JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Cloning and Characterization of an Environmental DNA-Derived Gene Cluster That Encodes the Biosynthesis of the Antitumor Substance BE-54017

Fang-Yuan Chang and Sean F. Brady*

Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, United States

Supporting Information

ABSTRACT: Soil is predicted to contain thousands of unique bacterial species per gram. Soil DNA libraries represent large reservoirs of biosynthetic diversity from which diverse secondary metabolite gene clusters can be recovered and studied. The screening of an archived soil DNA library using primers designed to target oxytryptophan dimerization genes allowed us to identify and functionally characterize the first indolotryptoline biosynthetic gene cluster. The recovery and heterologous expression of an environmental DNA-derived gene cluster encoding the biosynthesis of the antitumor substance BE-54017 is reported here. Transposon mutagenesis identified two monooxygenases, AbeX1 and AbeX2, as being responsible for the transformation of an indolocarbazole precursor into the indolotryptoline core of BE-54017.

A number of structurally diverse and pharmacologically interesting natural products are thought to arise from the dimerization of tryptophan (Figure 1A,B).[1,2] Biosynthetic gene clusters for violacein[3] as well as representative bisindolylquinones (e.g., terrequinone A,[4]Figure 1A) and indolocarbazole-based metabolites[5−8] (Figure 1B) have been cloned and extensively characterized. However, most members of this diverse class of metabolites have not been investigated biosynthetically. Soil contains an extraordinarily diverse collection of both cultured and as yet uncultured bacteria.[9,10] DNA isolated from natural populations of soil bacteria [environmental DNA (eDNA)] should therefore represent a reservoir from which additional tryptophan dimer biosynthetic gene clusters can be cloned and characterized. Here we report the cloning and heterologous expression of an eDNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017 (**1**) (Figure 2).[11] BE-54017 is part of a small family of indolotryptoline (indolotetrahydro-β-carboline)-based natural products that show potent activity against tumor cell lines.[1,11] Transposon mutagenesis of the antitumor substance BE-54017 (*abe*) gene cluster revealed that the indolotryptoline core of BE-54017 arises from two successive oxidations of an indolocarbazole precursor.

To recover tryptophan dimer biosynthetic gene clusters from the environment, we screened a previously archived soil eDNA cosmid library by polymerase chain reaction (PCR) using degenerate primers designed to recognize conserved regions in known

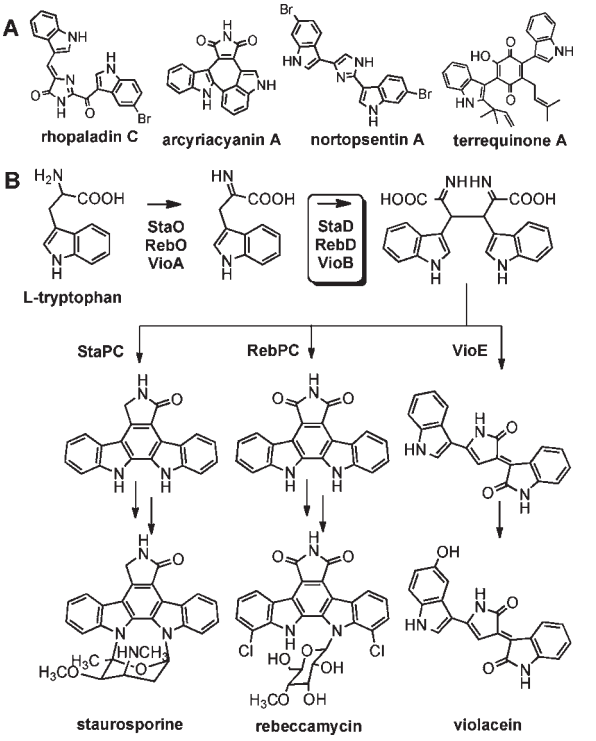

Figure 1. (A) Structurally diverse natural products predicted to arise from the dimerization of tryptophan. (B) Overview of indolocarbazole (staurosporine/rebeccamycin) and violacein biosynthesis. In these pathways, homologous enzymes (StaD/RebD/VioB) are responsible for the dimerization of oxytryptophan.

oxytryptophan dimerization genes (StaD, RebD, VioB, etc).[12,13] Oxytryptophan dimerization enzymes were chosen as our targets because this enzyme family is used in the biosynthesis of structurally diverse tryptophan dimers. Both indolocarbazole (e.g., staurosporine, rebeccamycin, K-252a, and AT2433) biosynthetic gene clusters and violacein biosynthetic gene clusters contain homologous enzymes that carry out the oxidation (StaO/RebO/VioA) and subsequent dimerization (StaD/RebD/VioB) of tryptophan (Figure 1B). Cosmid clones containing unique eDNA-derived StaD-like homologues were recovered from the eDNA library using whole-cell PCR on serially diluted library

Received: March 16, 2011
Published: May 04, 2011

dx.doi.org/10.1021/ja2022653 | J. Am. Chem. Soc. 2011, 133, 9996–9999

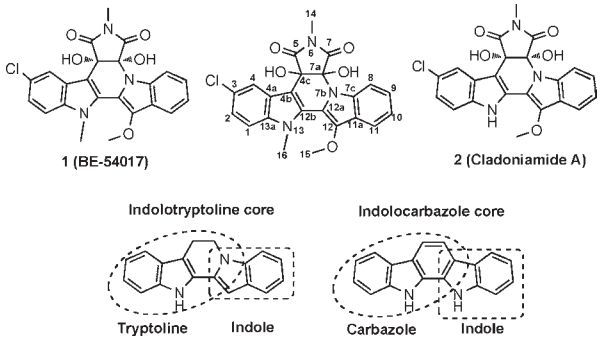

Figure 2. The eDNA-derived *abe* gene cluster encodes the biosynthesis of BE-54017 (**1**). BE-54017 and the cladoniamides share an indolotryptoline core. Here we show that indolotryptolines arise from indolocarbazole precursors.

aliquots. Each recovered cosmid was then de novo sequenced and annotated.

One eDNA-derived cosmid, AB1650, was found to contain a complete set of conserved indolocarbazole biosynthetic genes (*abeO, D, C, P*) as well as two monooxygenases (*abeX1, X2*), three methyltransferases (*abeM1, M2, M3*), and a halogenase (*abeH*) (Figure 3). The presence of the two predicted monooxygenases was unprecedented in known indolocarbazole biosynthetic pathways, suggesting that this gene cluster likely encodes the biosynthesis of an oxidized or rearranged indolocarbazole-based metabolite. To investigate this possibility, cosmid AB1650 was retrofitted with the genetic elements required for conjugation, selection, and stable integration into *Streptomyces* and then conjugated into *Streptomyces albus* for heterologous expression studies.

Cultures of *S. albus* transformed with AB1650 produced one major clone-specific metabolite as well as four minor clone-specific metabolites. The ^{1}H and ^{13}C NMR, UV, and high-resolution electrospray ionization mass spectrometry (HR-ESI-MS)[14] data for the major metabolite (**1**) were identical to data reported for the antitumor substance BE-54017 (Figure 2).[11] Minor metabolites **3** and **4** each differ from **1** by the loss of 28 mass units, corresponding to "CO".[14] Recently reported derivatives of the structurally related metabolite cladoniamide A (**2**), in which the *N*-methylsuccinimide substructure has been hydrolyzed, show a similar loss of "CO".[15] As would be expected for correspondingly hydrolyzed BE-54017 derivatives, compounds **3** and **4** both contain a new correlation spectroscopy (COSY) spin system between an amide nitrogen and a methyl singlet that now shows a heteronuclear multiple bond coherence (HMBC) correlation to only one of the two carbonyl carbons in the structure. HMBC correlations from the C4c hydroxyl proton to C7a, C5, C4c, and C4b in compound **3** and from the C7a hydroxyl proton to C7, C7a, and C4c in compound **4** allowed us to define the position of the *N*-methylamide in both structures. On the basis of comparisons of HR-ESI-MS[14] and 2D NMR data, compounds **5** and **6** were determined to be the deschloro analogues of compounds **1** and **4**, respectively. Compounds **3**−**6** are novel secondary metabolites (Figure 4).

It has been suggested that the indolotryptoline core seen in BE-54017 and the cladoniamides might arise either from the oxidation of an indolocarbazole precursor or from an indolocarbazole-independent pathway.[1,15] To elucidate the origin of the indolotryptoline substructure and assign specific functions to the

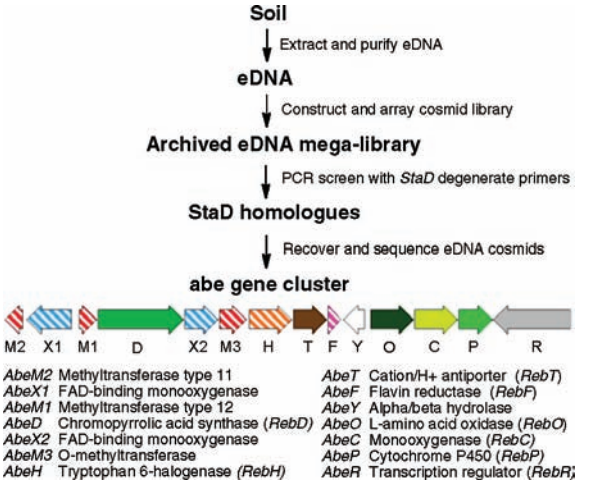

Figure 3. The BE-54017 biosynthetic gene cluster (*abe*) (GenBank accession code JF439215) was recovered from an eDNA library using primers that recognize StaD-like oxytryptophan dimerization genes.

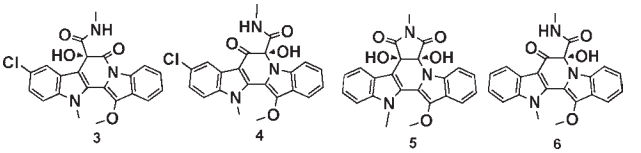

Figure 4. Minor compounds isolated from the culture broth of *S. albus* transformed with the *abe* gene cluster.

individual genes found in the *abe* gene cluster, we carried out a transposon mutagenesis study on cosmid AB1650. Individual transposon mutants were sequenced to identify clones with insertions in key biosynthetic genes (Figure 5). This collection of transposon mutants was then conjugated back into *S. albus*, and the major clone-specific metabolites found in the culture broth extracts of each mutant were structurally characterized.

Consistent with what is known about the biosynthesis of indolocarbazoles (Figure 1B), transposon insertions in predicted indolocarbazole biosynthetic genes resulted in either the absence of organic-extractable clone-specific metabolites (*abeO, D*) or the production of the known indolocarbazole intermediate 3-chloromonopyrrolic acid (*abeC, P*) (Figure 5). Homologues of the two predicted oxidoreductases, *abeX1* and *abeX2*, do not appear in any known bisindole biosynthetic gene clusters. Disruption of *abeX2* results in the accumulation of compound **8**,[16] and disruption of *abeX1* leads to the accumulation of the simple indolocarbazole 3-chloroarcyriaflavin (**7**).[17] The isolation of compound **7** confirmed the indolocarbazole origin of BE-54017.

Compound **8** contains the C4c/C7a diol seen in BE-54017 but has not undergone a rearrangement of the indolocarbazole core, while compound **7** contains neither the diol nor the flipped indole. AbeX1 and AbeX2 show homology to class-A flavoprotein monooxygenases, and therefore the oxidation reactions they carry out are predicted to proceed through epoxides (Figure 6).[18] The mutagenesis results coupled with homology arguments allowed us to construct a biosynthetic scheme for the formation of the indolotryptoline core in BE-54017 from indolocarbazole **7** (Figure 6). Our scheme is consistent with the proposed biogenesis of the cladoniamides outlined by Williams et al.[15] In this scheme, the diol is introduced by AbeX1 through the epoxidation

dx.doi.org/10.1021/ja2022653 |*J. Am. Chem. Soc.* 2011, 133, 9996–9999

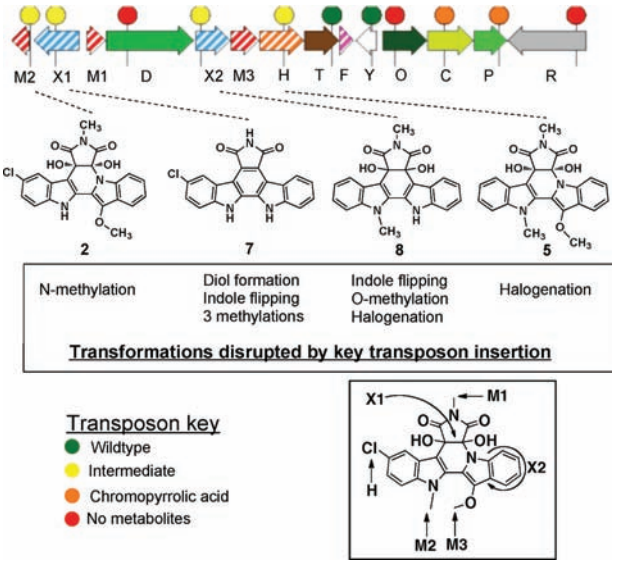

Figure 5. Major metabolites produced by select transposon mutants. Sites on BE-54017 where key biosynthetic enzymes are predicted to act are indicated in the inset.

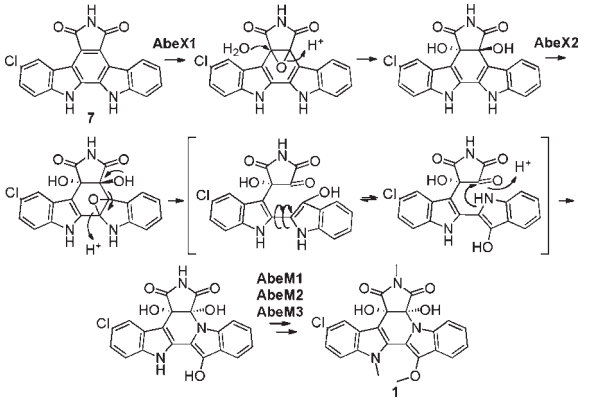

Figure 6. Two monooxygenases, AbeX1 and AbeX2, are responsible for the conversion of an indolocarbazole precursor into the indolotryptoline core of BE-54017. BE-54017 requires three methylations; the exact timing of these transformations is not known.

of the C4c/C7a double bond. AbeX2 is then responsible for promoting the rearrangement of the indole via the introduction of a second epoxide at C7b/C12a. The opening of this epoxide is accompanied by fragmentation of the C7a/C7b bond followed by rotation of the indole around C12a/C12b and finally the formation of the C7a/N7b bond.

Whether the proposed epoxide hydrolysis and indole rearrangement reactions occur spontaneously, are catalyzed by the monooxygenases themselves, or are catalyzed by other enzymes is not clear. While AbeY shares the same α/β-hydrolase fold superfamily as most epoxide hydrolases,[19] our studies suggest that it is dispensable in the biosynthesis of compounds **1** and **3**−**6**, as all five compounds are produced by the *abeY* knockout mutant. The appearance of small quantities of low-molecular-weight clone-specific metabolites in extracts from cultures of the *abeY* mutant suggests that while AbeY is not required, it may be involved in enhancing the efficiency of one or more biosynthetic transformations, similar to the reported function of StaC in staurosporine biosynthesis.[20] Alternatively, the role of AbeY may be complemented by the host's endogenous biosynthetic machinery.

A detailed accounting of the functionality that appears on the compounds produced in our transposon mutagenesis experiments allowed us to assign functions to the three methyltransferases and the predicted halogenase found in the *abe* gene cluster (Figure 5). Disruption of *abeH*, a homologue of the tryptophan halogenase found in the rebeccamycin pathway, led to the accumulation of deschloro derivative **5**, thereby confirming its role as a halogenase. The absence of the chloride substituent on compound **8** suggested that the transposon insertion in *abeX2* disrupts the expression of downstream genes in the same operon. *AbeM3* is positioned between *abeX2* and *abeH* and is therefore predicted to be transcriptionally silenced in this transposon mutant. Since both the N6 and N13 methylations appear on **8**, *abeM3* is predicted to be responsible for the methylation of the C12 hydroxyl in BE-54017. The regiospecificity of the two remaining *N*-methyltransferases, *abeM1* and *abeM2*, was inferred from the accumulation of the O12,N6-dimethylated derivative **2** (cladoniamide A) in the *abeM2* transposon mutant. AbeM1, AbeM2, and AbeM3 are therefore N6-, N13-, and O12-specific methyltransferases, respectively (Figure 5 inset).

No gene in the *abe* cluster could be linked to the hydrolysis of the *N*-methylsuccinimide, suggesting that this reaction is either carried out by the host or occurs spontaneously during the fermentation process.

Compounds **1**−**8** were assayed for antiproliferative activity against human colon cancer HCT116 cells. While **3**−**8** were not active below 8 μg/mL, **1** and **2** exhibited potent antiproliferative activities [IC_{50} (μg/mL): 0.079 for **1**, 0.0088 for **2**]. Crystallographic studies have shown that the planar structure of indolocarbazoles is important for topoisomerase/DNA (rebeccamycin) and kinase (staurosporine) binding.[21,22] The C4c/C7a diol seen in indolotryptolines causes the *N*-methylsuccinimide to bend out of the bisindole plane, which may afford these compounds the ability to bind different cellular targets for their antitumor activity. Considerable progress in generating novel indolocarbazole analogues by combinatorial biosynthesis has been made, and the identification of AbeX1 and AbeX2 provides new tools for producing structurally and functionally diverse bisindole metabolites.[23]

The cloning of DNA extracted directly from environmental samples provides a means of exploring the biosynthetic capacity of thousands of bacterial genomes simultaneously. Archived eDNA libraries represent permanent resources from which diverse gene clusters can be recovered and studied. In this study, the screening of an archived eDNA library using primers designed to recognize oxytryptophan dimerization genes allowed us to clone and characterize the first indolotryptoline biosynthetic gene cluster.

■ ASSOCIATED CONTENT

Supporting Information. Experimental details, annotation table for the *abe* gene cluster, tables of NMR assignments, and 1D and 2D NMR spectra for compounds **1**−**8**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
sbrady@rockefeller.edu

■ ACKNOWLEDGMENT

This work was supported by NIH Grant GM077516. S.F.B. is a Howard Hughes Medical Institute Early Career Scientist.

■ REFERENCES

(1) Ryan, K. S.; Drennan, C. L. *Chem. Biol.* **2009**, *16*, 351.
(2) Sanchez, C.; Mendez, C.; Salas, J. A. *Nat. Prod. Rep.* **2006**, *23*, 1007.
(3) August, P. R.; Grossman, T. H.; Minor, C.; Draper, M. P.; MacNeil, I. A.; Pemberton, J. M.; Call, K. M.; Holt, D.; Osburne, M. S. *J Mol. Microbiol. Biotechnol.* **2000**, *2*, 513.
(4) Balibar, C. J.; Howard-Jones, A. R.; Walsh, C. T. *Nat. Chem. Biol.* **2007**, *3*, 584.
(5) Onaka, H.; Taniguchi, S.; Igarashi, Y.; Furumai, T. *J. Antibiot. (Tokyo)* **2002**, *55*, 1063.
(6) Sanchez, C.; Butovich, I. A.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. *Chem. Biol.* **2002**, *9*, 519.
(7) Gao, Q.; Zhang, C.; Blanchard, S.; Thorson, J. S. *Chem. Biol.* **2006**, *13*, 733.
(8) Kim, S. Y.; Park, J. S.; Chae, C. S.; Hyun, C. G.; Choi, B. W.; Shin, J.; Oh, K. B. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 1119.
(9) Rappe, M. S.; Giovannoni, S. J. *Annu. Rev. Microbiol.* **2003**, *57*, 369.
(10) Torsvik, V.; Goksoyr, J.; Daae, F. L. *Appl. Environ. Microbiol.* **1990**, *56*, 782.
(11) Nakase, K.; Nakajima, S.; Hirayama, M.; Kondo, H.; Kojiri, K.; Suda, H. JP 2000178274, 2000.
(12) Feng, Z.; Kim, J. H.; Brady, S. F. *J. Am. Chem. Soc.* **2010**, *132*, 11902.
(13) Primers: StaDVF, GTSATGMTSCAGTACCTSTACGC; StaDVR, YTCVAGCTGRTAGYCSGGRTG.
(14) HR-ESI-MS (m/z): (**1**) $C_{23}H_{19}ClN_3O_5$ $[M + H]^+$: calcd, 452.1013; found, 452.10193. (**2**) $C_{22}H_{17}ClN_3O_5$ $[M + H]^+$: calcd, 438.0778; found, 438.0844. (**3**) $C_{22}H_{19}ClN_3O_4$ $[M + H]^+$: calcd, 424.1064; found, 424.1048. (**4**) $C_{22}H_{19}ClN_3O_4$ $[M + H]^+$: calcd, 424.1064; found, 424.1068. (**5**) $C_{23}H_{20}N_3O_5$ $[M + H]^+$: calcd, 418.1403; found, 418.1383. (**6**) $C_{22}H_{20}N_3O_4$ $[M + H]^+$: calcd, 390.1454; found, 390.1457. (**7**) $C_{20}H_9ClN_3O_2$ $[M - H]^-$: calcd, 358.0462; found, 358.0385.
(15) Williams, D. E.; Davies, J.; Patrick, B. O.; Bottriell, H.; Tarling, T.; Roberge, M.; Andersen, R. J. *Org. Lett.* **2008**, *10*, 3501.
(16) The predicted molecular formula for **8** [HR-ESI-MS (m/z): $C_{22}H_{17}N_3O_4Na$ $[M + Na]^+$: calcd, 410.1117; found, 410.1119] differs from **5** by "CH_2O". A comparison of the ^1H NMR spectra of **5** and **8** showed that **8** does not contain the C12 methoxy protons and that it has a new NH hydrogen. Most of the HMBC correlations seen in **8** also appeared in **5**, but new HMBC correlations from the C7a OH to C7b and from the NH to C7b, C12a, and C11a indicated that the indole in **8** has not been flipped.
(17) Sanchez, C.; Zhu, L.; Brana, A. F.; Salas, A. P.; Rohr, J.; Mendez, C.; Salas, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 461.
(18) van Berkel, W. J.; Kamerbeek, N. M.; Fraaije, M. W. *J. Biotechnol.* **2006**, *124*, 670.
(19) Widersten, M.; Gurell, A.; Lindberg, D. *Biochim. Biophys. Acta* **2010**, *1800*, 316.
(20) Howard-Jones, A. R.; Walsh, C. T. *J. Am. Chem. Soc.* **2006**, *128*, 12289.
(21) Staker, B. L.; Feese, M. D.; Cushman, M.; Pommier, Y.; Zembower, D.; Stewart, L.; Burgin, A. B. *J. Med. Chem.* **2005**, *48*, 2336.
(22) Prade, L.; Engh, R. A.; Girod, A.; Kinzel, V.; Huber, R.; Bossemeyer, D. *Structure* **1997**, *5*, 1627.
(23) Salas, J. A.; Mendez, C. *Curr. Opin. Chem. Biol.* **2009**, *13*, 152.