Organic LETTERS

Letter

pubs.acs.org/OrgLett

Isolation of a Metabolite from the *pks* Island Provides Insights into Colibactin Biosynthesis and Activity

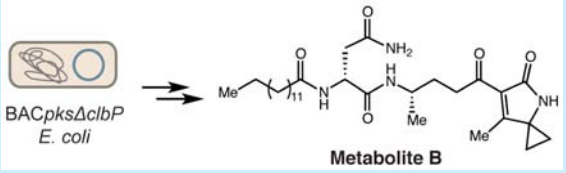
Carolyn A. Brotherton,[†] Matthew Wilson,[†] Gary Byrd,[‡] and Emily P. Balskus*[,†]

[†]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

[‡]Small Molecule Mass Spectrometry Facility, Faculty of Arts and Sciences Division of Science, Harvard University, Cambridge, Massachusetts 02138, United States

Supporting Information

ABSTRACT: Colibactin is a structurally uncharacterized, genotoxic natural product produced by commensal and pathogenic strains of *E. coli* that harbor the *pks* island. A new metabolite has been isolated from a *pks*$^+$ *E. coli* mutant missing an essential biosynthetic enzyme. The unusual azaspiro[2.4] bicyclic ring system of this molecule provides new insights into colibactin biosynthesis and suggests a mechanism through which colibactin and other *pks*-derived metabolites may exert genotoxicity.



Colibactin is a human-associated bacterial natural product that has been linked to disease. Exposure to certain strains of *Escherichia coli*, including gut commensals and extraintestinal pathogens, induces double-strand breaks in the DNA of host cells.[1] This genotoxic activity has been linked to a biosynthetic gene cluster (the *pks* island) that encodes a hybrid nonribosomal peptide synthetase (NRPS)−polyketide synthase (PKS) assembly line (Figure 1A).[1a] Recent studies have shown that *pks*$^+$ *E. coli* promote tumorigenesis in a mouse model of colitis-associated colorectal cancer (CRC) and that both colitis and CRC patients have increased carriage of *pks*$^+$ strains.[2] These findings support the theory that colibactin production affects cancer progression in humans. The *pks* island is also found in the genomes of sponge- and honeybee-associated bacteria, suggesting it may influence other symbioses.[3] Despite intense interest in its activity, the active genotoxin encoded by the *pks* island (colibactin) has not been isolated, severely limiting efforts to study its biosynthesis, mode of action, and role in host−microbe interactions.

We have gained structural information about colibactin by studying the biosynthetic enzymes encoded by the *pks* island. We recently characterized the chemical events associated with the colibactin self-resistance mechanism (Figure 1B).[4] This process involves the biosynthesis of an inactive metabolite or metabolites (precolibactin) containing an *N*-acylated-D-asparagine introduced by the first module of the assembly line (ClbN). Following elongation by ClbB and the remaining assembly-line enzymes, the prodrug motif is hydrolyzed by the periplasmic peptidase ClbP, an activity that is necessary for production of active genotoxin (colibactin). This biosynthetic proposal has been supported by recent isolation efforts which yielded the prodrug scaffold (**1**) and a potential assembly-line truncation product (Metabolite A, **2**) (Figure 1C).[5] These studies also highlight the challenges of isolating the active product(s) from the *pks* island.

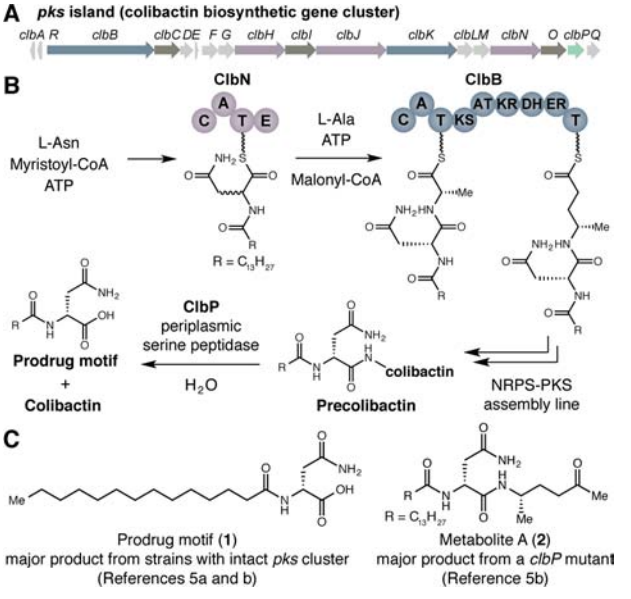


Figure 1. (A) The *pks* island. NRPS (purple), PKS (brown), hybrid NRPS/PKS (blue), and peptidase ClbP (green) are highlighted. (B) Established chemical events in colibactin biosynthesis. C = condensation, A = adenylation, T = thiolation, E = epimerization, KS = ketosynthase, AT = acyltransferase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase, CoA = coenzyme A. (C) Structurally characterized metabolites from *pks*$^+$ *E. coli* strains.

To aid our ongoing efforts to understand colibactin biosynthesis, we sought to identify additional metabolites derived from this gene cluster. We describe here the isolation of Metabolite B (**3**), a new molecule produced by a Δ*clbP E. coli* mutant. This

Received: February 10, 2015
Published: March 10, 2015

DOI: 10.1021/acs.orglett.5b00432
Org. Lett. 2015, 17, 1545−1548

Table 1. NMR Data of Metabolite B in DMSO-d_6

fragment	carbon	δC (type)	δH, multiplicity, J (Hz)	COSY	HMBC (^{1}H to ^{13}C)	ROESY
I	1	14.0 (CH_3)	0.85, t (7.3)	2	2, 3	
	2	22.1 (CH_2)	1.30−1.13, m	1		
	3	31.3 (CH_2)	1.30−1.13, m			
	4−10	29.10, 29.06, 28.99, 28.88, 28.75 (CH_2)	1.30−1.13, m			
	11	28.65 (CH_2)	1.30−1.13, m	12		
	12	25.2 (CH_2)	1.45, m	11, 13	11	13
	13	35.2 (CH_2)	2.08, m	12	11, 12, 14	12, 14NH
	14	172.1 (C)	–	–		
			NH, 7.88, d, (8.1)	15	14	13, 16b
	15	49.8 (CH)	4.47, m	14NH, 16a, 16b	16, 17, 18	18NH
	16	37.4 (CH_2)	a 2.45, dd (7.7, 15.1)	15, 16b	15, 17, 18	
			b 2.30, dd (7.7, 15.1)	15, 16a	15, 17, 18	14NH
	17	171.4 (C)	–			
			a NH, 7.24, s	17bNH		
			b NH, 6.81, s	17aNH		
	18	170.3 (C)	–			
			NH, 7.48, d, (8.2)	19	18	15, 20, 21
II	19	44.1 (CH)	3.71, m	18NH, 20		20, 21
	20	20.5 (CH_3)	1.00, d, (7.3)	19	21	18NH, 19, 21
	21	29.87 (CH_2)	1.60, m	22a, 22b		18NH, 19, 20, 22a, 22b
	22	38.4 (CH_2)	a 2.92, ddd (6, 8.7, 17.3)	21, 22b	21, 23	21
			b 2.82, ddd (6, 8.7, 17.3)	21, 22a	21, 23	21
III	23	197.7 (C)	–			
	24	169.2 (C)	–			
	25	128.9 (C)	–			
	26	10.8 (CH_3)	1.96, s		24, 25, 27	28a, 29a
	27	45.5 (C)	–			
	28	13.7 (CH_2)	a 1.49, m	28b, 29a, 29b	29, 30	26
			b 1.41, m	28a, 29a, 29b		30NH
	29	13.7 (CH_2)	a 1.49, m	28a, 28b, 29b	28, 30	26
			b 1.41, m	28a, 28b, 29a		30NH
	30	169.6 (C)	–			
			NH, 8.51, s		24, 25, 27	28b, 29b

metabolite contains a spiro-cyclopropane, a structural motif rare among NRPS and PKS products and one never previously observed in an *E. coli* metabolite. The structure of **3** suggests that DNA or protein alkylation may be important for genotoxicity. Overall, this work demonstrates the *pks* island encodes novel biosynthetic activity, sheds light on colibactin's biological activity, and expands our knowledge of the biosynthetic capabilities of *E. coli*.

We decided to isolate molecules from a *clbP* mutant for several reasons. Our characterization of the colibactin self-resistance mechanism led us to hypothesize that this deletion would lead to the accumulation of inactive *pks* metabolites containing the N-terminal prodrug motif. We postulated that these metabolites might display increased stability relative to those lacking the motif, as the primary amine released by prodrug hydrolysis could react as a nucleophile in intra- or intermolecular decomposition pathways. Additionally, we envisioned using the structure of the prodrug motif to identify candidate metabolites for isolation.

We began our work by comparing the metabolite profiles of whole-culture methanol extracts of *E. coli* DH10B expressing the wild-type *pks* cluster (BAC*pks*) or the *pks* cluster with *clbP* knocked out (BAC*pks*Δ*clbP*) using the web-based XCMS program (Table S1).[6] Several metabolites were significantly enriched in BAC*pks*Δ*clbP* extracts. The most abundant was the known compound **2** (m/z 462.3386, $[M + Na]^+$ of 439.3410) (Figure 1C), which was identified by Crawford and co-workers as the most abundant metabolite associated with two Δ*clbP* strains.[5b] Another highly abundant molecule, Metabolite B (**3**) (m/z 569.3775, $[M + Na]^+$ of 546.3781), had not been isolated previously. Consistent with earlier studies, the only metabolite significantly enriched in the BAC*pks* extracts was the hydrolyzed prodrug motif (**1**) (m/z 343.2549, $[M + H]^+$ of 342.2519).[5a] When we fed BAC*pks*Δ*clbP* cultures with fully deuterated myristic acid (myr-d_{27}), both **2** and **3** were labeled (Figure S2). Through LC−MS/MS analysis we discovered that, in addition to forming fragments corresponding to loss of the prodrug motif, unlabeled and myr-d_{27} labeled **3** shared a highly unsaturated fragment (m/z 205.1326), which we determined from accurate mass measurement had a molecular formula of $C_{12}H_{17}N_2O$ (**9**) (Figures S3 and S4). From this information, we hypothesized that **3** might contain a heterocycle and would provide an intriguing target for isolation.

To facilitate this effort we sought to increase production of *pks* metabolites in our expression strains. We evaluated the effects of different variables on metabolite production in BAC*pks* cultures using the concentration of the hydrolyzed prodrug motif as a proxy for flux through the biosynthetic pathway (Figure S5). We found that overexpression of the *pks* transcriptional regulator

DOI: 10.1021/acs.orglett.5b00432
Org. Lett. 2015, 17, 1545−1548

clbR resulted in a 5-fold increase in prodrug motif accumulation.[7] Growth in terrific broth (TB) led to a further 4-fold increase compared to growth in Luria−Bertani (LB) media. With these results in hand, we chose to isolate **3** from a DH10B strain grown in TB and harboring both BAC*pks*Δ*clbP* and pTrcHisA-ClbR-N-His_6. We grew 60 L of culture, obtaining 70 g of dried cell mass. Extraction with methanol provided 10.6 g of crude extract that was further fractionated using silica gel flash column chromatography. The 1:1 ethyl acetate/methanol fraction was further purified using two rounds of reversed phase HPLC to obtain ∼5 mg of **3**.

High-resolution mass measurement of the protonated molecular ion for **3** (*m*/*z* 547.3851) yielded a molecular formula of $C_{30}H_{50}N_4O_5$ with 8 degrees of unsaturation. The ^{1}H NMR spectrum in DMSO-d_6 (Figure S7) displayed resonances suggesting the presence of five NH protons at δ8.51 (s), δ7.88 (d, *J* = 8.1 Hz), δ7.48 (d, *J* = 8.2 Hz), δ7.24 (s), and δ6.81 (s); two methines at δ4.47 (m) and δ3.71 (m); a methyl singlet at δ1.96; a methyl doublet at δ1.00 (*J* = 7.3 Hz); a methyl triplet at δ0.85 (*J* = 7.3 Hz); seven methylenes at δ2.92 (ddd, *J* = 6, 8.7, 17.3 Hz), δ2.82 (ddd, *J* = 6, 8.7, 17.3 Hz), δ2.45 (dd, *J* = 7.7, 15.1 Hz), δ2.30 (dd, *J* = 7.7, 15.1 Hz), δ2.08 (m), δ1.60 (m), and δ1.45 (m); and a broad multiplet corresponding to 20 protons at δ1.30−1.13. ^{13}C NMR analysis in DMSO-d_6 (Figure S8) revealed resonances suggesting the presence of a carbonyl at δ197.7; five amide or electron-deficient sp^2-hybridized carbons at δ172.1, δ171.4, δ170.3, δ169.6, and δ169.2; an olefin carbon at δ128.9; three carbons adjacent to amide nitrogens at δ49.8, δ45.5, and δ44.1; three carbons adjacent to carbonyls at δ38.4, δ37.4, and δ35.2; ten alkyl carbons at δ31.3, δ29.87, δ29.10, δ29.06, δ28.99, δ28.88, δ28.75, δ28.65, δ25.2, and δ22.1; and four methyl groups or upfield methylenes at δ20.5, δ14.0, δ13.7, and δ10.8. Complete analysis by both one-dimensional (1D) and two-dimensional (2D) NMR allowed us to complete the structural assignment of Metabolite B (Table 1, Figure 2A).

From our analyses, we concluded that **3** is comprised of an *N*-myristoyl-D-asparagine residue (**I**, Figure 2B) joined by an amide bond to a saturated linker derived from L-alanine (**II**, Figure 2B), which is connected by a carbonyl to an unsaturated heterocycle (**III**, Figure 2B). We propose the absolute stereochemistry of **3** based on analogy to that of *pks* metabolites **1** and **2**, as well as the known biosynthetic activities of ClbN and ClbB. Connectivities between and within fragments **I**−**III** were established by correlation spectroscopy (gCOSY) correlations and homonuclear multiple bond correlation (gHMBCAD) spectroscopy (Table 1 and Figure 2B). While the left-hand portion of **3** (**I** and **II**) is similar to known *pks* metabolites, the right-hand fragment (**III**) is a novel structure composed of an unusual azaspiro[2.4]-heptenone heterocycle.

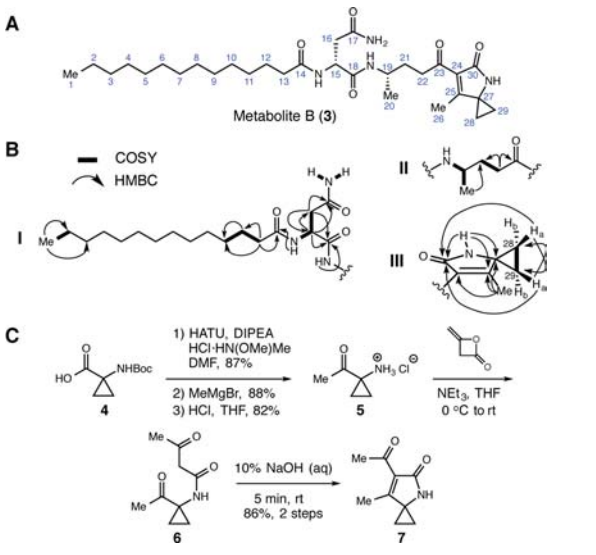


Figure 2. Metabolite B (**3**) possesses an unusual spiro-cyclopropane ring system. (A) The deduced structure of **3**. (B) Key COSY and HMBC correlations establishing the structure of **3**. (C) Synthesis of model compound **7**.

The spiro-cyclopropane ring was the most challenging structural assignment encountered in our analysis. The proton resonances at δ1.49 (28Ha and 29Ha) and δ1.41 (28Hb and 29Hb) showed COSY correlations only to each other. According to heteronuclear single quantum coherence spectroscopy (gHSQCAD), these protons were attached to a carbon with the same chemical shift (δ13.7, 28C and 29C) and the same phase as the methine and methyl carbons. Analysis by distortionless enhancement by polarization transfer (DEPT) ^{13}C NMR unequivocally confirmed this resonance as a methylene (Figures S12−S14). In the DEPT spectrum, the carbon resonance at δ13.7 matched the phase of the other methylene carbons, and in the ^{1}H-coupled DEPT spectrum, the peak appeared as a negative triplet ($^1J_{C-H}$ of 169 Hz).

Distinguishing between potential constitutional isomers of fragment **III** was achieved through analysis of NOE correlations and chemical synthesis. Nuclear Overhauser effect spectroscopy (ROESYAD) revealed multiple through-space correlations within fragment **III**, including correlations from δ1.96 (26H) to δ1.49 (28Ha and 29Ha) as well as correlations from δ1.41 (28Hb and 29Hb) to δ8.51 (30NH) (Figure S15). These data suggested the lactam nitrogen and the allylic methyl carbon were both adjacent to the spiro carbon (27C). To provide additional support for this structural assignment, we synthesized a model of fragment **III** (**7**) in five steps from Boc-protected 1-amino-1-cyclopropane carboxylic acid (**4**) (Figure 2C). The key step in this sequence was an intramolecular Knoevenagel condensation that proceeded rapidly and cleanly to provide the desired spirocycle in excellent yield.[8] The ^{1}H and ^{13}C NMR spectra of this molecule closely matched those of fragment **III** (Figures S16−S17, Table S3). From this comparison we can conclude that **3** and **7** very likely possess related connectivity.

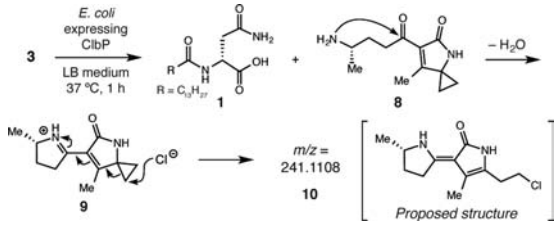
The chemical structure of Metabolite B provides information about some of the biosynthetic events that follow construction of the prodrug scaffold and elongation by ClbB (Scheme S1). Specifically, the construction of **3** could begin with a thiolation (T) domain-tethered thioester generated by ClbB. Decarboxylative Claisen condensation with malonyl-CoA, amide bond formation with 1-amino-1-cyclopropane carboxylic acid, and a second Claisen condensation with malonyl-CoA would generate an enzyme-tethered intermediate that could undergo an intramolecular Knoevenagel condensation similar to that used in the synthesis of **7**. Thioester hydrolysis and decarboxylation would provide **3**. There are no homologues of known cyclopropane biosynthetic enzymes encoded in the *pks* cluster, suggesting that this structural feature may be made using new biosynthetic logic.[9] We do not yet know if **3** is produced by wild-type *pks*$^+$ *E. coli* and whether it is an intermediate or a shunt product in colibactin biosynthesis.

To gain preliminary insights into whether the bicyclic ring system present in Metabolite B is also found in precolibactin, we tested the ability of peptidase ClbP to cleave **3** in an *in vivo* LC-

DOI: 10.1021/acs.orglett.5b00432
Org. Lett. 2015, 17, 1545−1548

MS assay (Scheme 1, Figures S18–S20). We measured the amount of **1** in culture extracts of ClbP-expressing *E. coli* strains

Scheme 1. Metabolite B is Processed by Peptidase ClbP



incubated with **3**, known ClbP substrate *N*-myristoyl-D-asparagine-L-alanine-*O*-methyl ester,[4] or DMSO. *E. coli* expressing full-length ClbP-C-His_6 hydrolyzed both **3** and the control substrate. In assays with **3**, we did not identify masses corresponding to the accumulation of primary amine **8** or imine **9**. Instead we observed a product **10** that had a mass consistent with a chloride adduct (m/z 241.1108). Neither **1** nor **10** was found in assays with cells expressing the inactive mutant ClbP-S95A-C-His_6 or an empty vector. The ability of ClbP to process Metabolite B supports the hypothesis that this spirocyclic ring system is present in precolibactin and/or other ClbP substrates produced by *pks*$^+$ *E. coli*.

Our results raise important questions about the role of ClbP in generating active genotoxin and the mechanism by which colibactin production leads to DNA damage. We postulate that prodrug cleavage could enhance the activity of **3** and related *pks* metabolites in two ways: by releasing metabolites from the *E. coli* membrane through removal of the lipophilic *N*-acyl chain and by generating a more potent electrophile, an α,β-unsaturated iminium with an adjacent cyclopropyl group, that could alkylate DNA or a target protein.[10] Intriguingly, this activation strategy bears resemblance to, but is unique from, those involved in modulating the activities of other DNA-alkylating natural products that contain cyclopropanes,[11] including the CC-1065/duocarmycins[12] (conformational change upon DNA binding) and the illudins/acylfulvenes[13] (activation via reductive metabolism). Based on this precedent, we have tentatively proposed a cyclopropane-opened structure for **10**.

In conclusion, we have successfully isolated a complex metabolite associated with the *pks* island that will guide further studies of colibactin biosynthesis and biological activity. The structure of Metabolite B raises intriguing questions about how the azaspiro[2.4] bicyclic ring system is biosynthesized by the *pks* assembly line enzymes and how this structure may contribute to colibactin's genotoxicity. Beyond its impact in the area of natural product discovery, this work will aid future efforts to understand how *pks*$^+$ *E. coli* impact human health and disease.

■ ASSOCIATED CONTENT

Supporting Information

Full experimental and characterization data, including ^{1}H and ^{13}C NMR for all new compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: balskus@chemistry.harvard.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Gregory Heffron (Harvard Medical School, Boston, MA) for assistance with NMR experiments, Sunia Trauger (Small Molecule Mass Spectrometry Facility, Harvard University, Cambridge, MA) for help with LC–MS analyses, and Li Zha (Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA) and Pedro Leão (Center for Marine and Environmental Research, University of Porto, Porto, Portugal) for helpful discussions. We acknowledge financial support from the Damon Runyon-Rachleff Innovation Award, the Smith Family Award for Excellence in Biomedical Research, and the Packard Fellowship for Science and Engineering.

■ REFERENCES

(1) (a) Nougayrède, J. P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Science* **2006**, *313*, 848. (b) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrède, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11537.

(2) Arthur, J. C.; Perez-Chanona, E.; Mühlbauer, M.; Tomkovich, S.; Uronis, J. M.; Fan, T. J.; Campbell, B. J.; Abujamel, T.; Dogan, B.; Rogers, A. B.; Rhodes, J. M.; Stintzi, A.; Simpson, K. W.; Hansen, J. J.; Keku, T. O.; Fodor, A. A.; Jobin, C. *Science* **2012**, *338*, 120.

(3) (a) Bondarev, V.; Richter, M.; Romano, S. *Environ. Microbiol.* **2013**, *15*, 2095–2113. (b) Engel, P.; Vizcaino, M. I.; Crawford, J. M. *Appl. Environ. Microbiol.* **2015**, *81*, 1502.

(4) Brotherton, C. A.; Balskus, E. P. *J. Am. Chem. Soc.* **2013**, *135*, 3359.

(5) (a) Bian, X.; Fu, J.; Plaza, A.; Herrmann, J.; Pistorius, D.; Stewart, A. F.; Zhang, Y.; Müller, R. *ChemBioChem* **2013**, *14*, 1194. (b) Vizcaino, M. I.; Engel, P.; Trautman, E.; Crawford, J. M. *J. Am. Chem. Soc.* **2014**, *136*, 9244.

(6) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035.

(7) Krumbholz, G. Ph.D. Thesis, Julius-Maximilians-Universität Würzburg, 2010.

(8) Kato, T.; Sato, M.; Yoshida, T. *Chem. Pharm. Bull.* **1971**, *19*, 292.

(9) Thibodeaux, C. J.; Chen, W.; Liu, H. *Chem. Rev.* **2012**, *112*, 1681.

(10) (a) Wong, H. N. C.; Hon, M.; Tse, C.; Yip, Y.; Tanko, J.; Hudlicky, T. *Chem. Rev.* **1989**, *89*, 165. (b) Bose, G.; Bracht, K.; Bednarski, P. J.; Lalk, M.; Langer, P. *Bioorg. Med. Chem.* **2006**, *14*, 4694.

(11) (a) Wolkenberg, S. E.; Boger, D. L. *Chem. Rev.* **2002**, *102*, 2477. (b) Gersch, M.; Kreuzer, J.; Sieber, S. A. *Nat. Prod. Rep.* **2012**, *29*, 659.

(12) (a) Gosh, N.; Sheldrake, H. M.; Searcey, M.; Pors, K. *Curr. Top. Med. Chem.* **2009**, *9*, 1494. (b) Boger, D. L.; Garbaccio, R. M. *Acc. Chem. Res.* **1999**, *32*, 1043. (c) Boger, D. L.; Johnson, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3642.

(13) Tanasova, M.; Sturla, S. *Chem. Rev.* **2012**, *112*, 3578.

DOI: 10.1021/acs.orglett.5b00432
Org. Lett. 2015, 17, 1545–1548