrocyclophanes are a family of cytotoxic natural products from photosynthetic cyanobacteria that contain a highly unusual [7.7]paracyclophane ring system.[3] This scaffold is unique among natural products, and the biochemical transformations involved in its assembly are currently unknown. Here we report the identification of a gene cluster from the cyanobacterium *Cylindrospermum licheniforme* that is involved in cylindrocyclophane production. Using a combination of *in vitro* enzyme characterization and feeding studies, we have discovered that assembly of the paracyclophane macrocycle employs unusual biosynthetic logic, including functionalization of an unactivated carbon center.

Moore et al. reported the isolation and structural characterization of cylindrocyclophanes A–F in the early 1990s (Figure 1A).[3a,b] Subsequent feeding experiments suggested a unique biosynthetic pathway for constructing the [7.7]paracyclophane skeleton (Figure 1B). The symmetric labeling pattern resulting from acetate incorporation revealed the polyketide origin of these natural products and led Bobzin and Moore to hypothesize that their biosynthesis involved a C–C bond-forming, head-to-tail dimerization reaction of an alkyl resorcinol intermediate (Figure 1C).[4] This type of tailoring event, which constructs two sp^2–sp^3 C–C linkages and two stereocenters, has limited precedent among biological transformations.[5] Though feeding
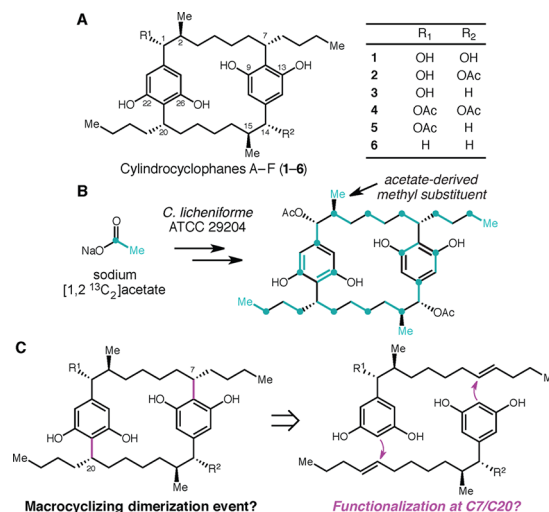
experiments did not provide insight into the exact structure of the monomeric precursor, Bobzin and Moore proposed that the C7/C20 position was pre-functionalized for eventual C–C bond formation during monomer construction on a polyketide synthase (PKS) assembly line.

The unusual molecular framework of the cylindrocyclophanes has inspired several elegant total syntheses.[6] Though synthetic chemists have recognized and admired the complexity-generating power of the proposed biosynthetic route to the [7.7]paracyclophane skeleton, no chemical synthesis to date has been able to exploit this disconnection. Understanding cylindrocyclophane biosynthesis at the molecular level should therefore not only reveal new enzymatic reactions and biosynthetic logic but also inspire chemists seeking to emulate the elegance of Nature's synthetic approach.

Our strategy for locating the cylindrocyclophane biosynthetic gene cluster involved whole genome sequencing followed by bioinformatic searches for specific gene products likely to be involved in natural product assembly. In addition to confirming the polyketide nature of the cylindrocyclophanes, the original
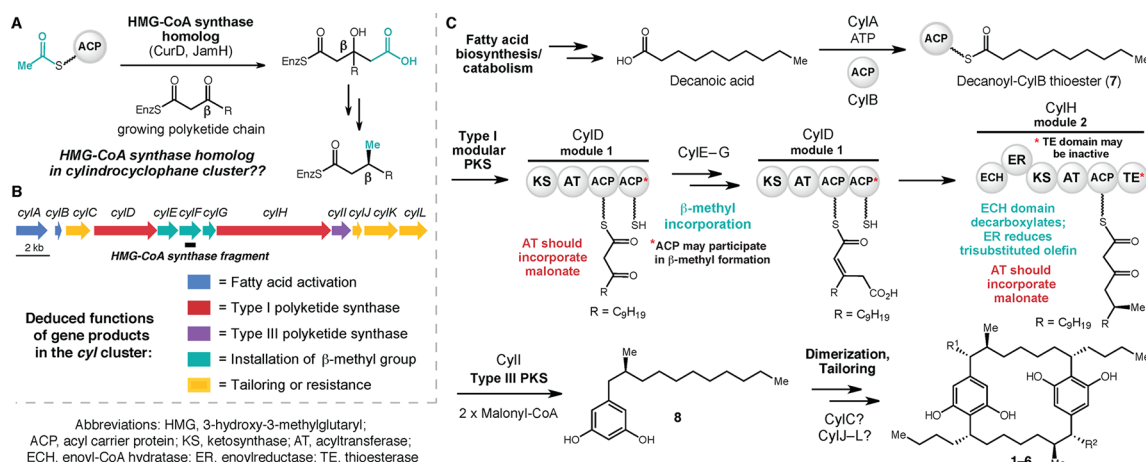
Figure 2. Discovery of the *cyl* gene cluster and proposed biosynthesis of the cylindrocyclophanes. (A) HMG-CoA synthase homologs involved in β-Me group installation on polyketide scaffolds. (B) The putative *cyl* gene cluster. Each arrow represents the direction of transcription of an open reading frame. See SI for complete cluster annotation. (C) Biosynthetic hypothesis for the assembly of the cylindrocyclophanes.

feeding studies revealed an acetate-derived β-methyl group within the backbone of cylindrocyclophanes A—F (Figure 1B).[4] This relatively unusual structural feature is introduced into polyketide scaffolds using a highly conserved pathway containing an enzyme homologous to hydroxymethylglutaryl-CoA (HMG-CoA) synthase (Figure 2A).[7] Reasoning that a closely related enzyme should be involved in cylindrocyclophane assembly, we used degenerate PCR to amplify putative HMG-CoA synthase homologs from genomic DNA of *C. licheniforme* ATCC 29412, a producer of cylindrocyclophanes A—F.[3b,8] A 636 bp PCR product was obtained, and its translated sequence revealed close homology (92 and 90% identity at the amino acid level) to JamH and CurD, cyanobacterial HMG-CoA synthases involved in β-Me installation during jamaicamide[9] and curacin[10] biosynthesis (Figure S1). Because genes responsible for natural product assembly in microbes are typically co-localized, we hypothesized that the *C. licheniforme* HMG-CoA synthase fragment should be clustered with additional genes involved in cylindrocyclophane biosynthesis.

Searches for this nucleotide sequence in *C. licheniforme* genome sequence data generated by 454 pyrosequencing revealed an identical sequence on contig 1024. In addition to a full-length HMG-CoA synthase homolog, this 12.1 kb genome fragment contained additional open reading frames (ORFs) hypothesized to encode enzymes possessing functions consistent with involvement in cylindrocyclophane assembly, including a PKS and other enzymes involved in β-Me installation. Searches for the remaining genes associated with addition of this substituent identified two more contigs connected to cylindrocyclophane biosynthesis (111 and 112).[11] We confirmed the proximities of all three contigs by sequencing regions of the genome spanning adjacent contigs, permitting assembly of a single gene cluster. The putative cylindrocyclophane (*cyl*) biosynthetic gene cluster contains 12 ORFs (Figure 2B), including genes encoding the enzymatic machinery required for installing the β-Me substituent (*cylE, F, G, H*). The cluster also has two distinct types of PKS machinery, a short type I modular PKS assembly line (*cylD, H*) and a type III PKS (*cylI*). Two additional genes, predicted fatty acid adenylating enzyme *cylA* and freestanding acyl carrier protein (ACP) *cylB*, likely produce enzymes involved in initiating biosynthesis. The remaining ORFs include homologs of a phospholipid methyltransferase (*cylJ*), a predicted flavin-dependent oxidoreductase

(*cylL*), a protein with a hemolysin-like binding region (*cylK*), and a hypothetical protein (*cylC*). Detailed annotation of the cluster (Table S1) revealed all of the enzymatic reactivity necessary for biosynthesis of a putative cylindrocyclophane monomeric precursor (**8**) and enabled us to generate a detailed hypothesis for its assembly (Figure 2C).

One of the most striking and unexpected features of the *cyl* gene cluster is the abbreviated nature of the type I PKS assembly line. CylD and CylH contain just two intact modules and are therefore expected to perform only two elongation reactions. Iterative function of these enzymes is unlikely as they both lack all of the domains needed to form the fully reduced polyketide backbone found in the final natural products. Given the size and structures of the cylindrocyclophanes, we propose that biosynthesis originates with recruitment of a free fatty acid. Activation of decanoic acid by CylA and loading onto CylB to form thioester **7** could be followed by transfer to the first type I PKS CylD. This PKS contains ketosynthase (KS) and acyltransferase (AT) domains along with two tandem ACP domains, an arrangement that is often associated with the β-Me group modification.[7] We hypothesize that CylD performs one elongation with malonyl-CoA and the resulting β-ketothioester is further processed by CylE, F, and G, which are predicted to catalyze addition of acetate to the β-ketone and dehydration to form an α,β-unsaturated thioester intermediate. We were unable to locate a candidate freestanding ACP for use in β-Me installation but noted that the second ACP domain on CylD is most closely related in sequence to freestanding ACPs of this type, perhaps indicating that it fills this role. The final two enzymatic activities needed to complete construction of the chiral β-Me substituent are encoded as domains on the second type I PKS CylH. CylH contains enoyl-CoA hydratase (ECH) and enoylreductase (ER) domains, which should catalyze decarboxylation and reduction of the CylD product.[7] The additional KS, AT, ACP, and thioesterase (TE) domains of CylH constitute another PKS module and should catalyze a second elongation with malonyl-CoA. Interestingly, multiple sequence alignments of the CylH TE domain with TE domains from characterized type I PKS assembly lines indicate that it may lack a conserved catalytic aspartate residue and could therefore be non-functional.[12] We propose that type III PKS CylI terminates the assembly line by utilizing the CylH-bound β-ketothioester as a starter unit for aromatic ring formation. Type III PKSs typically use CoA

18519

dx.doi.org/10.1021/ja308318p | *J. Am. Chem. Soc.* 2012, 134, 18518−18521

thioesters as substrates and catalyze successive Claisen condensations, cyclization of the resulting polyketide intermediate, and aromatization. The action of CylI on the nascent product of the CylD-CylH assembly line could generate the resorcinol-type aromatic ring found in the cylindrocyclophanes. Overall, the PKS machinery encoded by this biosynthetic gene cluster is consistent with its role in cylindrocyclophane assembly.

We solidified the link between the *cyl* gene cluster and cylindrocyclophane biosynthesis by characterizing the activities of several of the encoded enzymes *in vitro*. Specifically, we examined CylA and CylB, the enzymes predicted to be involved in fatty acid recruitment, and CylI, the type III PKS. CylA, CylB, and CylI were individually cloned from *C. licheniforme*, overexpressed in *E. coli*, and purified. The ability of CylA to activate and load decanoic acid onto holo-CylB was probed using both HPLC and LC-MS (Figure 3A). Incubation of holo-CylB
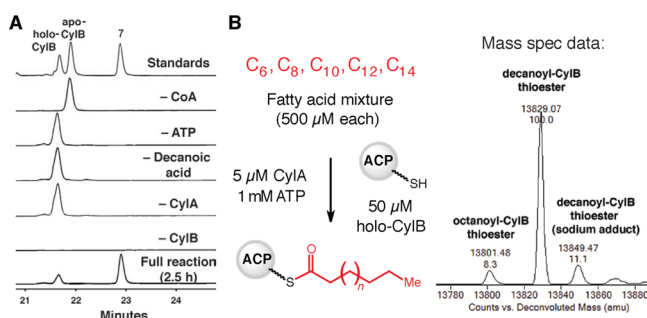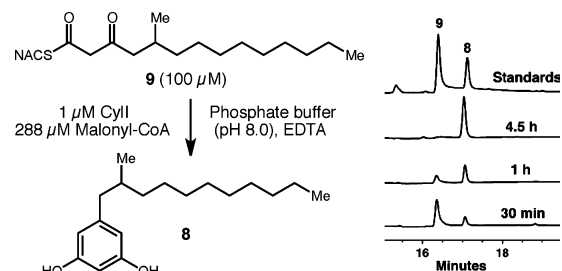

Figure 3. *In vitro* characterization of CylA and CylB. (A) HPLC assay showing the loading of decanoic acid onto holo-CylB by CylA. (B) LC-MS competition assay reveals selectivity for C_{10} fatty acid activation by CylA and CylB.

with decanoic acid, CylA, and ATP resulted in formation of the anticipated decanoyl-CylB thioester. This activity was dependent on the presence of both ATP and CylA. CylA was not able to load decanoyl-CoA onto holo-CylB, which suggests that CylA is a fatty acyl-AMP ligase rather than a fatty acyl-CoA ligase (Figure S8). The size selectivity of this fatty acid activation and loading process was evaluated using an LC-MS-based competition assay. Incubating CylA and holo-CylB with equal amounts of C_6, C_8, C_{10}, C_{12}, and C_{14} saturated fatty acids resulted in highly selective formation of decanoyl-CylB thioester (Figure 3B). We observed only a small amount (<10%) of product derived from octanoic acid. Overall, these experiments demonstrate that CylA and CylB are capable of recruiting decanoic acid for cylindrocyclophane biosynthesis in a size-selective manner. This exquisite selectivity is consistent with the observation that all members of the cylindrocyclophane natural product family share the same core carbon skeleton. Activation of a free fatty acid for use as a starter unit is rare among type I modular PKS assembly lines but has been demonstrated previously in the biosynthesis of mycobacterial fatty acids[13] and several hybrid PKS/non-ribosomal peptide synthetase systems.[9,14]

Having confirmed that the *in vitro* activities of enzymes involved in initiating cylindrocyclophane biosynthesis supported our biosynthetic hypothesis, we next explored the chemistry underlying assembly line termination. Evaluating the reactivity of type III PKS CylI required accessing a suitable substrate via chemical synthesis. We hypothesized that CylI would catalyze formation of a resorcinol-type aromatic ring via reaction of an ACP-tethered β-ketothioester intermediate with two molecules of malonyl-CoA (Scheme S1). We synthesized an *N*-acetylcyst-

amine (NAC) thioester analog (**9**) of the proposed CylI substrate and evaluated its reactivity in the presence of CylI and malonyl-CoA. A single new product peak was observed by HPLC (Figure 4); isolation of this product from a large-scale enzymatic

Figure 4. *In vitro* characterization of CylI. HPLC time course for formation of resorcinol **8** (230 nm).

reaction and characterization using NMR (^{1}H, ^{13}C) and HRMS confirmed formation of resorcinol **8**. The enzymatically synthesized material was identical in all respects to a product standard prepared via chemical synthesis. The reactivity of type III PKS CylI supports its proposed role in cylindrocyclophane biosynthesis, terminating a type I modular PKS assembly line in a highly unusual fashion via aromatic ring formation. While a few type III PKS enzymes and domains have been reported to accept acyl-ACP substrates generated by iterative type I[15] and type II[16] fatty acid synthases, to our knowledge CylI is the first type III PKS predicted to utilize an acyl-ACP substrate from a modular assembly line.

The *in vitro* characterization of CylA, CylB, and CylI connects the *cyl* gene cluster to assembly of an alkyl resorcinol intermediate that could be a monomeric precursor to the cylindrocyclophanes. To further confirm that the fatty acid activation pathway encoded by this cluster is directly connected to cylindrocyclophane production, we performed feeding studies with *C. licheniforme*, evaluating incorporation of isotopically labeled precursors into the cylindrocyclophane skeleton using LC-MS of culture extracts. Feeding of d_{19}-decanoic acid produced both d_{18}- and d_{36}-cylindrocyclophane F (Figure 5).
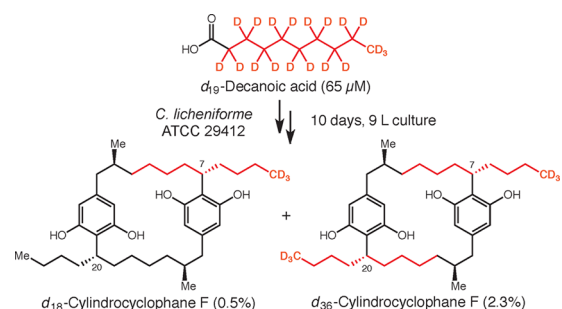
Figure 5. Incorporation of labeled decanoic acid into cylindrocyclophane F by *C. licheniforme*.

The masses corresponding to these single and double incorporation products were not found in extracts from cultures fed with unlabeled decanoic acid. Interestingly, the double incorporation product predominates, which may indicate that synthesis of a monomeric precursor and dimerization are tightly coupled, a hypothesis that is consistent with our inability to locate masses of putative monomeric intermediates in culture extracts. Overall, the results of this feeding experiment verify that

18520

dx.doi.org/10.1021/ja308318p | *J. Am. Chem. Soc.* 2012, 134, 18518–18521

decanoic acid is a biosynthetic precursor to the cylindrocyclophanes and is assimilated into both halves of the natural product skeleton. It also confirms that the eventual C7/C20 carbon atom of the paracyclophane scaffold enters the biosynthetic pathway as an unactivated methylene group. Construction of this unique macrocyclic architecture therefore requires a C–H functionalization event at some point in the biosynthetic pathway.

One intriguing possibility for macrocycle formation is direct oxidative sp^2–sp^3 C–C bond-forming dimerization of **8** involving the resorcinol aromatic ring and the unactivated alkyl chain. A conceptually related transformation is utilized in the biosynthesis of the prodiginine alkaloid streptorubin B and has been linked to the activity of Reiske oxygenase RedG.[5] Notably, no RedG homolog is found in the *cyl* cluster or in the *C. licheniforme* genome sequencing data. If cylindrocyclophane biosynthesis does employ direct oxidative C–C bond formation, it must therefore be carried out using distinct enzymatic chemistry. We hypothesize that one or more of the remaining enzymes encoded by the *cyl* cluster (CylC, CylJ–L) may be involved in paracyclophane formation, and the stage is now set to elucidate the details of this process through further biochemical characterization.[17]

In summary, we have obtained the first molecular insights into the biosynthesis of the cylindrocyclophanes through identification of a biosynthetic gene cluster, *in vitro* characterization of biosynthetic enzymes, and feeding studies. Assembly of potential biosynthetic intermediate **8** involves a remarkable combination of polyketide biosynthetic machinery: size-selective recruitment of a fatty acid building block by CylA and CylB, elaboration of the fatty acid by a type I modular PKS, and termination of the assembly line by type III PKS CylI. This unusual initiation and off-loading logic could potentially be applied in combinatorial biosynthesis. Incorporation of labeled decanoic acid into the cylindrocyclophane skeleton *in vivo* confirmed the role of this metabolite in biosynthesis and established that assembly of the macrocycle involves functionalization of an unactivated carbon center. Elucidating the biosynthetic logic underlying paracyclophane construction and tailoring, including the precise nature of the key C–C bond-forming event, are important future challenges that should enrich our understanding of how Nature constructs complex molecular architecture.

ASSOCIATED CONTENT

Supporting Information

Experimental details and characterization data. This material is available free of charge via the Internet at http://pubs.acs.org. Nucleotide sequence data have been deposited into GenBank (accession no.: JX477167).

AUTHOR INFORMATION

Corresponding Author
balskus@chemistry.harvard.edu

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) (a) Walsh, C. T.; Fischbach, M. F. *J. Am. Chem. Soc.* **2010**, *132*, 2469. (b) Weeks, A. M.; Chang, M. C. Y. *Biochemistry* **2011**, *50*, 5404.

(2) Wender, P. A.; Miller, B. L. *Nature* **2009**, *460*, 197.

(3) (a) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E.; Brinen, L. S.; Kato, Y.; Clardy, J. *J. Am. Chem. Soc.* **1990**, *112*, 4061. (b) Moore, B. S.; Chen, J. L.; Patterson, G. M. L.; Moore, R. E. *Tetrahedron* **1992**, *48*, 3001. (c) Bui, H. T. N.; Jansen, R.; Pham, H. T. L.; Mundt, S. *J. Nat. Prod.* **2007**, *70*, 499. (d) Chlipala, G. E.; Sturdy, M.; Krunic, A.; Lantvit, D. D.; Shen, Q.; Porter, K.; Swanson, S. M; Orjala, J. *J. Nat. Prod.* **2010**, *73*, 1529.

(4) Bobzin, S. C.; Moore, R. E. *Tetrahedron* **1993**, *49*, 7615.

(5) Sydor, P. K.; Barry, S. M.; Odulate, O. M.; Barona-Gomez, F.; Haynes, S. W.; Corre, C.; Song, L.; Challis, G. L. *Nat. Chem.* **2011**, *3*, 388.

(6) (a) Smith, A. B., III; Kozmin, S. A.; Paone, D. V. *J. Am. Chem. Soc.* **1999**, *121*, 7423. (b) Smith, A. B., III; Kozmin, S. A.; Adams, C. M.; Paone, D. V. *J. Am. Chem. Soc.* **2000**, *122*, 4984. (c) Hoye, T. R.; Humpal, P. E.; Moon, B. *J. Am. Chem. Soc.* **2000**, *122*, 4982. (d) Yamakoshi, H.; et al. *Org. Biomol. Chem.* **2009**, *7*, 3772. (e) Nicolaou, K. C.; Sun, Y.; Korman, H.; Sarlah, D. *Angew. Chem., Int. Ed.* **2010**, *49*, 5875.

(7) Calderone, C. T. *Nat. Prod. Rep.* **2008**, *25*, 845.

(8) We independently confirmed that this strain produces cylindrocyclophane F. For isolation details see SI.

(9) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M.; Gerwick, W. H. *Chem. Biol.* **2004**, *11*, 817.

(10) Chang, Z. X.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J. Y.; Sherman, D. H.; Gerwick, W. H. *J. Nat. Prod.* **2004**, *67*, 1356.

(11) Descriptions of the bioinformatic analyses used to locate the gene cluster and the strategy for cluster assembly can be found in the SI.

(12) The sequence of the CylH TE domain is most closely related to type II editing thioesterases, which are typically freestanding enzymes that catalyze hydrolysis at reduced rates. This homology suggests another possibility for assembly line termination; the TE could be functional but be outcompeted for off-loading by CylI.

(13) (a) Trivedi, O. A.; Arora, P.; Sridharan, V.; Tickoo, R.; Mohanty, D.; Rajesh, S.; Gokhale, R. S. *Nature* **2004**, *428*, 441. (b) Arora, P.; Goyal, A.; Natarajan, V. T.; Rajakumara, E.; Verma, P.; Gupta, R.; Yousuf, M.; Trivedi, O. A.; Mohanty, D.; Tyagi, A.; Sankaranarayanan, R.; Gokhale, R. S. *Nat. Chem. Biol.* **2009**, *5*, 166.

(14) Hansen, D. B.; Bumpus, S. B.; Aron, Z. D.; Kelleher, N. L.; Walsh, C. T. *J. Am. Chem. Soc.* **2007**, *129*, 6366.

(15) (a) Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P. *Nat. Chem. Biol.* **2006**, *2*, 494. (b) Miyanaga, A.; Funa, N.; Awakawa, T.; Horinouchi, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 871.

(16) (a) Song, L.; Barona-Gomez, F.; Corre, C.; Xiang, L.; Udwary, D. W.; Austin, M. B.; Noel, J. P.; Moore, B. S.; Challis, G. L. *J. Am. Chem. Soc.* **2006**, *128*, 14754. (b) Chemler, J. A.; Buchholz, T. J.; Geders, T. W.; Akey, D. L.; Rath, C. M.; Chlipala, G. E.; Smith, J. L.; Sherman, D. H. *J. Am. Chem. Soc.* **2012**, *134*, 7359. (c) Nakano, C.; Ozawa, H.; Akanuma, G.; Funa, N.; Horinouchi, S. *J. Bacteriol.* **2009**, *191*, 4916.

(17) Feeding studies to assess whether resorcinol **8** is an intermediate in cylindrocyclophane biosynthesis were attempted, but were complicated by the toxicity of **8** to *C. licheniforme* at low concentrations (<10 μM); lack of incorporation could be attributed either to **8** being off-pathway or to cell death. This observed toxicity does provide a possible explanation for tight coupling of monomer generation and macrocyclization.