Microbial Genomics as a Guide to Drug Discovery and Structural Elucidation: ECO-02301, a Novel Antifungal Agent, as an Example[†]

James B. McAlpine,* Brian O. Bachmann,[‡] Mahmood Piraee, Steve Tremblay, Anne-Marie Alarco, Emmanuel Zazopoulos, and Chris M. Farnet

Ecopia BioSciences, Inc., 7290 Frederick Banting Street, Montréal, Québec, H4S 2A1, Canada

Analysis of the genome of *Streptomyces aizunensis* NRRL B-11277 indicated its potential to produce a compound of novel and highly predictable structure. The structure was predicted with sufficient accuracy to allow straightforward detection of the specific metabolite in HPLC profiles of fermentation extracts and hence to guide the isolation. The spectroscopic work was reduced to a confirmation of structure rather than a first principle determination. The compound, ECO-02301 (**1**), demonstrated potent antifungal activity. This work exemplifies not only the discovery of novel antibiotics from well-characterized organisms but also the utility of genomics as a further tool, complementary to spectroscopy, to enable rapid determination of complex structures.

Natural products have been, and continue to be, a leading source of molecules for drug discovery, but new technologies are required to increase the probability of identifying new structures. Over the past 15 years the knowledge and understanding of the genes involved in secondary metabolite biosynthesis have progressed to the point where we are now able to discern the complex chemical structure of a secondary metabolite, which will result from the enzymes produced following expression of a particular set of genomic sequences. This is particularly true for metabolites produced by multimodular enzymes such as the type I polyketide synthases (PKSs)[1] or nonribosomal peptide synthetases (NRPSs),[2] where the order of biosynthetic steps is directly reflected in the gene organization. This allows for a completely new approach to the discovery of novel antibiotics and other secondary metabolites and adds another source of data for use in the elucidation of structure. Extrapolation of advances over the last 10 years renders it probable that the genomics of secondary metabolite biosynthesis will provide another tool in determination of structure comparable to the role spectroscopic/spectrometric interpretation has assumed over the last 60 years.

The actinomycetes, an order of filamentous bacteria, have proven to be a very rich source of secondary metabolites. As these organisms are prokaryotes, they provide an excellent ground for metabolite identification via genomic DNA sequencing. This approach is rendered feasible by the fact that all of the genes encoding the large number of enzymes required for the synthesis of a typical secondary metabolite are clustered in a tight locus.[3] These loci can be identified by examination of a modest number of genome sequence tags (GSTs) using a shotgun DNA sequencing approach termed genome scanning.[4] A GST with a signature of any of a large number of known biosynthetic genes then serves as a probe to identify a cosmid containing that GST and thence to allow sequencing of the entire locus from which it came. With this approach we typically find more than 10 loci coding for complex secondary metabolites in any given actinomycete and, perhaps more surprising, that even well-studied strains, including strains producing commercial antibiotics, contain several loci providing the biosynthetic capability to produce hitherto unreported secondary metabolites.

Here we report on the use of genome mining in the discovery and structural elucidation of a novel antifungal antibiotic from *Streptomyces aizunensis* NRRL B-11277, exemplifying not only the discovery of novel antibiotics from well-characterized organisms but also the utility of genomics as a further tool, complementary to spectroscopy, to enable rapid determination of complex structures.

Results and Discussion

Streptomyces aizunensis NRRL B-11277 was reported to produce bicyclomycin in published fermentation screening studies.[5] Surprisingly, genome scanning identified 11 natural product gene clusters, in addition to the bicyclomycin cluster, coding for the biosynthesis of a wide range of structural types. One of these gene clusters included 35 open reading frames (ORFs), including a large, type I PKS system predicted to generate a long polyketide backbone containing two polyene chromophores (Figure 1a). Phylogenetic analysis of the terminal thioesterase (TE) predicted a linear polyketide product, as this domain showed greater similarity to TE domains of PKS systems synthesizing linear polyketides than to TE domains of systems synthesizing macrolide polyketides (Figure 2). Four ORFs (18, 26, 27, and 32) in the cluster provided genes coding for enzymes with the capacity to convert arginine into γ-aminobutyryl CoA and to load the activated γ-aminobutyryl group onto the ACP domain of the loading module of the PKS system, thus implicating this group as the likely starter unit for polyketide biosynthesis (Figure 1a). Similar schemes have been proposed for the initiation of polyketides using starter units derived from arginine or ornithine.[6] Two ORFs (34, 35) provided the proteins for the synthesis of 5-aminolevulinate and its conversion via a coenzyme A ester to aminohydroxycyclopentenone, which is condensed onto the carboxy terminus of the polyketide by the enzyme coded for by ORF 33 (Figure 1a). This pathway is supported by the biosynthetic studies of Floss, Ōmura, and colleagues, who postulated that the aminohydroxycyclopentenone moiety found in the asukamycins is formed by intramolecular
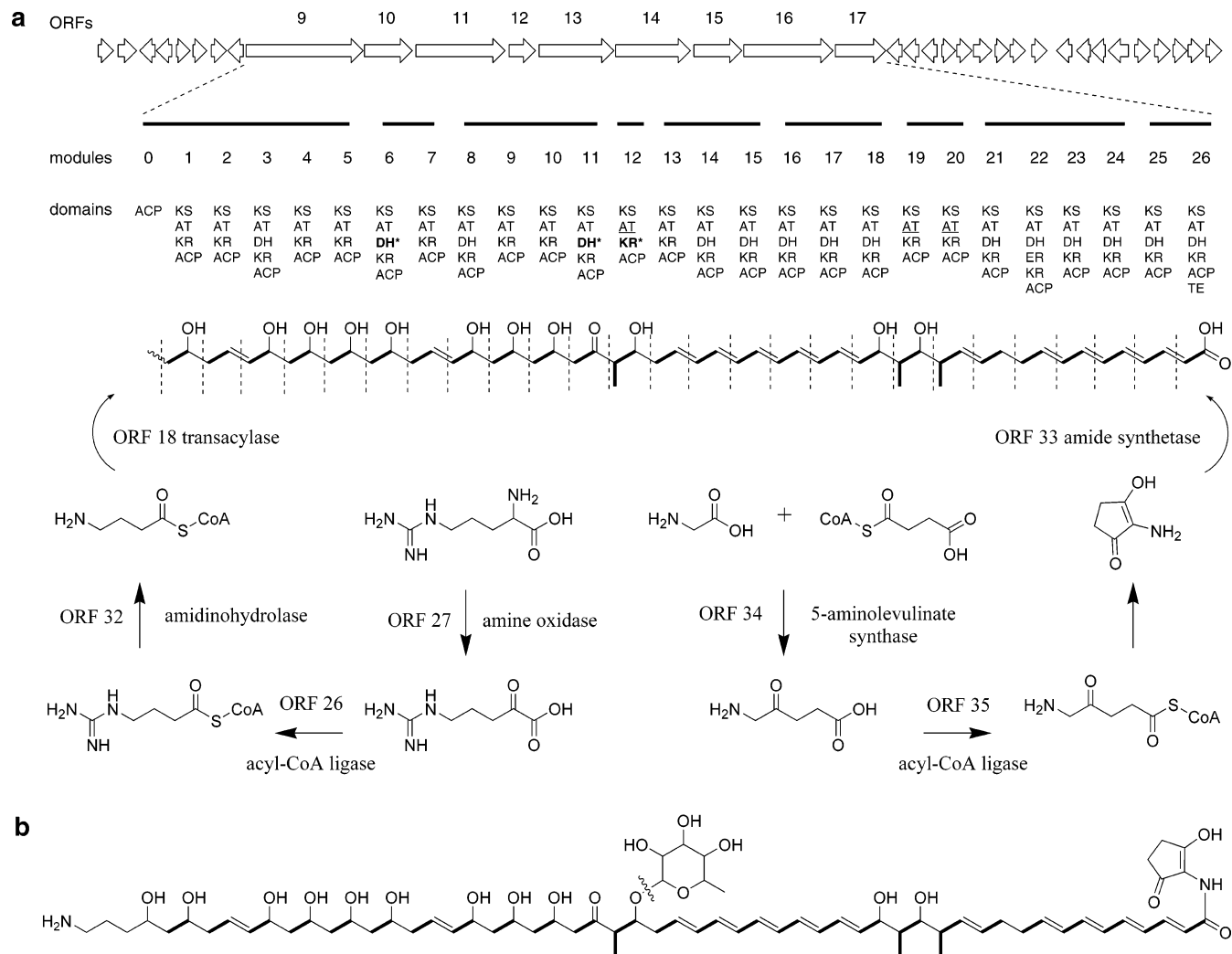
Figure 1. (a) Genomic arrangement of the locus coding for ECO-02301 (**1**) biosynthesis. Open reading frames (ORFs) 9 through 17 encode the type I polyketide synthetase, and the domains of the modules from these ORFs are shown. ACP = acyl carrier protein, KS = ketosynthase, AT = acyl transferase, KR = ketoreductase, DH = dehydratase, ER = enoylreductase, TE = thioesterase. The domains in bold type and asterisked are nonfunctional. (The KR in module 12 has a glutamine instead of a tyrosine at position 149 of the catalytic triad, and the DHs in modules 6 and 11 contain substitutions of charged amino acids arginine and glutamic acid, respectively, at position 68 for noncharged aliphatic amino acids.) The ATs all specify malonyl CoA except for those underscored, which specify methylmalonyl CoA. (b) The structure of ECO-02301 (**1**) as deduced from genomic analysis of the biosynthetic cluster. (Glycosidic linkage by HMBC.)
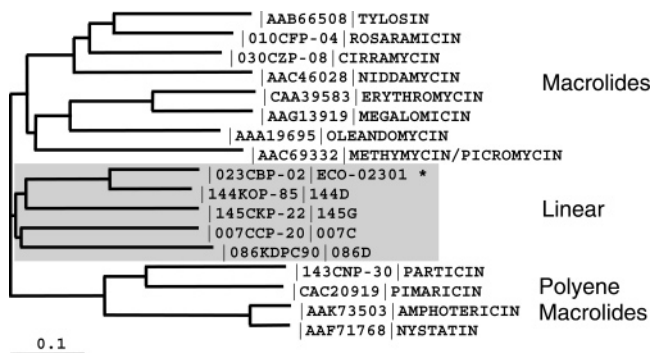

Figure 2. Phylogenetic analysis of several thioesterases from Ecopia's database.

cyclization of 5-aminolevulinate.[7] Other genes in the cluster coded for a glycosyl transferase (ORF 9) and four enzymes (ORFs 22−25) required for conversion of glucose to a 6-deoxyhexose, indicating that the product likely contained a single sugar residue. Full analysis of the gene cluster led to the prediction of a linear polyene polyketide metabo-

lite of structure **1** (Figure 1b), where the position of the glycosidic linkage to the polyketide backbone was uncertain. Irrespective of the placement of the sugar, searches against chemical databases indicated that the structure of the predicted metabolite would be novel. The predicted structure also suggested physicochemical properties (e.g., mass approximately 1297 Da and characteristic UV absorbance imparted by the pentaene chromophore) that could be used to detect the corresponding metabolite in fermentation beers with a high degree of certainty.

To obtain expression of these genes, *S. aizunensis* NRRL B-11277 was grown in shaken flasks in nearly 50 different media designed for the production of secondary metabolites. At harvest an equal volume of MeOH was added to the broths, which were then vortexed and centrifuged and the supernatant liquid was concentrated 10-fold to give an exhaustive extract, which was analyzed by HPLC monitored by DAD/UV and positive and negative ion MS. Fractions were collected and tested for antimicrobial activity. A few of the extracts yielded compounds with similar chromatographic retention times, UV absorption λ_{max} at
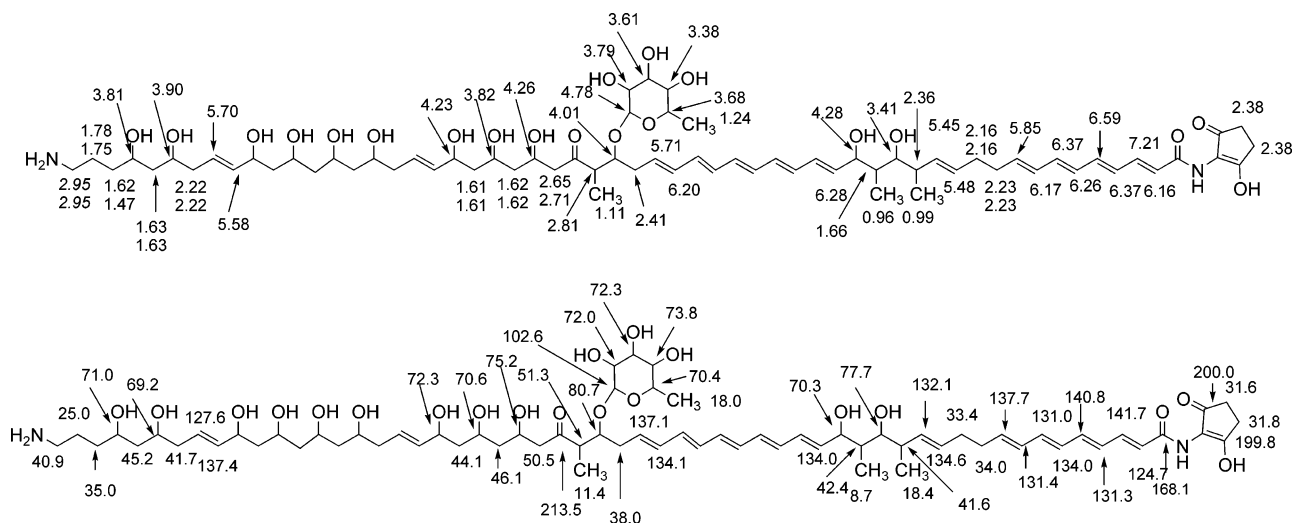
Figure 3. Planar structure of ECO-02301 (**1**) showing ^{1}H NMR (top) and ^{13}C NMR (bottom) signal assignments for spectra of the compound in methanol-d_4.

319, 333, and 351 nm, and strong MS peaks at *m/z* 1298 (positive ion mode) and 1296 (negative ion mode). These physical characteristics were entirely consistent with a compound of formula **1**. Fractions containing this metabolite exhibited strong activity against *Candida albicans*.

A high-yielding medium was chosen, and the organism was regrown on a larger scale. ECO-02301 (**1**) was extracted from the mycelial pellet with MeOH and acetone, and from the broth with Diaion HP-20 resin. After the resin had been washed with MeOH/H$_2$O it was eluted with MeOH. The crude extracts were purified by HPLC on a Waters Xterra C-18, and ECO-02301 was obtained as a yellow solid of MW (HRMS) 1297.7566 (C$_{70}$H$_{109}$N$_2$O$_{20}$ requires 1297.7568), $\lambda_{max}(\epsilon)$ 319 (110 000), 334 (146 000), and 351 (128 000) nm and was the subject of a series of 1D and 2D NMR measurements in methanol-d_4, including a gDQCOSY, gHSQC, gHMBC, TOCSY, gHSQCTOXY, and several 1D TOCSY experiments.

Analysis of these experiments led to the structure and assignments shown in Figure 3. Although considerable overlap of signals complicated assignments of all of the signals to specific protons and carbons, those that could be made unambiguously confirmed the structure predicted from the genomics. A major cross-peak between the well-separated proton resonance at 4.01 ppm and the anomeric carbon at 102.6 ppm placed the sugar as shown in Figure 3. A well-resolved carbon spectrum with high signal-to-noise ratio showed that the unassigned methylene carbons were at 42.0, 45.3, 45.4 and 46.6. gHSQC analysis indicated that these were attached to protons at 2.24, 1.62, 1.50 and 1.68, and 1.55, respectively. Similarly the unassigned carbinols were at 66.2, 66.2 (resolved), 67.2, and 69.0 attached to protons at 4.06, 4.08, 4.22, and 3.89, and the unassigned olefinic carbons were at 129.1, 131.0, 131.9, 133.3, 133.7, 134.3, 134.8, 136.5, and 138.0 attached to protons at 5.72, 5.72, 6.28, 6.25, 6.28, 6.25, 6.19, 5.53, and 5.86, respectively. Almost certainly the traditional approach of changing solvents, to find, a posteriori, a situation with an appropriate dispersion of chemical shifts, would have worked and allowed a stepwise assignment of sufficient signals to define the two incompletely assigned structural segments (one of which is defined by the UV spectrum). However this laborious approach, requiring complete reassignment of all signals, was obviated by the genomic analysis, and the planar structure was elucidated beyond reasonable doubt.

The amidohydroxycyclopentenone signals were not straightforward and reflected the tautomeric equilibrium of this moiety. The upfield methylene signal and the downfield carbonyl signals were only 10% of the intensity of those from the other tautomer. The signal from C-1 of this moiety was not detected, a phenomenon that was previously ascribed[8] to tautomerization for the same structural unit in the moenomycin type antibiotic, AC326-α.

These results clearly demonstrate the advanced state of the genomics of microbial secondary metabolite biosynthesis and the potential of a new approach to natural product discovery. The analysis of the genome of a well-characterized microorganism indicated its potential to produce a novel and highly defined structure. The structure was predicted with sufficient accuracy to allow straightforward detection of the specific metabolite in HPLC profiles of fermentation extracts and hence to guide its isolation. The spectroscopic work was reduced to a confirmation of structure rather than a first principle determination. This work demonstrates the utility of genome mining in the discovery of natural products having new structures, and the particular example of ECO-02301 (**1**) demonstrates the significant role that genomic analysis can play in the structure determination of novel natural products. Finally we established that ECO-02301 (**1**) exhibited potent antifungal activity against a variety of fungi pathogenic to humans (Table 1).

Experimental Section

General Experimental Procedures. Analytical HPLC was carried out with a Waters Alliance 2690 instrument equipped with a Micromass ZQ electrospray source and Waters 996 diode array UV detectors. Semipreparative HPLC was done on a Waters 1525 instrument with a Waters 2996 diode array UV detector. All NMR spectra were measured in MeOH-d_4 on a Varian Unity Inova 500 MHz spectrometer operating at 500 MHz for ^{1}H and 125 MHz for ^{13}C nuclei and using a 5 mm probe for the former and a 3 mm probe for the latter. HRMS were measured by ESI on a Ionspec Ultima 7T Fourier transform mass spectrometer from a solution in 50% aqueous CH$_3$CN/0.1% HCOOH.

Genome Scanning. The genome of *S. aizunensis* NRRL B-11277 was analyzed by genome scanning technique as described.[4] The DNA and protein sequences that comprise the ECO-02301 gene cluster are deposited in GenBank under accession numbers AY899213 and AY899214.

Table 1. MICs of ECO-02301 (**1**) against Various Strains of Pathogenic Fungi[a]

	MIC (μg/mL)	
strain	ECO-02301	amphotericin B
C. albicans ATCC10231	4	1
C. glabrata ATCC 90030	4	1
C. lusitaniae ATCC 200953	4	1
C. tropicalis ATCC 200955	4	2
C. krusei LSPQ 0309	4	1
S. cerevisiae ATCC 9763	4	1
Aspergillus fumigatus ATCC 204305	16	2
Aspergillus flavus ATCC 204304	≥ 8	4
Cryptococcus neoformans ATCC 32045	4−8	0.5

[a] The antifungal potency of ECO-02301 (**1**) was evaluated by determining the minimum inhibitory concentration (MIC) of the compound according to the NCCLS guidelines (*Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts; Approved Standard.* NCCLS document M27-A, Vol. 17, No. 9, 1997; *Reference Method for Broth Dilution Antifungal Susceptibility Testing of Filamentous Fungi; Approved Standard.* NCCLS document M38-A, Vol. 22, No. 16). Amphotericin B was used as comparator.

Fermentation. *S. aizunensis* NRRL B-11277 was maintained as frozen glycerol stock at −80 °C. Inoculum for fermentation was obtained by streaking freshly thawed frozen stock onto tomato paste-oatmeal agar (ATCC medium 1360) and incubating at 28 °C until a thick growth of vegetative mycelium was apparent (5 to 7 days). Two to three loopfuls of this mycelium was homogenized in 1.3 mL of ITSB medium (Trypticase soy broth (30 g), yeast extract (3 g), $MgSO_4 \cdot 7H_2O$ (2 g), glucose (5 g), and maltose (4 g) per L of distilled H_2O), and this was used to inoculate two 25 mL portions of ITSB medium in 125 mL Erlenmeyer flasks each containing three glass beads. These were incubated at 28 °C for 65−70 h on a rotary shaker at 250 rpm, and 10 mL aliquots were used as inocula for 2 L baffled flasks containing 500 mL of production medium (soluble starch (10 g), glucose (12 g), Pharmamedia (10 g, Sigma), corn steep liquor (5 g, Sigma), and Proflo oil (4 mL, Traders Protein) per L of distilled H_2O adjusted to pH 7.2). Production fermentation was carried out at 28 °C for 7 days on rotary shakers at 250 rpm.

Extraction and Isolation. At harvest the contents of 4 × 2 L flasks were combined and cooled to 4 °C, the pH was adjusted to 3.5 with 1 N HCl, and the mixture was allowed to stand at that temperature overnight. The whole broth was centrifuged at 2500 rpm for 30 min, and the supernatant was decanted from the mycelial pellet, which was extracted with MeOH (800 mL) stirred for 1 h. The mixture was centrifuged, and the supernatant was decanted. The precipitate was further extracted similarly with acetone (800 mL), the combined extracts were diluted with H_2O (200 mL), the pH was adjusted to 6.5, Diaion HP-20 (100 mL) was added, and MeOH was removed on a rotary evaporator. The sample was loaded onto a prepacked column (1 × 10 in.) of Diaion HP-20, and the decanted broth, after pH adjustment to 6.5, was passed through the column. The column was washed with H_2O (200 mL) and then with aliquots (60 mL each) of 5, 15, 25, 50, and 75% MeOH. The crude product was eluted with 100% MeOH (300 mL). Solvent was removed from the MeOH eluates, and the residue was dissolved in MeOH (2 mL) and diluted with H_2O (2 mL). The mixture was centrifuged, and the supernatant liquid was chromatographed, in 500 μL aliquots, on a Waters Xterra C-18 HPLC column (19 × 150 mm) developed with an aqueous NH_4OAc/CH_3CN gradient and monitored by DAD/UV. Eco-02301 eluted around 20.5 min in a 30 min gradient from 5 to 50% CH_3CN. Pooled conservative collection of the center of the UV peak gave ECO-02301 (**1**) (8.6 mg).

Supporting Information Available: This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Hopwood, D. A. *Chem. Rev.* **1997**, *97*, 2465−2497. Stauton, J.; Weissman, K. *Nat. Prod. Rep.* **2001**, *18*, 380−416.
(2) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2615−2673. Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275−287.
(3) Martin, J. F.; Liras, P. *Annu. Rev. Microbiol.* **1989**, *43*, 173−206.
(4) Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet C. M. *Nat. Biotechnol.* **2003**, *21*, 187−190.
(5) Miyamura, S.; Ogasawara, N.; Otsuka, H.; Niwayama, S.; Tanaka, H.; Take, T.; Uchiyama, T.; Ochiai, H.; Abe, K.; Koizumi, K.; Asao, K.; Matsuki, K.; Hoshino, T. *J. Antibiot.* **1972**, *25*, 610−612. Miyamura, S.; Ogasawara, N.; Otsuka, H.; Niwayama, S.; Tanaka, H.; Take, T.; Uchiyama, T.; Ochiai, H. *J. Antibiot.* **1973**, *26*, 479−484.
(6) Moore, B. S.; Hertweck, C. *Nat. Prod. Rep.* **2002**, *19*, 70−99.
(7) Nakagawa, A.; Wu, T. S.; Keller, P. J.; Lee, J. P.; Ōmura, S.; Floss, H. G. *J. Chem. Soc., Chem. Commun.* **1985**, *1985*, 519−521.
(8) He, H.; Shen, B.; Korshalla, J.; Siegel, M. M.; Carter, G. T. *J. Antibiot.* **2000**, *53*, 191−195.

NP0401664