Published in final edited form as:

Chembiochem. 2014 April 14; 15(6): 815–821. doi:10.1002/cbic.201300756.

Characterization of an environmental DNA derived gene cluster that encodes for the bisindolylmaleimide methylarcyriarubin

Fang-Yuan Chang^[a] [Prof.] and Sean F. Brady^{*,[a]}

^[a]Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University 1230 York Avenue, New York, NY 10065

Abstract

Bisindolylmaleimides represent a naturally occurring class of metabolites that are of interest because of their protein kinase inhibition activity. From a metagenomic library constructed using soil DNA, we identified the four gene *mar* cluster, a bisindolylmaleimide gene cluster that encodes for methylarcyriarubin (**1**) production. Heterologous expression of the *mar* gene cluster in *Escherichia coli* revealed that the Rieske dioxygenase, MarC, facilitates the oxidative decarboxylation of a chromopyrrolic acid (CPA) intermediate to yield the bisindolylmaleimide core. The characterization of the *mar* cluster defines a new role for CPA in the biosynthesis of structurally diverse bacterial tryptophan dimers.

Keywords

biosynthesis; bisindolylmaleimide; metagenomics; natural products; Rieske dioxygenase

Bisindolylmaleimide natural products share a common 3,4-di-1H-indol-3-yl-1H-pyrrole-2,5-dione core structure^[1] (Scheme 1, red). The discovery of the simplest bisindolylmaleimide natural product, arcylarubin A,^[2] from a slime mold (*Arcyria denudate*) spearheaded the extensive study of this class of compounds, with more than 2,400 and 4,000 bisindolylmaleimide-related references in the PubMed and SciFinder databases, respectively. Bisindolylmaleimide analogs, with activities in cancer,^[3] diabetes,^[4] cardiovascular^[5] and neurodegenerative^[6] disease models have now been synthesized, some of which have advanced into clinical trials^[7] (Scheme 1A). Surprisingly, bisindolylmaleimide biosynthesis has remained uncharacterized.

In addition to being isolated from slime molds, bisindolylmaleimides have also now been found in bacterial culture broth extracts, suggesting the presence of bisindolylmaleimide biosynthetic gene clusters in bacterial genomes.^[8] Soil is thought to contain thousands of bacterial species per gram, with as much as 99% of these bacteria being recalcitrant to culturing using standard methods.^[9] The study of natural microbial communities using culture independent methods provides a means of systematically exploring large numbers of biosynthetic gene clusters arising from diverse bacterial species. Here, we describe the first

*Fax: (+1) 212-327-8280 sbrady@rockefeller.edu.

Supporting information for this article is available on the WWW under <http://www.chembiochem.org>.

functional characterization of bisindolylmaleimide biosynthesis using a gene cluster recovered from a soil environmental DNA (eDNA) library (Scheme 1B). Our *in vivo* analysis of the eDNA-derived methylarcyriarubin (*mar*) gene cluster reveals that a novel dioxygenase, MarC, is responsible for the formation of the bisindolylmaleimide core through the oxidative decarboxylation of a chromopyrrolic acid intermediate.

Based on its bisindole chemical structure, bisindolylmaleimide is predicted to arise from the dimerization of two tryptophans. Among structurally characterized bacterial tryptophan dimers, the biosyntheses of violacein and a number of indolocarbazole compounds have been described.^[10] In each case, the biosynthesis begins with the oxidation of L-tryptophan by an indole-3-pyruvic acid imine (IPA imine) synthase (e.g. VioA, RebO, StaO) and then proceeds through the dimerization of IPA imine by a chromopyrrolic acid (CPA) synthase (e.g. VioB, RebD, StaD) to produce CPA (Scheme 2). Because of the structural resemblance of CPA to bisindolylmaleimide, it can be hypothesized that bisindolylmaleimide biosynthesis starts in a similar manner and then diverges to form the bisindolylmaleimide scaffold by the oxidative decarboxylation of CPA. Previously reported *in vivo* and *in vitro* expression studies using indolocarbazole and violacein biosynthetic genes have found that bisindolylmaleimides (e.g. arcyrarubin A) are neither consumed as substrates nor accumulated as intermediate products in these biosynthetic pathways,^[11] suggesting that the diverse collection of known tryptophan dimer modification enzymes is unlikely to be sufficient for producing bisindolylmaleimide structures from CPA.

CPA synthase gene sequences are highly conserved across all known bacterial tryptophan dimer gene clusters. In previous studies, we have used degenerate PCR primers designed to recognize conserved regions in CPA synthase genes to screen soil eDNA libraries for potential CPA homologs.^[12] In these studies, novel CPA-related amplicons were then used to guide the recovery of eDNA clones containing gene clusters that were bioinformatically predicted to encode a diverse collection of CPA-derived metabolites. Among these is the previously uncharacterized *mar* gene cluster that is encoded on cosmid NM343. Cosmid NM343 was recovered from an eDNA library constructed using Chihuahuan desert soil collected in New Mexico (Figure 1, S1). The *mar* cluster is predicted to contain four genes: a CPA synthase (*marB*), a dioxygenase (*marC*), a *vioE* homolog (*marE*) and a methyltransferase (*marM*). The dioxygenase MarC is unprecedented in known tryptophan dimer biosynthetic pathways and we hypothesized that it might direct the formation of a core tryptophan dimer structure that differed from those that had been biosynthetically characterized to date. To investigate this possibility, we chose to study the *mar* cluster using heterologous expression methods.

Initially, the native *mar* cluster was introduced into a variety of model hosts (e.g. *Escherichia coli*, *Streptomyces* spp., *Burkholderia* spp.) for expression studies, but no clone-specific metabolites were detected in the culture broths. In an effort to address potential transcriptional inefficiencies of *mar* gene cluster promoters in these hosts, we individually cloned the *mar* biosynthetic genes in front of T7 promoters and introduced these constructs into *E. coli*. Unfortunately, this also failed to result in the production of any detectable clone-specific small molecules by *E. coli* cultures (Figure 2, MarBCEM).

Although all of the genetic information required to encode the biosynthesis of a bacterial natural product is generally found clustered on a bacterial chromosome, there are a number of examples where a required biosynthetic intermediate is encoded elsewhere in the genome. For example, genes encoding for the biosynthesis of deoxysugar precursors, which are required for many glycosylated compounds, are often found elsewhere in the genome and must, therefore, be introduced *in trans* in heterologous expression studies.^[13] With this in mind, we investigated the possibility that the biosynthesis of a precursor required by the *mar* pathway might not be encoded within the *mar* cluster.

Functionally characterized CPA synthases from other tryptophan dimer biosynthetic clusters have been found to accept oxidized tryptophan (IPA imine), but not tryptophan itself, as a substrate.^[14] Neither the *mar* cluster nor the *E. coli* genome contains an IPA imine synthase homolog. Therefore, if MarB functions as a CPA synthase as predicted by its high sequence identity to known CPA synthases, an IPA imine synthase would have to be supplied *in trans* for *mar* biosynthesis to proceed in a heterologous expression setting. A number of sequenced bacterial genomes contain isolated predicted IPA imine synthase genes, suggesting that IPA imine production may be encoded outside secondary metabolism in a variety of bacteria. Therefore, we co-expressed the IPA imine synthase *vioA* from the violacein cluster with the rest of the *mar* biosynthetic genes in *E. coli*. This resulted in the production of a clone-specific metabolite (**1**) (Figure 2, VioA + MarBCEM), which we had not observed in any previous tryptophan dimer studies, along with the expected tryptophan dimer intermediates, IPA (**3**) and CPA (**4**).

Compound **1** (1.6 mg/L) was purified from large-scale cultures of *E. coli* transformed with the VioA + MarBCEM expression constructs. The structure of **1** was solved using a combination of NMR, UV and HRMS data (Figure S2, S3) and determined to be the methylated bisindolylmaleimide methylarcyriarubin (**1**). Although methylarcyriarubin has been made synthetically, to the best of our knowledge, this is the first reported case of methylarcyriarubin being isolated as a natural product.

To elucidate the role of the individual *mar* biosynthetic genes in the biogenesis of methylarcyriarubin (**1**), culture broth extracts from *E. coli* strains expressing different combinations of the *mar* biosynthetic genes were characterized in detail. In *E. coli* lacking *marM* (Figure 2, VioA+MarBCE), arcyriarubin A (**2**), the desmethyl form of methylarcyriarubin (**1**), was produced along with **3** and **4**. Without the dioxygenase *marC* (Figure 2, VioA+MarBEM; VioA+MarB), accumulation of **3** and **4** was observed. Replacement of MarB with the characterized CPA synthase from the violacein pathway (VioB) resulted in an identical metabolic profile (Figure 2, VioA+MarB; VioAB). No difference in metabolic profile was observed between *E. coli* expressing all the *mar* biosynthetic genes (Figure 2, VioA+MarBCEM) and a strain lacking the predicted *vioE* homolog, *marE* (Figure 2, VioA+MarBCM). Based on these expression studies, the biosynthesis of methylarcyriarubin is predicted to share the same two initial biosynthetic transformations as all other biosynthetically characterized bacterial tryptophan dimers (Scheme 3). Specifically, tryptophan is first oxidized to IPA imine by an IPA imine synthase that is found outside the *mar* cluster in the endogenous host's genome. MarB then dimerizes IPA imine to give CPA. The *mar* biosynthetic pathway is then predicted to diverge from

known tryptophan dimer pathway in that MarC appears to function as a bisindolylmaleimide synthase by converting CPA into arcyrarubin A. MarM is predicted to methylate the bisindolylmaleimide to yield methylarcyriarubin (**1**).

MarC is functionally similar to RebC from rebeccamycin biosynthesis in that they both produce a maleimide moiety from CPA. RebC is an FAD-binding monooxygenase that reacts in tandem with RebP to produce maleimide indolocarbazole.^[15] Based on cocrystallization studies,^[16] the likely substrate of RebC was found to be 7-carboxy-K252c, which is produced by RebP via the C-5 and C-5' aryl-aryl coupling of CPA. This led to a proposed mechanism for RebC involving hydroxylation of 7-carboxy-K252c at the α -carbon of the carboxyl group to facilitate decarboxylation and yield a pyrrole-diol moiety, followed by an oxidative rearrangement to generate the maleimide (Scheme 4). While MarC is also responsible for generating a maleimide moiety, it is predicted, based on sequence homology, to be a Rieske type dioxygenase. Rieske type dioxygenase contains a [2Fe-2S] iron-sulfur cluster, instead of FAD, as the cofactor.^[17] In MarC dependent bisindolylmaleimide biosynthesis, we propose that MarC hydroxylates the C2-C3 and C2'-C3' olefins of CPA in two successive oxidations to facilitate two decarboxylations and generate, without relying on a second enzyme, the pyrrole-diol that can similarly undergo oxidative rearrangement to yield the maleimide (Scheme 4). While our data supports this simple model, the involvement of unknown host factors cannot be ruled out.

Based on our *in vivo* analyses, no function can yet be assigned to MarE, a predicted VioE homolog. Gel analysis of the soluble protein extract of *marE* harboring *E. coli* confirms that soluble MarE is produced in this system (Figure S4). In violacein biosynthesis, VioE is predicted to produce prodeoxyviolacein (**5**) by "interacting" with a transient intermediate of unknown structure that is produced by the violacein CPA synthase.^[18] VioE is a small protein (191 aa) that lacks any functionally characterized homologs, any known catalytic residues or any recognized cofactor-binding motifs. Accordingly, the mechanistic details of its role in violacein biosynthesis remain unclear. MarE shares high sequence identity to VioE, however *in vivo* MarE cannot complement the function of VioE in the production of prodeoxyviolacein (**5**) (Figure 2, VioA+MarBE versus VioAE+MarB). Whether MarE is inactive in the *mar* cluster, plays a role that is redundant in *E. coli*, or functions in the biosynthesis of an as yet unidentified metabolite is subject for further investigation.

A diverse collection of bisindolylmaleimides have been chemically synthesized and tested for bioactivity, with a particular focus on kinase inhibitory activity.^[19] Arcyrarubin A (**2**) is a submicromolar inhibitor of protein kinase C (IC₅₀: 0.55 μ M).^[20] Interestingly, the addition of the *N*-methyl to give methylarcyriarubin (**1**) abolishes protein kinase C inhibitory activity (IC₅₀: > 100 μ M),^[20] but leads to activity against mitogen-stimulated protein kinase p70^{s6k}/p85^{s6k} (IC₅₀: 8 μ M).^[21] A number of synthetic bisindolylmaleimide derivatives, including ruboxistaurin, enzastaurin, and MKC-1, have undergone or are currently in clinical trials as potent and specific kinase inhibitors for use as cancer and diabetes therapies.^[7]

Indolocarbazole tryptophan dimers, which differ from bisindolylmaleimides by the coupling of the C-5 and C-5' indole carbons, have also been extensively explored as kinase

inhibitors.^[22] The additional C-C coupling forces indolocarbazoles to bind in a planar conformation, while the more flexible bisindolylmaleimides have been observed to bind in a nonplanar conformation.^[23] This conformational flexibility is believed to be responsible for trends observed in tryptophan bioactivities. Bisindolylmaleimides tend to be less potent, but more specific kinase inhibitors compared to their indolocarbazole counterparts.^[20, 24] Known indolocarbazole biosynthetic enzymes have been used, both *in vitro* in chemoenzymatic synthesis^[25] and *in vivo* in combinatorial biosynthesis,^[26] to generate many un-natural indolocarbazole analogs. Considering the significant clinical relevance of the bisindolylmaleimides, the identification of the *mar* biosynthetic gene cluster should facilitate the generation of additional collections of potentially pharmaceutically relevant tryptophan dimers using biosynthetic approaches.

The dimerization of tryptophan by the tandem action of an IPA imine synthase and a CPA synthase generates a tremendously versatile intermediate (CPA) that appears to serve as a substrate for the biosynthesis of a diverse array of dimer substructures.^[27] The characterization of the *mar* cluster adds a new branch to the meta-biosynthetic scheme of bacterial tryptophan dimers (Scheme 5). In this case, the Rieske type dioxygenase, MarC, is predicted to oxidatively decarboxylate CPA to produce the bisindolylmaleimide core. As only a small subset of known tryptophan dimer natural products have had their biosynthetic gene clusters characterized thus far, it is likely that the current global biosynthetic scheme is still incomplete and that there are more divergent tryptophan dimer pathways awaiting discovery.

Experimental Section

Soil environmental DNA (eDNA) library construction

An eDNA library containing $>1.5 \times 10^7$ unique clones was constructed from a Chihuahuan Desert (New Mexico; NM library) soil sample as described previously.^[28] In brief, the soil was sifted to remove large particulates and heated for 2 hr at 70 °C in lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 1.5 M NaCl, 1% (w/v) CTAB, 2% (w/v) SDS, pH 8.0). The soil particulates were removed by centrifugation (30 min, 4000 × g, 4 °C). Crude eDNA was precipitated by the addition of isopropanol (0.7 vol) and collected by centrifugation (30 min, 4000 × g, 4 °C), followed by a wash step with 70% ethanol (10 min, 4000 × g, 4 °C). The resulting eDNA was resuspended in TE buffer and subjected to agarose gel electrophoresis (1% agarose gel, 16 hr, 20 V). Electroelution (2 hr, 100V) of the high molecular weight (HMW; >25 kb) compression band yielded purified eDNA. HMW eDNA was concentrated (100 KDa molecular weight cut off), blunt-ended (End-It), ligated into cosmid vector (pWEB-TNC or pWEB436), packaged into λ phage (MaxPlax) and transfected into *E. coli* (EC100, Epicentre). The transfected cells were inoculated and arrayed into 48-well plates such that each well consisted of $4-5 \times 10^3$ clones in LB medium (5 mL) with antibiotic (100 µg/mL ampicillin or 50 µg/mL apramycin) and grown overnight. Matching DNA miniprep and glycerol stock pairs were made for each well and arrayed such that sets of 8 wells were combined to generate unique library “row pools.”

Screening and recovery of *mar* cluster harboring clone NM343

The *mar* cluster was found on a cosmid (NM434) recovered from a screen of the New Mexico eDNA library.^[12a] In this screen, degenerate PCR primers StaDV-F and StaDV-R (Table 1) were designed to recognize conserved regions in known bacterial chromopyrrolic acid (CPA) synthase genes (accession no.: *vioB* AF172851.1, *staD* AB088119.1, *rebD* AJ414559.1, *inkD* DQ399653.1, *atmD* DQ297453.1.). These primers were then used to screen the New Mexico eDNA library “row pools” for CPA genes. PCR reaction mixtures were as follows: FailSafe PCR Buffer G (10 μ L; Epicentre), StaDV-F and StaDV-R primers (0.5 μ L each; final concentration of 2.5 μ M each), template “row pool” eDNA (0.5 μ L; 100 ng), and *Taq* DNA polymerase (0.2 μ L; New England Biolabs). PCR cycling conditions: 1 cycle of 95°C for 5 min; 7 cycles of 95°C for 30 sec, 65°C for 30 sec with 1°C decrement per cycle to 59°C, 72°C for 40 sec; 30 cycles of 95°C for 30 sec, 58°C for 30 sec, 72°C for 40 sec; 1 cycle of 72°C for 7 min; 4°C hold. Amplicons of the correct predicted size (~561 base pairs) were gel-purified, re-amplified and sequenced using the StaDV-F and StaDV-R primers. Those that were confirmed to be CPA synthase gene sequences based on BLASTX homology search (NCBI) were then used to guide the recovery of the eDNA clone harboring each of the unique CPS sequences identified in the arrayed library. The *mar* cluster containing clone was recovered from well number 343 of the New Mexico library (NM343) using a serial dilution approach with primers NM343-F and NM343-R (Table 1).^[12a] Clone NM343 was then *de novo* sequenced at the Sloan Kettering Institute DNA Sequencing Core Facility using 454 pyrosequencing technology (Roche). The sequence data was assembled using GS De Novo Assembler (Roche), annotated using FGENESB (SoftBerry) for gene prediction and BLASTP (NCBI) for protein homology searches and deposited in the GenBank database under the accession number KF551863.

Cloning of genes from the *mar* and violacein (*vio*) gene clusters

Individual genes were amplified from the *mar* and *vio* clusters using clones NM343 and CSL51^[29] as templates, respectively, using *Phusion Hot Start Flex* DNA polymerase kit (New England Biolabs). Primers are shown in Table 1. PCR cycling conditions: 1 cycle of 95°C for 5 min; 30 cycles of 95°C for 10 sec, 62°C for 30 sec, 72°C for 30 sec/kb; 1 cycle of 72°C for 7 min; 4°C hold. Gel purified amplicons were restriction digested and cloned into the following Duet (Novagen) vectors: MarB NcoI/SalI sites of pCOLADuet-1; MarC NcoI/HindIII sites of pETDuet-1; MarE NdeI/KpnI sites of pETDuet-1; MarM NcoI/SalI sites of pCDFDuet-1; VioA NdeI/MfeI sites of pCOLADuet-1; VioB NcoI/HindIII sites of pCOLADuet-1; VioE NdeI/KpnI sites of pETDuet-1.

Heterologous expression of *mar* and *vio* biosynthetic genes

For expression studies, electrocompetent *E. coli* BL21 cells were transformed with Duet vectors harboring various combinations of *mar* and *vio* biosynthetic genes and grown in LB medium with required antibiotic combination for selection (30 μ g/mL kanamycin, 100 μ g/mL ampicillin, 100 μ g/mL spectinomycin). Gene expression was induced in cultures grown to an OD₆₀₀ of 0.5 with the addition of IPTG (final concentration of 0.1 mM). Thirty-six hours (200 rpm, 25 °C) post induction, the cultures were extracted with ethyl acetate that was acidified to pH ~3-4 with the addition of hydrochloric acid. Extracts dried *in vacuo*

were dissolved in methanol and subjected to reversed phase LC/MS analysis (150 × 4.6 mm, 5 μm × Bridge C18: linear gradient of 80:20 water:methanol to 0:100 water:methanol with 0.1% formic acid). Commercially available methylarcyriarubin (i.e. Bisindolylmaleimide V, Santa Cruz Biotechnology), arcyriarubin A (i.e. Bisindolylmaleimide IV, Santa Cruz Biotechnology), and indole-3-pyruvic acid (Sigma-Aldrich) were run on the LC/MS, under the same conditions. Because chromopyrrolic acid and prodeoxyviolacein were not commercially available, these standards were prepared from the heterologous expression of well-defined biosynthetic genes from the violacein pathway. Chromopyrrolic acid and prodexoyviolacein were produced from VioAB and VioABE expressing *E. coli* cultures, respectively. They were each purified from culture broth extracts by HPLC using conditions based on previous violacein pathway studies.^[18] Analytical LC/MS data was acquired using Micromass ZQ mass spectrometer (Waters).

Large-scale production and isolation of methylarcyriarubin (1) from *E. coli* cultures expressing *mar* genes

Cultures of VioA + MarBCEM expressing *E. coli* BL21 cells (2 L) grown for 36 hr (200 rpm, 25 °C) after IPTG induction was extracted with ethyl acetate (4 L). This extract was initially fractionated by silica gel RediSep flash chromatography (RediSepRf 12 gram silica flash column: 3 min 100% chloroform, 27 min linear gradient from 100% chloroform to 85:15 chloroform:methanol). Compound **1** eluted with 99:1 chloroform:methanol. Compound **1** was then purified (1.6 mg/L) from the 99:1 fraction by 65:35 water:acetonitrile isocratic reversed phase HPLC (150 × 10 mm, 5 μm × Bridge C18). LTQ-Orbitrap mass spectrometer (Thermo Scientific) and 600-MHz spectrometer (Bruker) were used to acquire HRMS and NMR data, respectively, for structure elucidation studies.

MarE protein expression analysis

A liquid culture (100 mL) of *E. coli* harboring *vioE*/pETDuet-1 was grown to an OD₆₀₀ of 0.5. The culture was subsequently split into two, with one uninduced and the other induced with the addition of IPTG (final concentration of 0.1 mM). Two hours (200 rpm, 37 °C) post induction, an aliquot (1 mL) was removed from each culture and the cells were pelleted by centrifugation (1 min, 13,000 × g). The cells were resuspended in native purification buffer (200 μL; 0.5 M NaCl, 50 mM NaH₂PO₄, pH 8.0) and lysed by sonication (30 cycles of 1 sec pulse on and 2 sec pulse off, 45% amplitude; Sonic Dismembrator, Fisher Scientific). The cell debris and the insoluble proteins were collected by centrifugation (15 min, 13,000 × g, 4 °C). Aliquots (20 μL) of the supernatant were mixed with SDS loading buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA, 0.02% bromophenol blue), heated (10 min, 95 °C) and run on a polyacrylamide gel (4-20% Mini-PROTEAN TGX Gel with Precision Plus Protein Dual Color Standards, Bio-Rad). Gels were stained with coomassie (Coomassie Brilliant Blue R-250 Staining Solution, Bio-Rad) and imaged using Gel Doc X+ System (Bio-Rad).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH GM-77516. S.F.B. is an HHMI Early Career Scientist.

References

- [1]. Gill M, Steglich W. *Fortschr Chem Org Naturst*. 1987; 51:1–317. [PubMed: 3315906]
- [2]. Steglich W, Steffan B, Kopanski L, Eckhardt G. *Angew Chem Int Ed Engl*. 1980; 19:459–460.
- [3] a). Zhang J, Yang PL, Gray NS. *Nat Rev Cancer*. 2009; 9:28–39. [PubMed: 19104514] b) Driggers EM, Hale SP, Lee J, Terrett NK. *Nat Rev Drug Discov*. 2008; 7:608–624. [PubMed: 18591981] c) Pajak B, Orzechowska S, Gajkowska B, Orzechowski A. *Adv Med Sci*. 2008; 53:21–31. [PubMed: 18635421]
- [4]. Anderson PW, McGill JB, Tuttle KR. *Curr Opin Nephrol Hypertens*. 2007; 16:397–402. [PubMed: 17693752]
- [5]. Katare RG, Zhitian Z, Sodeoka M, Sasaguri S. *Can J Physiol Pharmacol*. 2007; 85:979–985. [PubMed: 18066098]
- [6]. Asakai R, Aoyama Y, Fujimoto T. *Neurosci Res*. 2002; 44:297–304. [PubMed: 12413658]
- [7] a). Aiello LP, Davis MD, Girach A, Kles KA, Milton RC, Sheetz MJ, Vignati L, Zhi XE. *Ophthalmology*. 2006; 113:2221–2230. [PubMed: 16989901] b) Ma S, Rosen ST. *Curr Opin Oncol*. 2007; 19:590–595. [PubMed: 17906457] c) Zagouri F, Sergeantanis TN, Chrysikos D, Filipits M, Bartsch R. *Gynecol Oncol*. 2012; 127:662–672. [PubMed: 22967800]
- [8]. Li T, Du Y, Cui Q, Zhang J, Zhu W, Hong K, Li W. *Mar Drugs*. 2013; 11:466–488. [PubMed: 23389092]
- [9]. Torsvik V, Goksoyr J, Daae FL. *Appl Environ Microbiol*. 1990; 56:782–787. [PubMed: 2317046]
- [10] a). Pemberton JM, Vincent KM, Penfold RJ. *Current Microbiology*. 1991; 22:355–358. b) Onaka H, Taniguchi S, Igarashi Y, Furumai T. *Journal of Antibiotics*. 2002; 55:1063–1071. [PubMed: 12617516] c) Sanchez C, Butovich IA, Brana AF, Rohr J, Mendez C, Salas JA. *Chem Biol*. 2002; 9:519–531. [PubMed: 11983340] d) Chiu HT, Chen YL, Chen CY, Jin C, Lee MN, Lin YC. *Mol Biosyst*. 2009; 5:1180–1191. [PubMed: 19756308] e) Gao Q, Zhang C, Blanchard S, Thorson JS. *Chem Biol*. 2006; 13:733–743. [PubMed: 16873021]
- [11] a). Howard-Jones AR, Walsh CT. *J Am Chem Soc*. 2006; 128:12289–12298. [PubMed: 16967980] b) Balibar CJ, Walsh CT. *Biochemistry*. 2006; 45:15444–15457. [PubMed: 17176066] c) Asamizu S, Hirano S, Onaka H, Koshino H, Shiro Y, Nagano S. *Chembiochem*. 2012; 13:2495–2500. [PubMed: 23081937]
- [12] a). Chang FY, Ternei MA, Calle PY, Brady SF. *J Am Chem Soc*. 2013b) Chang FY, Brady SF. *Proc Natl Acad Sci U S A*. 2013; 110:2478–2483. [PubMed: 23302687] c) Chang FY, Brady SF. *J Am Chem Soc*. 2011; 133:9996–9999. [PubMed: 21542592]
- [13]. Salas JA, Mendez C. *Trends Microbiol*. 2007; 15:219–232. [PubMed: 17412593]
- [14]. Howard-Jones AR, Walsh CT. *Biochemistry*. 2005; 44:15652–15663. [PubMed: 16313168]
- [15] a). Howard-Jones AR, Walsh CT. *J Am Chem Soc*. 2007; 129:11016–11017. [PubMed: 17705392] b) Groom K, Bhattacharya A, Zechel DL. *Chembiochem*. 2011; 12:396–400. [PubMed: 21290541] c) Asamizu S, Shiro Y, Igarashi Y, Nagano S, Onaka H. *Biosci Biotechnol Biochem*. 2011; 75:2184–2193. [PubMed: 22056432]
- [16] a). Ryan KS, Howard-Jones AR, Hamill MJ, Elliott SJ, Walsh CT, Drennan CL. *Proc Natl Acad Sci U S A*. 2007; 104:15311–15316. [PubMed: 17873060] b) Goldman PJ, Ryan KS, Hamill MJ, Howard-Jones AR, Walsh CT, Elliott SJ, Drennan CL. *Chem Biol*. 2012; 19:855–865. [PubMed: 22840773] c) Ballou DP. *Proc Natl Acad Sci U S A*. 2007; 104:15587–15588. [PubMed: 17898164] d) Tsai SC. *Chem Biol*. 2012; 19:787–788. [PubMed: 22840764]
- [17]. Barry SM, Challis GL. *ACS Catal*. 2013; 3
- [18] a). Sanchez C, Brana AF, Mendez C, Salas JA. *Chembiochem*. 2006; 7:1231–1240. [PubMed: 16874749] b) Hirano S, Asamizu S, Onaka H, Shiro Y, Nagano S. *J Biol Chem*. 2008; 283:6459–6466. [PubMed: 18171677] c) Ryan KS, Balibar CJ, Turo KE, Walsh CT, Drennan CL. *J Biol Chem*. 2008; 283:6467–6475. [PubMed: 18171675]

- [19] a). Teicher BA. Clin Cancer Res. 2006; 12:5336–5345. [PubMed: 17000666] b) Brehmer D, Godl K, Zech B, Wissing J, Daub H. Mol Cell Proteomics. 2004; 3:490–500. [PubMed: 14769951]
- [20]. Davis PD, Hill CH, Lawton G, Nixon JS, Wilkinson SE, Hurst SA, Keech E, Turner SE. J Med Chem. 1992; 35:177–184. [PubMed: 1732526]
- [21]. Marmy-Conus N, Hannan KM, Pearson RB. FEBS Lett. 2002; 519:135–140. [PubMed: 12023032]
- [22]. Nakano H, Omura S. J Antibiot (Tokyo). 2009; 62:17–26. [PubMed: 19132059]
- [23] a). Komander D, Kular GS, Schuttelkopf AW, Deak M, Prakash KR, Bain J, Elliott M, Garrido-Franco M, Kozikowski AP, Alessi DR, van Aalten DM. Structure. 2004; 12:215–226. [PubMed: 14962382] b) Gassel M, Breitenlechner CB, Konig N, Huber R, Engh RA, Bossemeyer D. J Biol Chem. 2004; 279:23679–23690. [PubMed: 14996846]
- [24]. Toullec D, Pianetti P, Coste H, Bellevergue P, Grandperret T, Ajakane M, Baudet V, Boissin P, Boursier E, Loriolle F, Duhamel L, Charon D, Kirilovsky J. J Biol Chem. 1991; 266:15771–15781. [PubMed: 1874734]
- [25] a). Zhang CS, Weller RL, Thorson JS, Rajski SR. J Am Chem Soc. 2006; 128:2760–2761. [PubMed: 16506729] b) Zhang C, Albermann C, Fu X, Peters NR, Chisholm JD, Zhang G, Gilbert EJ, Wang PG, Van Vranken DL, Thorson JS. Chembiochem. 2006; 7:795–804. [PubMed: 16575939]
- [26] a). Sanchez C, Zhu L, Brana AF, Salas AP, Rohr J, Mendez C, Salas JA. Proc Natl Acad Sci U S A. 2005; 102:461–466. [PubMed: 15625109] b) Salas AP, Zhu L, Sanchez C, Brana AF, Rohr J, Mendez C, Salas JA. Mol Microbiol. 2005; 58:17–27. [PubMed: 16164546]
- [27]. Speck K, Magauer T. Beilstein J Org Chem. 2013; 9:2048–2078. [PubMed: 24204418]
- [28]. Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF. Proc Natl Acad Sci U S A. 2013; 110:11797–11802. [PubMed: 23824289]
- [29]. Brady SF, Chao CJ, Handelsman J, Clardy J. Org Lett. 2001; 3:1981–1984. [PubMed: 11418029]
- [30] a). Brenner M, Rexhausen H, Steffan B, Steglich W. Tetrahedron. 1988; 44:2887–2892. b) Wang K, Liu Z. Synthetic Communications. 2009; 40:144–150.

Figure 1.
eDNA-derived *mar* biosynthetic gene cluster.

Figure 2.

HPLC-UV traces of culture broth extracts from *mar* gene cluster expression studies in *E. coli*. IPA imine is reported to undergo spontaneous deamination to form IPA (**3**).^[11c, 14] The peak marked with an asterisk (*) is an uncharacterized by-product of the co-expression of an IPA imine synthase with a CPA synthase in *E. coli*.

Scheme 1.

Bisindolylmaleimide compounds, with the 3,4-di-1H-indol-3-yl-1H-pyrrole-2,5-dione core structure colored in red. A) Synthetic derivatives that have entered into clinical trials for drug therapy. B) Compounds that are encoded by the eDNA-derived *mar* biosynthetic gene cluster.

Scheme 2.

Biosynthesis of bacterial tryptophan dimers, namely violacein and indolocarbazole compounds like rebeccamycin and staurosporine, sharing two common initial steps and diverging downstream to yield various chemical structures.

Scheme 3.

Biosynthesis of methylarcyriarubin. VioA-like IPA imine synthase marked with an asterisk (*) is predicted to be encoded in the endogenous host's genome.

Scheme 4.

Comparison of the proposed enzymatic oxidative mechanism between bisindolylmaleimide and indolocarbazole in the biosynthesis of a maleimide moiety.

Scheme 5.

Bacterial tryptophan dimer biosynthetic pathways that diverge from a common tryptophan dimer intermediate (highlighted in green). Bisindolylmaleimide pathway was found to also branch out via oxidation by MarC (highlighted in blue).

Table 1PCR primer list.^[a]

Name	Sequence
StaDV-F	GTSATGMTSCAGTACCTSTACGC
StaDV-R	YTCVAGCTGRTAGYCSGGRTG
NM343-F	GAGCAGCTCAAGCTGGTGTG
NM343-R	AAGGCCTCGGAATCTGCTG
MarB-F	GAG <u>Accatggca</u> ATGAGCATCCTGGAATTTCCGC
MarB-R	GAG <u>Agtcgac</u> CCTCACAAGAGTGGAACGG
MarC-F	GAG <u>Atcatgac</u> ATGCTGAGCGCCGAAGACA
MarC-R	GAG <u>Aaagctt</u> CTCATGCGGTCTCCTTGC
MarE-F	GAG <u>AcatATG</u> AGCGCCGCCCGC
MarE-R	GAG <u>Aggtacc</u> GAGGATTGTTGGTCTGCTGAC
MarM-F	GAG <u>Aacatgtct</u> ATGACAACTCAGGGAACGCC
MarM-R	GAG <u>Agtcgac</u> GCTCAGCGTTCTTTCTGTGC
VioA-F	GAG <u>AcatATG</u> ACAAACTATTCCGACATTTCG
VioA-R	GAG <u>Acaattg</u> GGAAATCCAGAATGCTCATGC
VioB-F	GAG <u>Accatggca</u> ATGAGCATCTTGATTTCCCC
VioB-R	GAG <u>AaagcttTGC</u> ATATCAAGCCTCTCTAGAC
VioE-F	GCGC <u>catATG</u> CCGATGCCTGTCCAC
VioE-R	GCGC <u>ggtacc</u> CACAAACGGAACAGGACTCAGT

^[a] Underlined sequences indicate the restriction sites added for cloning purposes.