Natural Products

DOI: 10.1002/anie.200901209

An Environmental DNA-Derived Type II Polyketide Biosynthetic Pathway Encodes the Biosynthesis of the Pentacyclic Polyketide Erdacin**

*Ryan W. King, John D. Bauer, and Sean F. Brady**

It is now well established that environmental samples contain a significantly more diverse collection of bacteria than that which is readily grown in the laboratory.[1–4] Large insert clone libraries of DNA extracted directly from environmental samples (environmental DNA, eDNA) provide a means to retrieve the natural product biosynthetic gene clusters found in the genomes of these previously inaccessible bacteria. One of the challenges associated with the discovery of new molecules from large eDNA libraries has been the identification of clones containing intact gene clusters that can be used to produce small molecules in heterologous hosts.[5–10] Here we report the recovery of a collection of type II (iterative, aromatic) polyketide synthase (PKS) containing clones from an eDNA cosmid library and the characterization of erdacin, a novel pentacyclic polyketide produced by *Streptomyces albus* transformed with one of these clones. This study provides one of the clearest indications yet that DNA derived from the large unstudied collection of bacteria present in most environmental samples has the potential to encode the biosynthesis of molecules that are substantively different from known metabolites and not simply derivatives of molecules previously identified in culture-based studies.

Type II PKS gene clusters contain a conserved minimal PKS that is composed of two ketosynthase genes (KS_α and KS_β) and an acyl carrier protein (ACP). The minimal PKS is responsible for the iterative condensation of malonyl-CoA into a nascent polyketide chain that can be cyclized, aromatized, reduced, rearranged, and functionalized into a vast assortment of structurally unique metabolites.[11–15] As the minimal PKS is used iteratively, type II PKS gene clusters are smaller than most natural product gene clusters that encode the biosynthesis of comparably complex metabolites. As a result, it is possible to capture functionally complete type II PKS pathways on individual cosmid clones. We hypothesized that eDNA cosmid clones containing type II PKS biosynthetic machinery would be a productive starting point for the discovery of novel secondary metabolites from eDNA cosmid libraries. While both PCR-based studies and high-throughput sequencing efforts indicate that eDNA samples contain a plethora of novel type II PKS genes, none of these previous studies have focused on the recovery of functionally intact type II PKS gene clusters, and, as a result, no new metabolites have been produced from eDNA-derived type II PKS gene clusters.[9,16–18]

Degenerate primers based on conserved regions of minimal PKS KS_α and ACP genes were used to amplify full-length KS_β sequences captured in an eDNA library constructed from desert soil collected in Utah (Figure 1).[9,19,20] Of the 21 unique KS_β sequences amplified from this eDNA library, only one showed greater than 80% identity to a previously reported KS_β gene. The low identity these sequences exhibit to known KS_β sequences suggested they are associated with gene clusters that are functionally distinct from any previously sequenced gene clusters and might, therefore, encode the biosynthesis of novel secondary metab-

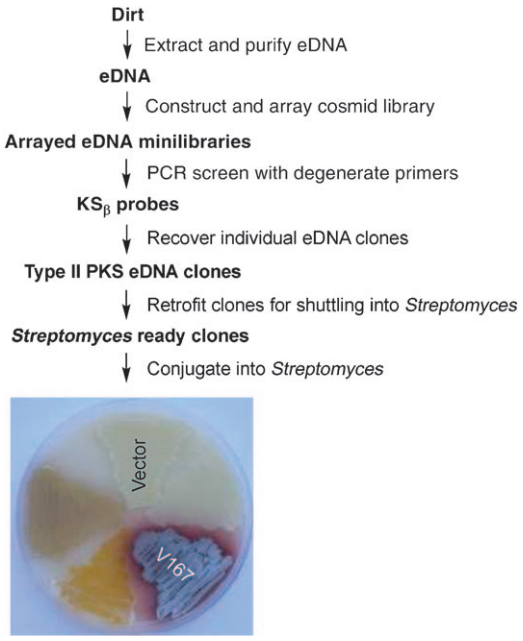

Figure 1. General scheme used for the recovery of type II PKS eDNA clones and the heterologous expression of erdacin (**1**) from clone V167. *S. albus* transformed with a vector control and eDNA-derived clones containing type II PKS biosynthetic genes are shown. The colored phenotype exhibited by some clones is indicative of the production of clone-specific small molecules.

[*] R. W. King, Dr. J. D. Bauer, Prof. S. F. Brady
Laboratory of Genetically Encoded Small Molecules
The Rockefeller University
1230 York Avenue, New York, NY 10065 (USA)
Fax: (+1) 212-327-8281
E-mail: sbrady@rockefeller.edu

[**] This work was supported by NIH GM077516, The Beckman Foundation, The Searle Foundation, and The Hershel Foundation. We thank Emil Lobkovsky for his assistance with our X-ray studies.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200901209.

olites. To explore this possibility, eDNA-derived KS_β sequences were used as probes to identify and recover ten cosmid clones containing type II PKS biosynthetic genes from the soil eDNA library. The recovered clones were retrofitted with the genetic elements necessary for conjugation and integration into *Streptomyces*; eight of the retrofitted clones were successfully integrated into both *S. lividans* and *S. albus*. Based on antibacterial assays, HPLC analysis, and color production (Figure 1), five of the eight clones appear to produce clone-specific metabolites in at least one host. V167, which produces large quantities (15–20 $mg\,L^{-1}$) of one major clone-specific metabolite in *S. albus*, was selected for further analysis.

The major clone-specific metabolite found in ethyl acetate extracts of V167 cultures grown in modified *Streptomyces*-supplemented minimal media was purified by reversed-phase HPLC, and given the trivial name erdacin (**1**).[21] Extensive 1D and 2D NMR analysis of erdacin, obtained from both ^{13}C-labeled and naturally abundant ^{12}C cultures, allowed us to unambiguously assign all but five carbon atoms (C-2, C-3, C-4, C-22, and C-23) in the final structure (Figure 2).[22] The position of the final five carbon atoms could not be assigned

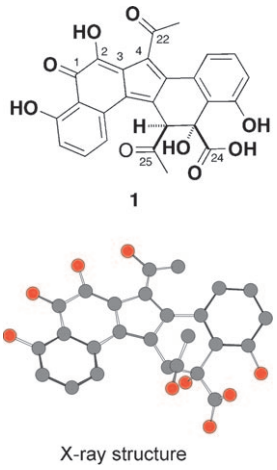

Figure 2. Chemical and computer-generated perspective diagrams of erdacin (**1**). Hydrogen atoms are not included in the computer-generated perspective drawing. The pentacyclic ring system found in **1** has not been previously reported from culture-dependent natural product studies.

with certainty because many of the NMR signals from this region of the structure were either weak or absent.

To complete the structural characterization of erdacin, crystals were obtained by slow evaporation from acetonitrile and water and then analyzed by single-crystal X-ray diffraction. The X-ray structure confirmed that erdacin is based on a novel pentacyclic (6-6-5-6-6) ring system. As drawn in Figure 2, the upper edge of the central five-membered ring is functionalized with an acetyl group, while the bottom edge of the third six-membered ring is functionalized with both an acetyl group and a carboxylate. To the best of our knowledge, erdacin is a novel natural product. We did not find the erdacin carbon skeleton nor its pentacyclic ring system to be a substructure of any known secondary metabolites characterized from culture-dependent studies. Pentacylic structures with the same number of five- and six-membered rings have been reported as natural products; however, the organization of the rings in known structures does not match that seen in erdacin.[23]

The weak NMR signals that prevented the unambiguous assignment of the complete erdacin structure by NMR spectroscopy alone likely result from an interaction between the C-2 hydroxy group and the C-22 carbonyl oxygen atom. Based on the final erdacin structure, a C-2 hydroxy-assisted tautomerization of the C-22 carbonyl group could explain both the broad NMR signals for this region of the structure and the rapid disappearance of the signals corresponding to the C-23 methyl protons that occurs in protic NMR solvents.

A hint as to the biosynthetic origin of this new structure comes from the characterization of juglomycin F (**4**),[24] a second less-abundant clone-specific metabolite found in extracts of V167 culture broth. The carbon skeleton of this known 13-carbon metabolite maps onto one half of the erdacin structure, thus suggesting that erdacin might arise from the heterodimerization of **4** (or a close relative of **4**) with a second 13-carbon subunit (Figure 3b). In an attempt to better understand the origin of **1**, the V167 insert was sequenced and transposon mutagenized. Figure 3a shows the position of key transposon insertions that result in easily discernable changes in the V167 colony phenotype. All of these transposons are found in a 20 kb region with 22 predicted open-reading frames that we have called the Erd gene cluster. The very high GC content (>90%) of some regions of the Erd gene cluster may explain our inability to obtain transposon insertions in every predicted Erd open-reading frame. The Erd gene cluster contains a minimal PKS, a collection of post PKS enzymes, regulatory proteins, a transporter, and a number of hypothetical proteins.

Transposon insertions in early post-minimal PKS biosynthetic genes (Erd4, -5, and -22) knockout the production of erdacin and unexpectedly lead to the accumulation of known octaketide shunt products instead of the 13-carbon (or 14-carbon) intermediates that are suggested by the structure of erdacin. The known early octaketide shunt products SEK4b, mutactin, and SEK34b (**2**) were found in extracts from Erd5, -4, and -22 mutants, respectively (see Figure 1 in the Supporting Information).[14] Transposon insertions in Erd20, a predicted monooxygenase, lead to the production of apparently reactive intermediates that generate a number of compounds with a range of molecular weights (*m*/*z* 240–315). One of the stable components of this mixture was fully characterized and found to be the decarboxylated octaketide prechrysophanol (**3**).[25] Based on the metabolites identified in extracts from both wild-type and mutant V167 cultures, it appears likely that the Erd gene cluster converts an octaketide produced by the minimal PKS into two distinct intermediates, which give rise to erdacin (Figure 3).

The formation of a juglomycin-like structure from an octaketide precursor has been reported previously.[26] While the isolation of juglomycin F (**4**) supports the existence of one 13-carbon precursor, the origin of the second 13-carbon

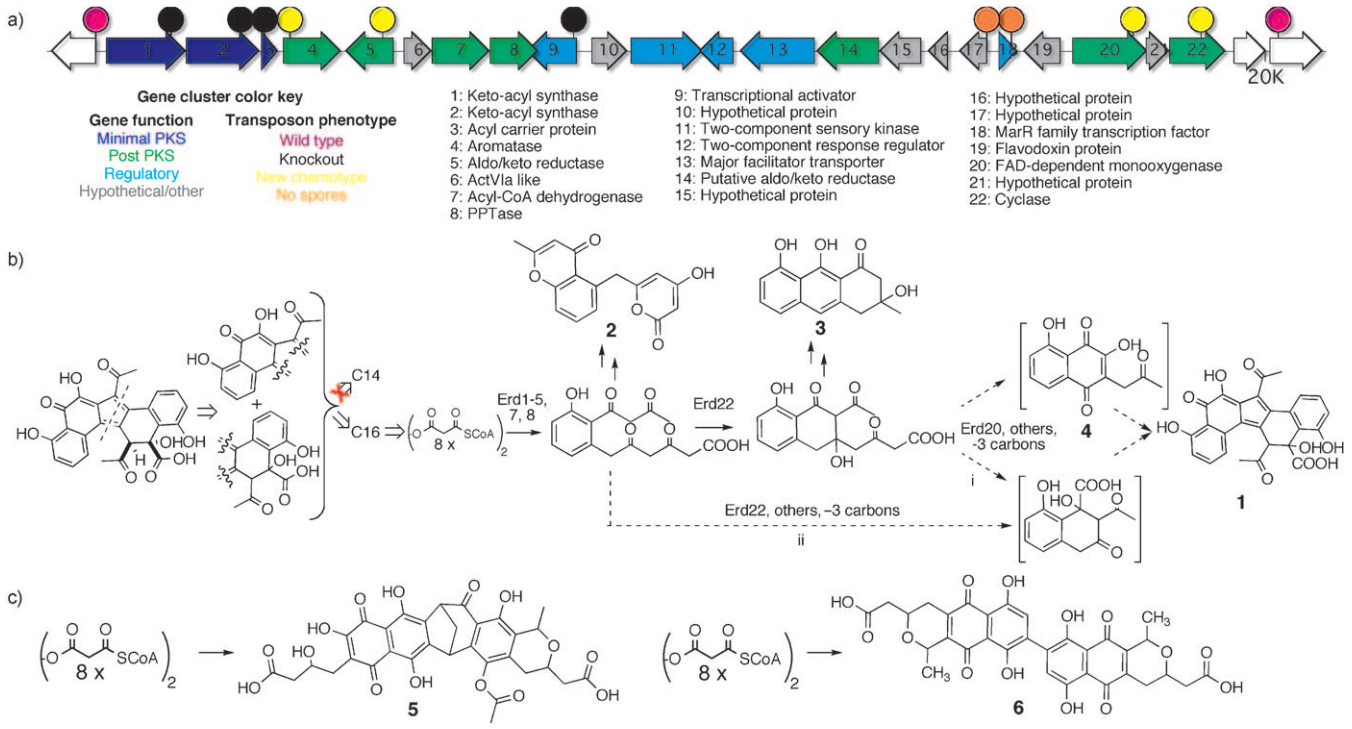

Figure 3. a) The Erd gene cluster (GenBank accession number FJ719113) and the location of key transposon insertions discussed in the text. b) A retrobiosynthetic analysis of erdacin suggests that it might arise from two 13-carbon monomers, however, only 16-carbon compounds accumulate in transposon knockouts of early post-miminal PKS genes. A proposed biosynthetic scheme for the formation of erdacin from an octaketide precursor is shown. Compounds **1**–**4** were isolated from wild-type and mutant V167 cultures. Shown in brackets are juglomycin F (**4**) and a hypothetical 13-carbon polyketide which together have carbon skeletons that accout for the upper and lower halves of erdacin (**1**), respectively. c) The Erd minimal PKS genes are related to minimal PKS genes involved in the biosynthesis of the two structurally distinct octaketide dimers naphthocyclinone (**5**) and actinorhodin (**6**).

subunit is not immediately apparent. Two possible routes for the formation of the second 13-carbon monomer with the necessary carbon skeleton are shown in Figure 3b (path i rearrangement, or path ii alternative cyclization). To distinguish between these two potential routes a [1,2-^{13}C]acetate stable isotope feeding experiment was carried out on cultures of V167. The appearance of the C-24 carboxylate as a doublet in the ^{13}C NMR spectra of erdacin derived from this feeding study indicates that C-24 and C-12 are incorporated as an intact acetate unit and therefore supports path ii as the likely route for the formation of the lower half of erdacin (Figure 4).

In the biosynthetic scheme shown in Figure 3b, both methyl ketones are predicted to arise from the decarboxylation of terminal acetates, instead of intact acetates as might be predicted from glancing at the structure of erdacin. In the ^{13}C NMR spectra of [1,2-^{13}C]acetate-labeled erdacin, the methyl carbon atoms are not coupled to the adjacent carbonyl groups, thus confirming that, as predicted, neither methyl ketone arises from the intact incorporation of acetate. Additional ^{13}C-^{13}C couplings observed in 2D INADEQUATE experiments (Figure 4b) performed on [1,2-^{13}C]acetate-labeled erdacin support the cyclization patterns suggested by both the transposon mutagenesis studies and the couplings observed in the ^{13}C NMR spectra of [1,2-^{13}C]acetate-labeled erdacin (Figure 4a).

While the exact nature of the late-stage polyketide intermediates used in the biosynthesis of erdacin is not known, it appears that erdacin arises from two distinct 13-carbon substructures that are derived from a common octaketide intermediate. In the proposed biosynthetic scheme, a common octaketide precursor would undergo two distinct cyclizations. At some point in the biosynthetic process, these octaketide intermediates would lose the same three carbon atoms (terminal carboxylate and terminal methyl ketone) to yield a pair of
13-carbon intermediates that make up the two halves of erdacin (Figure 4).

The Erd minimal PKS enzymes are most closely related to ACP, KS_α, and KS_β sequences from gene clusters that encode the biosynthesis of naphthocyclinon, actinorhodin, granaticin, and medermycin.[14,27–29] While granaticin and medermycin are simple octaketide structures, naphthocyclinone (**5**) and actinorhodin (**6**) are octaketide dimers that arise from two distinct dimerization strategies (Figure 3c).[14,27–29] Erdacin appears to result from a third previously unseen octaketide dimerizatoin strategy. By leveraging biosynthetic information from sequenced gene clusters we have been able to isolate an eDNA-derived gene cluster that is sufficiently different from known clusters and encodes the biosynthesis of a structurally unprecedented secondary metabolite.

Transposon insertions in either the hypothetical protein Erd17 or the MarR-like transcription factor Erd18 lead to mutants that no longer sporulate, which indicates that the Erd gene cluster may interfere with sporulation.[30] During our exploration of the potential bioactivities of erdacin, we found that it exhibits significant antioxidant activity. In a standard

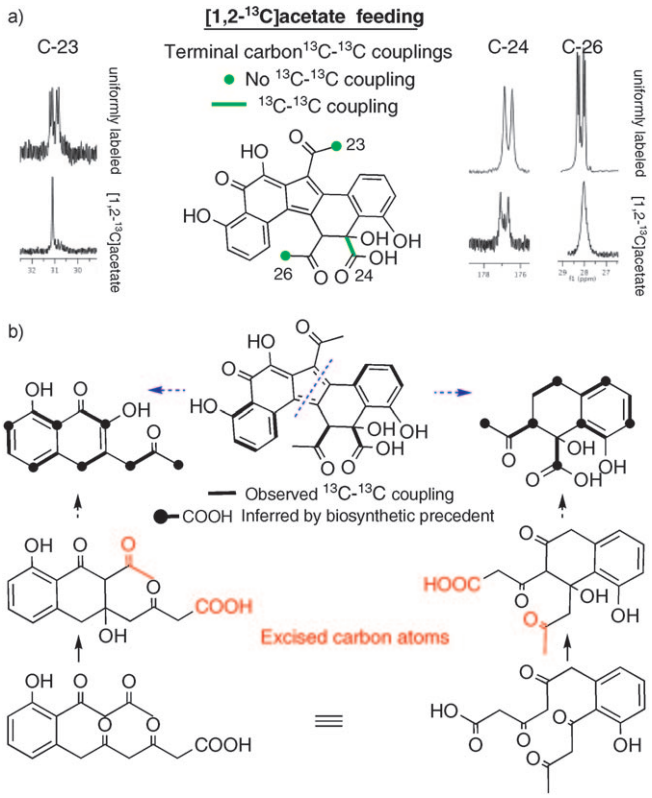

Figure 4. ^{13}C-^{13}C couplings (or in some cases the absence of coupling) observed in ^{13}C (a) and 2D INADEQUATE (b) NMR spectra of [1,2-^{13}C]acetate-labeled erdacin suggest that two distinct polyketide cyclization patterns are used in the biosynthesis of erdacin. Coupling between weak or overlapping NMR signals not clearly seen by 2D INADEQUATE were inferred based on polyketide biosynthetic precedent. The exact structure of the late-stage polyketide intermediates used in the biosynthesis of erdacin is not known.

copper reduction assay, erdacin exhibits antioxidant activity that is twice that reported for many well-known antioxidants, including ascorbic acid (vitamin C).[31] Studies are currently underway to further explore the bioactivity of erdacin.

The systematic recovery and large-scale investigation of eDNA clones containing minimal PKS gene clusters is likely to be a productive avenue for the culture-independent discovery of structurally unique secondary metabolites. More generally, this study suggests that "compact" natural product biosynthetic pathways that can be captured on individual cosmid sized clones (namely, type II PKS, type III PKS, aminoglycosides) are likely to be rewarding systems to examine in the search for novel metabolites from eDNA libraries. The discovery of erdacin, from the small sample of eDNA-derived clones investigated in this study, provides tangible evidence that the diverse collection of gene clusters predicted to be present in environmental bacteria likely encodes the biosynthesis of metabolites that are, in many cases, distinct from those that have been characterized by using traditional cultured-based discovery strategies.

Received: March 3, 2009
Revised: June 9, 2009
Published online: July 17, 2009

Keywords: biosynthesis · DNA · metagenomics · natural products · polyketides

[1] M. S. Rappe, S. J. Giovannoni, *Annu. Rev. Microbiol.* **2003**, *57*, 369.
[2] V. Torsvik, L. Ovreas, T. F. Thingstad, *Science* **2002**, *296*, 1064.
[3] T. P. Curtis, W. T. Sloan, J. W. Scannell, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10494.
[4] V. Torsvik, K. Salte, R. Sorheim, J. Goksoyr, *Appl. Environ. Microbiol.* **1990**, *56*, 776.
[5] J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16222.
[6] A. Schirmer, R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, C. R. Hutchinson, *Appl. Environ. Microbiol.* **2005**, *71*, 4840.
[7] S. Courtois, C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, J. L. Pernodet, *Appl. Environ. Microbiol.* **2003**, *69*, 49.
[8] S. F. Brady, J. D. Bauer, M. F. Clarke-Pearson, R. Daniels, *J. Am. Chem. Soc.* **2007**, *129*, 12102.
[9] K. T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, J. Davies, *J. Bacteriol.* **1997**, *179*, 7360.
[10] A. Ginolhac, C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T. M. Vogel, G. Perriere, P. Simonet, R. Nalin, *Appl. Environ. Microbiol.* **2004**, *70*, 5522.
[11] C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162.
[12] B. Shen, *Curr. Top. Chem.* **2000**, *209*, 1.
[13] R. McDaniel, S. Ebert-Khosla, D. A. Hopwood, C. Khosla, *Nature* **1995**, *375*, 549.
[14] R. McDaniel, S. Ebert-Khosla, D. A. Hopwood, C. Khosla, *J. Am. Chem. Soc.* **1994**, *116*, 10855.
[15] A. L. Matharu, R. J. Cox, J. Crosby, K. J. Byrom, T. J. Simpson, *Chem. Biol.* **1998**, *5*, 699.
[16] B. Wawrik, L. Kerkhof, G. J. Zylstra, J. J. Kukor, *Appl. Environ. Microbiol.* **2005**, *71*, 2232.
[17] M. Pang, G. A. Tan, N. Abdullah, C. Lee, C. Ng, *Biotechnology* **2008**, *7*, 660.
[18] B. Wawrik, D. Kutliev, U. A. Abdivasievna, J. J. Kukor, G. J. Zylstra, L. Kerkhof, *Appl. Environ. Microbiol.* **2007**, *73*, 2982.
[19] S. F. Brady, *Nat. Protoc.* **2007**, *2*, 1297.
[20] J. J. Banik, S. F. Brady, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17273.
[21] Four-day-old cultures of V167 grown in modified supplemented minimal media (30 °C at 250 rpm) were acidified to pH 3 and then extracted twice with an equal volume of ethyl acetate. The dried extract was dissolved in methanol, passed through a SPE previal C-18 plug, and then partitioned by reversed-phase HPLC (XBridge C18, 10 mm × 150 mm, 5 μm, 2 min of 20:80 acetonitrile/water with 0.1% formic acid followed by a gradient of increasing acetonitrile (20:80 to 50:50) over 24 min, 7 mL min^{-1}). Erdacin (**1**) eluted at 14.9 min. The trivial name erdacin is derived from "erda", the old english word for earth or soil.
[22] Erdacin (**1**): ^1H NMR (600 MHz, CD_3OD with 0.1% trifluoroacetic acid): $\delta = 7.40$ (m, 2H; H-17,18), 7.28 (t, $J = 7.7$ Hz, 1H; H-8), 7.16 (d, $J = 7.1$ Hz, 1H; H-7), 6.92 (d, $J = 7.7$ Hz, 1H; H-9), 6.73 (d, $J = 7.7$ Hz, 1H; H-19), 4.29 (s, 1H; H-13), 2.53 (s, 3H; H-23), 1.72 ppm (s, 3H; H-26); ^{13}C NMR (150 MHz, CD_3OD with 0.1% trifluoroacetic acid): $\delta = 205.5$ (C-22), 203.5 (C-25), 188.4 (C-1), 176.6 (C24), 165.3 (C-20), 156.7 (C-10), 153.7 (C-2), 143.3 (C-5), 138.8 (C-18), 135.6 (C-16), 134.6 (C-14), 133.4 (C-6), 131.3 (C-8), 131.2 (C-4), 130.6 (C-15), 127.7 (C-3), 123.8 (C-11), 120.6 (C-7), 119.1 (C-17), 118.7 (C-9), 118.4 (C-19), 114.1 (C-21), 75.5 (C-12), 64.6 (C-13), 30.3 (C-23), 27.9 ppm (C-26); HRESI-MS:

m/z 475.1037, calcd for $C_{26}H_{19}O_9$, 475.1029. ^1H and ^{13}C spectra for erdacin are included in the Supporting Information. CDCC 721152 contains the supplementary crystallography data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data request/cif.

[23] H. Lessmann, R. P. Maskey, S. Fotso, H. Lackner, H. Laatsch, *Z. Naturforsch. B* **2005**, *60*, 189.

[24] H. Lessmann, J. Krupa, H. Lackner, P. G. Jones, *Naturforsch. B* **1989**, *44*, 353.

[25] A. Yenesew, J. A. Ogur, H. Duddeck, *Phytochemistry* **1993**, *34*, 1442.

[26] M. Ozawa, T. Taguchi, T. Itoh, Y. Ebuizuka, K. I. Booker-Milburn, G. R. Stephenson, K. Ichinose, *Tetrahedron* **2003**, *59*, 8793.

[27] P. Brünker, K. McKinney, O. Sterner, W. Minas, J. E. Bailey, *Gene* **1999**, *227*, 125.

[28] K. Ichinose, D. J. Bedford, D. Tornus, A. Bechthold, M. J. Bibb, W. P. Revill, H. G. Floss, D. A. Hopwood, *Chem. Biol.* **1998**, *5*, 647.

[29] K. Ichinose, M. Ozawa, K. Itou, K. Kunieda, Y. Ebizuka, *Microbiology* **2003**, *149*, 1633.

[30] S. P. Wilkinson, A. Grove, *Curr. Issues Mol. Biol.* **2006**, *8*, 51.

[31] R. L. Prior, X. Wu, K. Schaich, *J. Agric. Food Chem.* **2005**, *53*, 4290.

