JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Assembly of Asperlicin Peptidyl Alkaloids from Anthranilate and Tryptophan: A Two-Enzyme Pathway Generates Heptacyclic Scaffold Complexity in Asperlicin E

Stuart W. Haynes,[†] Xue Gao,[‡] Yi Tang,*[,‡,§] and Christopher T. Walsh*[,†]

[†]Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, United States

[‡]Department of Chemical and Biomolecular Engineering, [§]Department of Chemistry and Biochemistry, University of California Los Angeles, 420 Westwood Plaza, Los Angeles, California 90095, United States

Supporting Information

ABSTRACT: Members of the asperlicin family of fungal metabolites produced by *Aspergillus alliaceus* are known potent CCK_A antagonists. Herein, we report the identification of the gene cluster responsible for directing their biosynthesis. We validate and probe the pathway by genetic manipulation, and provide the first biochemical characterization of the oxidative cyclization en route to the heptacyclic asperlicin E by reconstituting the activity of the FAD depend monooxygenase AspB. This report provides the first genetic characterization of a NRPS assembly line that efficiently activates two anthranilate building blocks and illustrates the remarkably efficient biosynthesis of the complex heptacyclic asperlicin E.

Aspergillus fungal strains produce a variety of peptidyl alkaloids that act as mycotoxins. Several of these fungal alkaloids incorporate anthranilate as a nonproteinogenic amino acid building block for two and three module nonribosomal peptide synthetase (NRPS) assembly lines.[1−4] Members of the asperlicin family of metabolites (Figure 1), produced by several strains of *Aspergillus alliaceus*, were of particular interest when discovered in the 1980s as selective antagonists of the cholecystokinin receptor CCK_A.[5,6] The major metabolite asperlicin, with its peptide scaffold biosynthetically morphed into a tetracyclic quinazoline [3,2-*a*][1,4] benzodiazepine-5,13-dione core, was a lead scaffold for subsequent high affinity, selective CCK_A ligands.

We have recently undertaken biosynthetic studies of related Aspergillus peptidyl alkaloids in the acetylaszonalenin, fumiquinazoline,[1,7−9] and tryptoquialanine[4] series. These studies centered on the utilization of anthranilate as a noncanonical aryl-*β*-amino acid starter unit in the NRPS assembly lines in tandem with L-tryptophan and L-alanine as additional building blocks. We deciphered the fungal code for anthranilate selection by NRPS modules to enable identification of the relevant biosynthetic gene clusters among a large number of NRPS candidates in aspergillus genomes.[8]

In this work, we have turned our attention to the *A. alliaceus* asperlicin producers for two reasons. First, we predict that the asperlicin NRPS assembly line should select and activate anthranilate for *both* the first and second residues (with Trp as the third building block) and that should set up formation of the *tetracyclic* quinazoline benodiazepinedione core, e.g., in asperlicins C and D (Figure 1) representing a dramatic morphing of an Ant-Ant-Trp tripeptide framework. Second, the presumed downstream asperlicin E metabolite has a fused angular, heptacyclic 6-6-7-6-5-5-6 ring scaffold, representing significant architectural complexity compared to the starting amino acids. Here, we report identification of the asperlicin gene cluster and characterization of a short, efficient two-enzyme pathway to asperlicin E.

A. alliaceus strain (ATCC 20656) was obtained from ATCC and grown in glucose minimal media (GMM) to induce asperlicin production, allowing detection of known secreted metabolites asperlicin C (expected 407.1503, found 407.1505), asperlicin D (expected 407.1503, found 407.1504), asperlicin E (expected 423.1452, found 423.1452) and the mature metabolite asperlicin (expected 536.2292, found 536.2293) by high resolution LC/MS analysis (structures in Figure 1). Authentic samples of enantiomeric pairs of both regioisomeric asperlicin C and asperlicin D were synthesized by microwave mediated methodology previously described for the synthesis of asperlicin C (see Supporting Information) and these were used as synthetic standards to confirm the identity of those two metabolites (Figures S5 and S6).[10] Genomic DNA was prepared and subjected to whole genome sequencing with approximately 575-fold coverage on an Illumina machine (Table S1).[11] A bioinformatic search for anthranilate-selective adenylation domains[8] turned up an orf we have designated as AspA (Figure 1, Tablea S2 and S3) as a proposed bimodular anthranilate-activating NRPS (predicted domains A-T-C-A-T-C_T , where A = Adenylation, T = Thiolation, and C = Condensation). Immediately adjacent is AspB (predicted to be an FAD-dependent oxygenase) and on the other DNA strand upstream AspC, predicted to be a valine-activating monomodular NRPS (A-T-C). This three gene organization is reminiscent of the fumiquinazoline biosynthetic genes we have previously described in *Aspergillus fumigatus* (Af12080, trimodular NRPS that synthesizes and activates an Ant-Trp-Ala tripeptide to perform a double cyclization/dehydration to the tricyclic fumiquinazoline F;[12] Af12060, FAD indole

Received: August 23, 2012
Published: October 2, 2012

   dx.doi.org/10.1021/ja308371z | *J. Am. Chem. Soc.* 2012, 134, 17444−17447

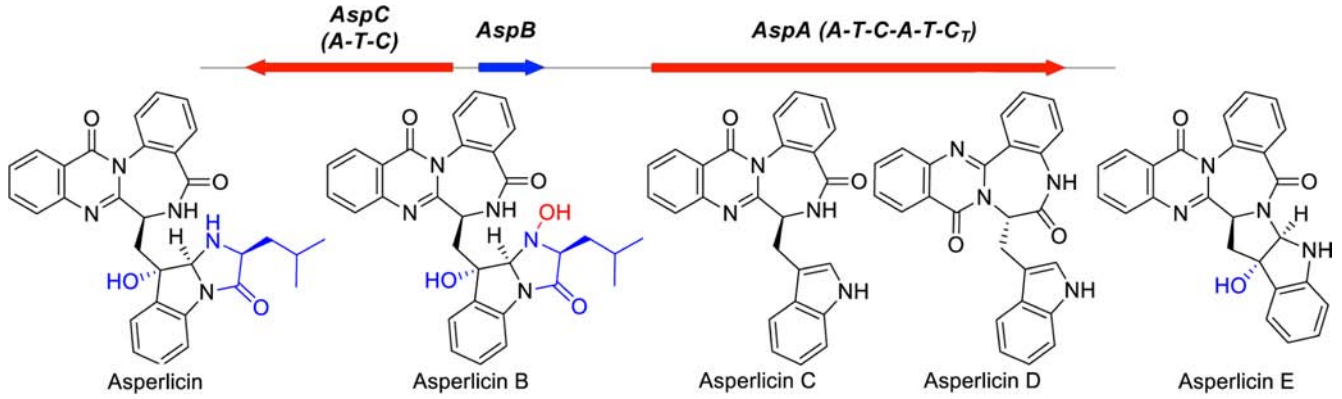

Figure 1. Schematic of *asp* cluster and structures of asperlicins.

epoxygenase; and Af12050, monomodular Ala-activating NRPS to create the FQA metabolite from FQF).[1] Notably however, the Asp cluster represents the first genetic characterization of a cluster responsible for the double incorporation of (the noncanonical aryl-*β*-amino acid) anthranilate. It was therefore surprising that AspA encodes only a bimodular NRPS (with predicted domains A-T-C-A-T-C_T) instead of the trimodular domain one would have anticipated for construction of a presumed Ant-Ant-Trp-enzyme tethered intermediate.

To validate the proposed function of this gene cluster, a double crossover deletion of the bimodular orf *aspA* was constructed (Figure S5)[13,14] in the producing *A. alliaceus* strain, resulting in loss of any detectable asperlicin metabolites, including asperlicins C, D, E and asperlicin itself (Figure S4b). Feeding of synthetic asperlicin C to the AspA mutant restored production of asperlicin and asperlicin E, whereas feeding of asperlicin D to this mutant did not result in the restoration of any downstream products (Figure S4c,d).

AspB, the predicted FAD binding enzyme, is highly homologous to Af12060 (Figure S1) from the fumiquinazoline pathway that we have previously purified and demonstrated to activate the pendant indole ring of FQF as an epoxide (that may be in equilibrium with the hydroxyiminium tautomer of the indole).[9] Intriguingly, in this case, the product of AspB (unlike the product of Af12060) is seen to undergo one of two possible fates: either immediate intramolecular capture en route to asperlicin E, or with the involvement of an additional enzyme (AspC), further processing to asperlicin. As predicted, by comparison to the fumiquinazoline pathway, when *aspB* is deleted via double crossover (via the same methodology to used to knock out *aspA*), production of asperilicin and asperlicin E is abolished and levels of asperlicin C are considerably increased (Figure S4e). AspB was then expressed and purified from *Escherichia coli* as a yellow-pigmented protein, characteristic of flavin-containing enzymes (Figure S3). We compared purified AspB and Af12060 for their ability to convert synthetic asperlicin C or D (and their enantiomers) in the presence of proposed cosubstrate O_2 to new products. Af12060 showed no activity (Figure S2). However, pure AspB indeed converted both enantiomers of synthetic asperlicin C to two distinct new products (Figure 2). High-resolution LC-MS analyses of both products were consistent with the formation of heptacyclic asperlicin E (or isomers) (expected 423.1452, found 423.1452), via intramolecular oxygenation and subsequent cyclization. Preparative incubations of asperlicin C with AspB allowed isolation of sufficient product for NMR spectroscopic characterization and confirmation of the AspB enzymatic

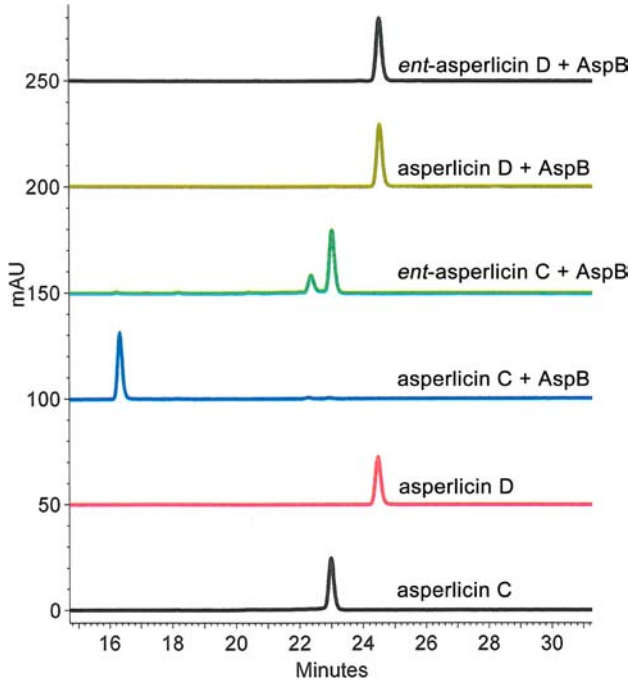

Figure 2. HPLC analysis (at 274 nm) of *in vitro* enzymatic reaction of AspB with enantiomeric pairs of asperlicin C and D.

product as asperlicin E by comparison with the previously characterized natural product (Figures S7 and S9). The instability of the product generated from the reaction of AspB with the enantiomer of asperlicin C (*ent*-asperlicin C) hindered full characterization. However, the data recorded (and the marked acceleration of its degradation under mildly acidic conditions) are consistent with the formation of an oxidized product, which is stalled prior to intramolecular cyclization en route to an asperlicin E type product (Figures S8 and S9). Interestingly, the apparent inactivity of *ent*-asperlicin C to intramolecular cyclization suggests that the stereochemistry of the Trp side chain is vital to correctly position the amide in the 7-membered diazepinone for intramolecular attack and, furthermore, is an indication of the precision of energetic fine-tuning of the cascade leading to the formation of asperlicin E.

By contrast asperlicin D was completely inactive with AspB as was fumiquinazoline F. Correspondingly, the genetic insertion to knockout AspB (noted above) results in the accumulation of much more asperlicin C than asperlicin D

dx.doi.org/10.1021/ja308371z | J. Am. Chem. Soc. 2012, 134, 17444−17447

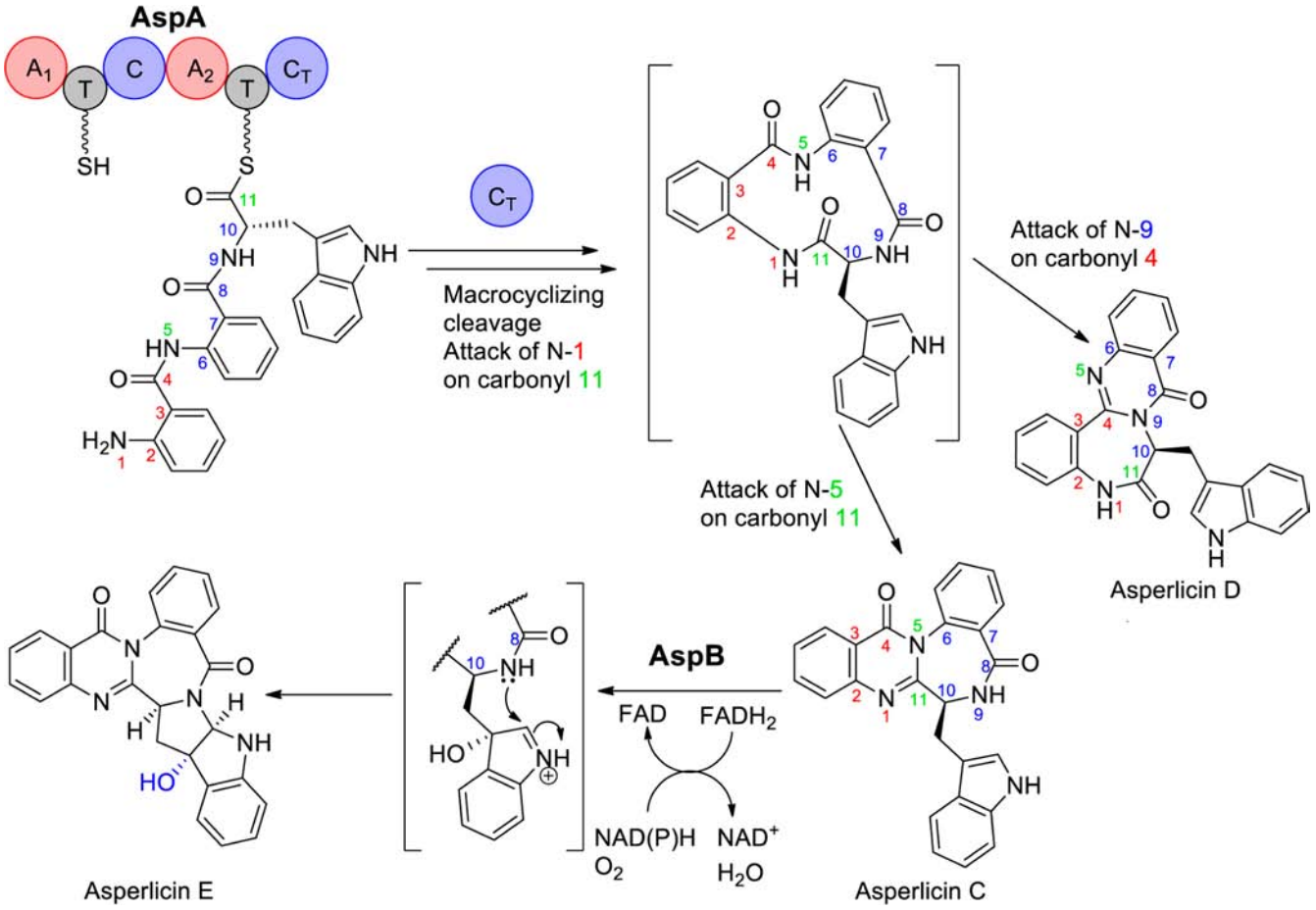

Figure 3. Asperlicin biosynthetic pathway en-route to heptacyclic asperlicin E, mediated by the action of AspA and AspB.

(Figure S4), which is consistent with the further metabolism of asperlicin C (but not asperlicin D) to further downstream products (asperlicin E and asperlicin) in the native producer.

Thus, only two enzymes, the bimodular NRPS AspA and the FAD-enzyme AspB, suffice to take two anthranilates and one tryptophan building block and assemble the constrained heptacyclic peptidyl alkaloid scaffold of asperlicin E. Figure 3 shows the proposed mechanism of these two enzymes: AspA generating a tetracyclic scaffold and AspB converting it via intramolecular capture of an oxygenated intermediate to the heptacyclic framework.

Given only one predicted Ant-activating module (module 1)[8] and one Trp activating module (module 2) in the bimodular AspA, we presume module 1 must act iteratively to select, activate, and sequentially load two anthranilyl units onto the T_1 domain. This could happen via an anthranilyl-anthranilyl-dipeptidyl thioester on T_1 before transfer to the amino group of Trp tethered on T_2, or could involve an anthranilyl thioester on T_1 to condense with an anthranilyl-Trp-thioester that has built up on T_2. The first condensation domain may be required for the formation of the iterative buildup of the tethered tripeptidyl thioester while the terminal C domain is likely required for the cyclization/release steps.[12] At this stage, it remains a possibility that a further enzyme (such as a standalone C-domain to act *in trans*) may be involved in the formation of Ant-Ant dipeptide if such an enzyme is not clustered with the Asp cluster. However, our bioinformatic investigation of the genome only identifies one Ant-activating A-domain. We are therefore confident that the Ant-activating A-domain of AspA must be acting iteratively.

Such an acyclic Ant-Ant-Trp-*S*-pantetheinyl-thioester tethered to the second thiolation (T_2) domain, as shown in Figure 3, must then undergo double cyclization. The first cyclization is likely to be from attack of the free amino group of Ant1 on the Trp3 thioester carbonyl (N_1 on carbonyl$_{11}$) to yield the 6, 11-bicyclic system while cleaving the tethered peptide chain from the NRPS. Subsequent transannular attack of the N_5 amide nitrogen on carbonyl$_{11}$ with dehydration would give the tetracyclic asperlicin C framework. Alternative capture of amide nitrogen N_9 on carbonyl$_4$ (with subsequent dehydration) yields an alternate 6-6-7-6 regiosiomeric tetracyclic core in asperlicin D (Figure 3). The presence of a mixture of asperlicin C and D regioisomers in producer extracts is consistent with an uncatalyzed transannular cyclization occurring after formation of the 6,11 bicyclic intermediate and post release from the NRPS active site. It appears as though the asperlicin D isomer is a dead end side product since it is not processed further by the flavoenzyme AspB.

We formulate the AspB mediated transformation of asperlicin C to asperlicin E via formation of the transient indole epoxide or a hydroxyiminium adduct (via a FAD-4a-OOH electrophilic oxygen donation to the indole 2, 3-double bond), with presumed alpha stereochemistry, as in the related Af12060-mediated FQF to FQA conversion.[1,9] That intermediate would undergo intramolecular capture by the amide in the 7-membered diazepinone ring of asperlicin C (but seemingly not in the diasteromer generated with *ent*-asperlicin C). The poor nucleophilicty of the amide -NH is overcome in this system by generating a strong electrophile and by the

dx.doi.org/10.1021/ja308371z | J. Am. Chem. Soc. 2012, 134, 17444−17447

effective high local concentration of the nucleophile due to the intramolecular nature of the addition. The action of AspB installs the −OH group and creates a N−C bond resulting in the linkage of the bicyclic indole moiety to what had been the tetracyclic core by the formation of a new 5-membered pyrroline ring. This generates the 6-6-7-6-5-5-6 heptacyclic framework of asperlicin E (where the underlined ring is the one just created).

This is a remarkable two-enzyme pathway: AspA and AspB start with three amino acids, two of them anthranilates and one tryptophan, and produce the highly constrained seven ring scaffold. Notable is the tandem use of two anthranilates to snap shut a 6-6-7-6 tetracyclic core, followed by the action of AspB and subsequent intramolecular capture that converts the bicyclic indole to a tricyclic pyrroloindole moiety, which is in turn fused to the four ring substructure.

The third enzyme encoded in the gene cluster, AspC, is proposed to be a valine activating monomodular NRPS, akin to the alanine activating NRPS that captures the epoxygenated FQF species and annulates it to fumiquinazoline A.[1] To generate asperlicin, AspC presumably provides a tethered valyl thioester to compete with the intramolecular capture of the hydroxyiminium ion that would yield asperlicin E as shown in Figure 3. Instead, the NH_2 of the valyl moiety tethered to AspC could capture the oxidized intermediate, thus annulating the indole moiety to yield asperlicin, a distinct scaffold where the seven rings systems are in separate tetracyclic and tricyclic substructures (Figure S10). Analysis of that competition will require purified active AspC in subsequent studies. Meanwhile, the *aspC* knockout mutant gives asperlicin E as the major product (Figure S10), consistent with shutting down the conversion of asperlicin C to asperlicin.

Anthranilate is the key nonproteinogenic building block, in combination with tryptophan, for Aspergilli to fashion a variety of complex peptidyl alkaloid scaffolds including fumiquinazolines,[2] tryptoquialanines,[4] azonalenins,[3] and ardeemins[15,16] in short, efficient biosynthetic pathways. The tandem iterative utilization of the nonproteinogenic anthranilate in the asperlicin pathway, and related pathways builds a unique tetracyclic core scaffold that can be further elaborated to architectural complexity in related scaffolds seen in the benzomalvins,[17,18] circumdatins,[19−22] and sclerotigenin.[23]

■ ASSOCIATED CONTENT

Supporting Information

Experimental procedures, synthesis of asperlicin C and D, isolation of genomic DNA, purification of all small molecules, purification of AspB, details of enzymatic reactions, protocol of genetic disruption, growth and extraction of fungal cultures, enzymatic reactions with Af12060, estimation of bound flavin to AspB, profiles of asperlicins with mutant strains, LC-HRMS comparisons of synthesized asperlicins with standards and NMR spectroscopic data. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

christopher_walsh@hms.harvard.edu; yitang@ucla.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Steven Malcolmson, Jared Parker and Thomas Gerken for their helpful advice and discussion. We also thank the NIH GM20011, GM49338 (to C.T.W.) and GM092217 (to Y.T.) for funding.

■ REFERENCES

(1) Ames, B. D.; Liu, X.; Walsh, C. T. *Biochemistry* **2010**, *49*, 8564.
(2) Takahashi, C.; Matsushita, T.; Doi, M.; Minoura, K.; Shingu, T.; Kumeda, Y.; Numata, A. *J. Chem. Soc., Perkin Trans. 1* **1995**, 2345.
(3) Yin, W. B.; Grundmann, A.; Cheng, J.; Li, S. M. *J. Biol. Chem.* **2009**, *284*, 100.
(4) Gao, X.; Chooi, Y. H.; Ames, B. D.; Wang, P.; Walsh, C. T.; Tang, Y. *J. Am. Chem. Soc.* **2011**, *133*, 2729.
(5) Chang, R.; Lotti, V.; Monaghan, R.; Birnbaum, J.; Stapley, E.; Goetz, M.; Albers-Schonberg, G.; Patchett, A.; Liesch, J.; Hensens, O.; et al. *Science* **1985**, *230*, 177.
(6) Liesch, J. M.; Hensens, O. D.; Zink, D. L.; Goetz, M. A. *J. Antibiot.* **1988**, *41*, 878.
(7) Ames, B. D.; Haynes, S. W.; Gao, X.; Evans, B. S.; Kelleher, N. L.; Tang, Y.; Walsh, C. T. *Biochemistry* **2011**, *50*, 8756.
(8) Ames, B. D.; Walsh, C. T. *Biochemistry* **2010**, *49*, 3351.
(9) Haynes, S. W.; Ames, B. D.; Gao, X.; Tang, Y.; Walsh, C. T. *Biochemistry* **2011**, *50*, 5668.
(10) Liu, J. F.; Kaselj, M.; Isome, Y.; Chapnick, J.; Zhang, B.; Bi, G.; Yohannes, D.; Yu, L.; Baldino, C. M. *J. Org. Chem.* **2005**, *70*, 10488.
(11) Ambry Genetics, California, USA. Assuming genome size of approximately 30 Mb (by analogy with previously fully assembled filamentous fungi).
(12) Gao, X.; Haynes, S. W.; Ames, B. D.; Wang, P.; Vien, L. P.; Walsh, C. T.; Tang, Y. *Nat. Chem. Biol.* **2012**, *8*, 823.
(13) Nayak, T.; Szewczyk, E.; Oakley, C. E.; Osmani, A.; Ukil, L.; Murray, S. L.; Hynes, M. J.; Osmani, S. A.; Oakley, B. R. *Genetics* **2006**, *172*, 1557.
(14) Szewczyk, E.; Nayak, T.; Oakley, C. E.; Edgerton, H.; Xiong, Y.; Taheri-Talesh, N.; Osmani, S. A.; Oakley, B. R. *Nat. Protoc.* **2006**, *1*, 3111.
(15) Hochlowski, J. E.; Mullally, M. M.; Spanton, S. G.; Whittern, D. N.; Hill, P.; McAlpine, J. B. *J. Antibiot.* **1993**, *46*, 380.
(16) Karwowski, J. P.; Jackson, M.; Rasmussen, R. R.; Humphrey, P. E.; Poddig, J. B.; Kohl, W. L.; Scherr, M. H.; Kadam, S.; McAlpine, J. B. *J. Antibiot.* **1993**, *46*, 374.
(17) Sun, H. H.; Barrow, C. J.; Sedlock, D. M.; Gillum, A. M.; Cooper, R. *J. Antibiot.* **1994**, *47*, 515.
(18) Sun, H. H.; Barrow, C. J.; Cooper, R. *J. Nat. Prod.* **1995**, *58*, 1575.
(19) Rahbaek, L.; Breinholt, J.; Frisvad, J. C.; Christophersen, C. *J. Org. Chem.* **1999**, *64*, 1689.
(20) Rahbaek, L.; Breinholt, J. *J. Nat. Prod.* **1999**, *62*, 904.
(21) Ookura, R.; Kito, K.; Ooi, T.; Namikoshi, M.; Kusumi, T. *J. Org. Chem.* **2008**, *73*, 4245.
(22) Zhang, D.; Yang, X.; Kang, J. S.; Choi, H. D.; Son, B. W. *J. Antibiot.* **2008**, *61*, 40.
(23) Joshi, B. K.; Gloer, J. B.; Wicklow, D. T.; Dowd, P. F. *J. Nat. Prod.* **1999**, *62*, 650.