



Arimetamycin A: Improving Clinically Relevant Families of Natural Products through Sequence-Guided Screening of Soil Metagenomes**

Hahk-Soo Kang and Sean F. Brady*

The chemical diversity encoded by natural microbial communities has been significantly under-explored due to limitations associated with the inability to culture the majority of environmental bacteria^[1] and the silencing of biosynthetic pathways under laboratory conditions.^[2] Soils are predicted to contain thousands of unique bacterial species, which potentially harbor tens of thousands of functionally unexplored gene clusters that encode for the biosynthesis of natural products.^[3] With the development of metagenomic cloning methods, it is now possible to use DNA extracted directly from soil (environmental DNA, eDNA) to construct libraries that capture the enormous biosynthetic diversity present in soil environments.^[4] These libraries provide a means of functionally examining unexplored soil biosynthetic gene clusters and are, therefore, appealing resources for sequence-guided natural product discovery programs.^[5] Based on the magnitude of the biosynthetic diversity captured in saturating soil eDNA libraries, we believed that it would be possible to use these libraries to identify novel members of clinically relevant natural product families with potentially improved biological activities. To this end, we used a natural product sequence-tag-driven approach to guide the discovery of an anthracycline-based aromatic polyketide that shows improved in vitro antiproliferative activity compared to the natural product anthracyclines that are currently in clinical use. In heterologous expression experiments the eDNA-derived *arm* cluster was found to encode for arimetamycin A, an anthracycline that is more potent than clinically used natural anthracyclines and retains activity against multidrug-resistant (MDR) cancer cells.

Aromatic polyketides comprise a large class of structurally and functionally diverse bacterial natural products.^[6] The biosyntheses of these metabolites, although differing in detail, all originate with the production of a polyacetate precursor by a conserved minimal polyketide synthase (min-PKS) that is composed of three proteins [β -ketoacyl synthase alpha (KS_{α}), β -ketoacyl synthase beta (KS_{β}), acyl carrier protein (ACP)] (Figure 1).^[6] KS_{α} and KS_{β} genes clade into groups that

correlate strongly with the specific aromatic polyketide structural classes that are encoded by the gene clusters in which they reside.^[7] Extrapolating from this observation, we hypothesized that it should be possible to systematically screen large eDNA libraries for biosynthetic gene clusters capable of encoding novel members of clinically relevant aromatic polyketide structural classes using only polymerase chain reaction (PCR)-amplified min-PKS gene fragments as sequence tags. To test the utility of this approach, cosmid DNA isolated from four previously archived and arrayed soil eDNA megalibraries (AB, UT, AZ, and TX) was screened by PCR using min-PKS-specific degenerate primers (Figure 1).^[8] PCR amplicons generated during the screening of individual sublibraries were sequenced and phylogenetically compared to the corresponding KS_{β} fragments from biosynthetic gene clusters known to encode for clinically relevant polyketide families. eDNA-derived KS_{β} sequences that fell into clades associated with the biosynthesis of a variety of clinically useful antibiotics and anticancer structural classes were identified in this analysis, suggesting that gene clusters capable of expanding the structural diversity seen within these molecule families were potentially present in our archived libraries (Figure 1).

A group of KS_{β} sequence tags of particular interest to us was one that fell into the clades with KS_{β} genes used in the biosynthesis of anthracycline-type molecules (Figure 1). The anthracyclines doxorubicin and daunorubicin, derived from the cultured bacterium *Streptomyces peucetius*, are both potent anticancer agents.^[9] They are used extensively in the treatment of diverse cancers including soft-tissue sarcomas, hematological malignancies, and a variety of other carcinomas. The clinical use of these drugs is hampered, however, by the development of resistance and cumulative cardiotoxicity issues.^[9a,b] Despite extensive synthetic efforts,^[10] natural compounds remain the most widely used compounds from this structural class. The results of min-PKS screening suggested that our arrayed soil eDNA libraries contained multiple biosynthetic gene clusters that could potentially encode additional natural anthracyclines.

Cosmid clones associated with three KS_{β} sequence tags (AZ129, AZ515, and TX19) that clade with known anthracycline KS_{β} sequences were recovered from the eDNA libraries, sequenced, annotated, and then compared to known anthracycline biosynthetic gene clusters (Figure 2). Two of these clones (AZ129 and TX19) were found to contain gene clusters distinct from any previously described anthracycline gene clusters. The gene cluster captured on the third clone (AZ515) is identical in both gene content and organization to the nogalamycin gene cluster.^[11] Sequencing of the TX19 clone revealed two min-PKSs, one containing a KS_{β} gene that clades

[*] Dr. H.-S. Kang, Prof. S. F. Brady
Laboratory of Genetically Encoded Small Molecules
Howard Hughes Medical Institute, The Rockefeller University
1230 York Avenue, New York, NY 10065 (USA)
E-mail: sbrady@rockefeller.edu
Homepage: <http://lab.rockefeller.edu/brady/>

[**] This work was supported by NIH GM077516. S.F.B. is an HHMI Early Career Scientist. We thank Dr. José A. Salas for the generous gift of pRham.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201305109>.

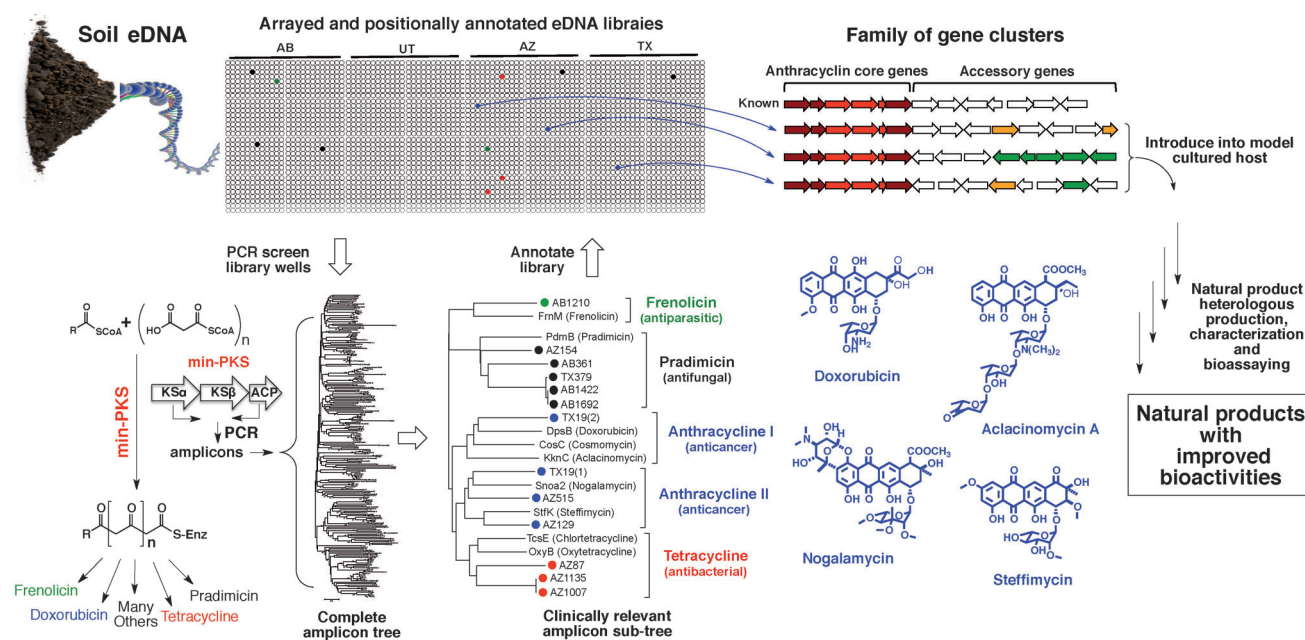


Figure 1. Sequence-guided discovery of clinically relevant polyketides from arrayed soil metagenomic libraries. DNA extracted directly from soil is used to construct saturating cosmid megalibraries (> 10 000 000 clones), which are arrayed as pools of unique 5000-membered sublibraries. Aromatic polyketide min-PKS gene cassettes are amplified in PCR reactions with degenerate primers using DNA from each unique sublibrary within the California (AB), Utah (UT), Arizona (AZ), and Texas (TX) megalibraries as template. The phylogenetic grouping of eDNA sequences with K_Sβ sequences from gene clusters that encode clinically relevant natural products is used to guide the recovery of clones containing gene clusters of interest from specific sublibraries. Heterologous expression of bioinformatically unique eDNA-derived gene clusters can be used to expand the structural diversity seen within known families of natural products and to potentially identify metabolites with improved activities.

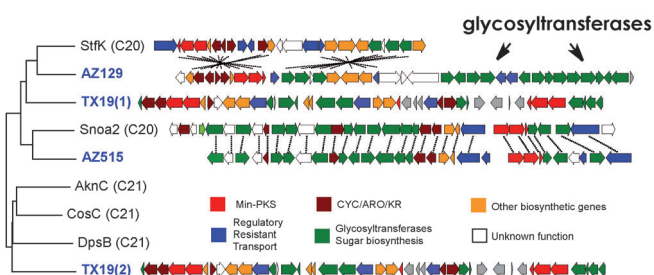


Figure 2. Anthracycline gene clusters. ClustalW phylogenetic tree of K_Sβ gene fragments from eDNA libraries (blue labels) and from culture-based studies (black labels). The gene clusters from clone AZ515 and clone AZ129 are compared to sequenced anthracycline gene cluster relatives (dashed lines connect functional homologues). The TX19 gene cluster does not closely resemble any anthracycline clusters deposited in public databases.

with type I (21-carbon core) anthracyclines (e.g. doxorubicin, cosmomycin, and aclacinomycin) and a second containing a K_Sβ gene with type II (20-carbon core) anthracyclines (e.g. nogalamycin and steffimycin) (Figure 2). No previously sequenced anthracycline gene clusters are predicted to encode for the biosynthesis of both type I and type II anthracycline core structures. The gene cluster captured on clone AZ129 is most closely related to the biosynthetic gene cluster encoding for the *S. steffisburgensis* derived steffimycin family of anthracyclines.^[12] However, it contains two additional glycosyltransferases as well as a set of genes predicted to encode for the biosynthesis of aminodeoxy sugars, which are not seen in known steffimycin structures (Figure 2).

The steffimycins are a small subgroup of type II anthracyclines (C20) that has gained only limited attention due to the poor antitumor activity of previously described steffimycin congeners.^[13] Extensive structure–activity relationship studies within the anthracycline family suggest that the sugar moiety plays a critical role in the observed antitumor activity of anthracyclines by interacting with the minor groove of DNA and topoisomerase II in a ternary complex.^[14] In fact, these studies suggest that the ternary complex interactions mediated by the sugar moiety and the fourth ring of the anthracycline core are likely more critical for determining antitumor potency than the intercalation of DNA by the anthraquinone portion of the anthracycline core. The weak biological activity observed for known steffimycins is, at least in part, likely due to the fact that all steffimycins reported to date are functionalized with rhamnose-like sugar moieties, while most clinically relevant anthracyclines contain aminodeoxy sugar derivatives (e.g. L-daunosamine and L-rhodosamine).^[13c] Owing to the importance of the sugar moiety for anthracycline bioactivity and the presence of the unprecedented set of genes for aminodeoxy sugar biosynthesis found on clone AZ129, we elected to study this gene cluster using heterologous expression methods.

To permit heterologous expression studies in *Streptomyces* spp., the AZ129 cosmid was retrofitted with the 6.8 kb DraI fragment from pOJ436 which contains an origin of transfer, an apramycin resistance gene, and the ΦC31 phage integration system. Retrofitted AZ129 was then transferred by intergenic conjugation into *S. albus* to generate *S. albus*/AZ129. In fermentation analyses this strain was found to

produce the known steffimycin aglycone biosynthetic intermediate, 8-demethoxy-10-deoxysteffermycinone (compound **4**).^[12] The production of an aglycone was not unexpected as AZ129, like the steffimycin cluster, is missing two key genes (dNDP-D-glucose synthase and a dNDP-D-glucose-4,6-dehydratase) that are required to initiate aminodeoxy sugar biosynthesis. These sugar biosynthesis genes appear in the primary metabolic background of some, but not all, *Streptomyces* spp.^[15] The early genes required for deoxy sugar biosynthesis were therefore provided in trans (pIJRham) by cloning the rhamnose biosynthetic operon (oleL, oleS, oleE, and oleU)^[16] from oleandromycin biosynthesis under the control of the constitutive ermE* promoter within the conjugative integrative shuttle expression vector pIJ10257.^[17] Conjugation of this construct into *S. albus*/AZ129 yielded *S. albus*/AZ129/pIJRham.

LCMS analysis of organic extracts obtained from cultures of *S. albus*/AZ129/pIJRham showed the presence of three major clone-specific metabolites with masses suggesting the successful addition of either one ($M_w = 544$ and 558) or two ($M_w = 756$) sugars onto the anthracycline aglycone core. These three metabolites were isolated by C_{18} reversed-phase HPLC from large-scale ethyl acetate extracts and named arimetamycins A–C (**1–3**: 2.5 mg L^{-1} for **1**, 0.5 mg L^{-1} for **2** and 3 mg L^{-1} for **3**). The structures of **1–3** (Figure 3) were determined by HRESIMS, and 1D and 2D NMR analysis including ^1H , ^{13}C , COSY, TOCSY, HMQC, HMBC, and ROESY NMR experiments (Discussions 1 and 2, Figures 1 and 2, and Tables 1 and 2 in the Supporting Information).

The arimetamycins feature a 20-carbon (type II) anthracycline core (compound **4**, Figure 4) with two distinct glycosylation patterns (Figure 3 a). The dissaccharide arimetamycin A (**1**) contains two rare *N,N*-dimethyl aminodeoxy

sugar moieties, brasilirose and lemonose. To the best of our knowledge, the lemonomycin^[18] and nocathiacin^[19] families are the only natural products reported to date that contain a lemonose sugar moiety. All previously reported anthracyclines in the steffimycin family are functionalized with a single rhamnose-based sugar moiety.^[13c] The monoglycosylated arimetamycins B (**2**) and C (**3**) contain *L*-olivose-based sugars which differ from *L*-rhamnose by the absence of the C-2' hydroxy group. Arimetamycin B (**2**) contains an actual *L*-olivose sugar while arimetamycin C (**3**) contains a methylated version of this sugar, 3'-OMe-*L*-olivose.

Arimetamycins A–C (**1–3**) were evaluated for anticancer activity against representative colon, lung, and breast cancer cell lines. The lung cancer cell line (H69AR) is multidrug resistant (MDR), showing high-level resistance to anthracyclines (50-fold) as well as many other anticancer agents including vincristine, etoposide, and mitoxantrone (Figure 3 b).^[20] The disaccharide-functionalized arimetamycin A (**1**) exhibits nanomolar IC_{50} values against all cell lines including the H69AR MDR cell line. The monosaccharide-functionalized arimetamycins B (**2**) and C (**3**) show only moderate activity against these cell lines (micromolar IC_{50}), which is comparable to the activity reported for previously described anthracyclines in the steffimycin family.^[21] IC_{50} values obtained for arimetamycin A are consistently lower (2–50 fold) than those observed for either doxorubicin or daunorubicin, and unlike doxorubicin or daunorubicin, arimetamycin A is highly active against both MDR and drug-sensitive cell lines. P-glycoprotein expression-dependent MDR is the most clinically significant resistance mechanism affecting the utility of anthracyclines.^[9a,b] The sugar moiety has been shown to play a key role in anthracycline bioactivity, in particular *N*-alkylation of the sugar has proved important.^[22] Arimetamycin A provides a new *N*-alkylated lead structure for use in the potential development of therapies for MDR cancers.

The biosynthesis of the arimetamycins can be rationalized on the basis of the predicted functions for the genes found within the AZ129 clone (Figure 4 a and Figure 3 in the Supporting Information). In our proposed biosynthetic scheme for the arimetamycins, the type II (C20) anthracycline **4** is predicted to be synthesized by enzymes encoded by nine genes (*arm1–9*) that show a high sequence identity (67–85%) to their counterparts from the steffimycin gene cluster.^[12] In this biosynthetic proposal, intermediate **4** is predicted to enter into two distinct glycosylation pathways (Figure 4 c) resulting in the formation of both the dissaccharide-containing **1** and the monosaccharide-containing **2** and **3**. The tentative assignment of glycosyl transfer specificities shown in our biosynthetic scheme is based on the high sequence identity observed between *arm12* and the steffimycin glycosyltransferase (*stfG*) and the position of the *arm25/32* glycosyl transfer genes in the sugar biosynthesis cassette. As with all heterologous expression based discovery studies, the genetic background, in this case *S. albus* expressing rhamnose biosynthesis, may play an unrecognized role in the generation of the specific compounds observed in these studies.

The structures of compounds **1–3** together with the overall organization of the *arm* cluster and the proposed branched

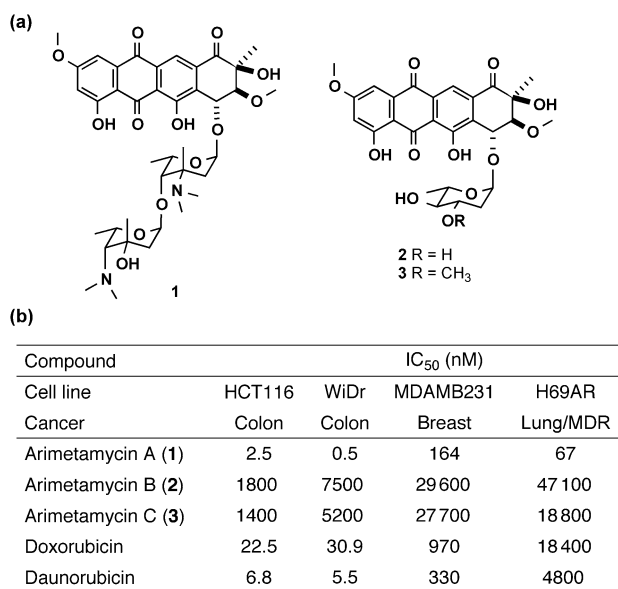


Figure 3. Structure and activity of the arimetamycins. a) Chemical structures of arimetamycins A–C (**1–3**). Relative configuration is shown only for the aglycone moiety. b) Cytotoxicity assay against four human cancer cell lines. Previously reported IC_{50} values for steffimycin against the MDAMB231 and HT-29 (WiDr) cell lines are 4870 and 6790 nM, respectively.^[21]

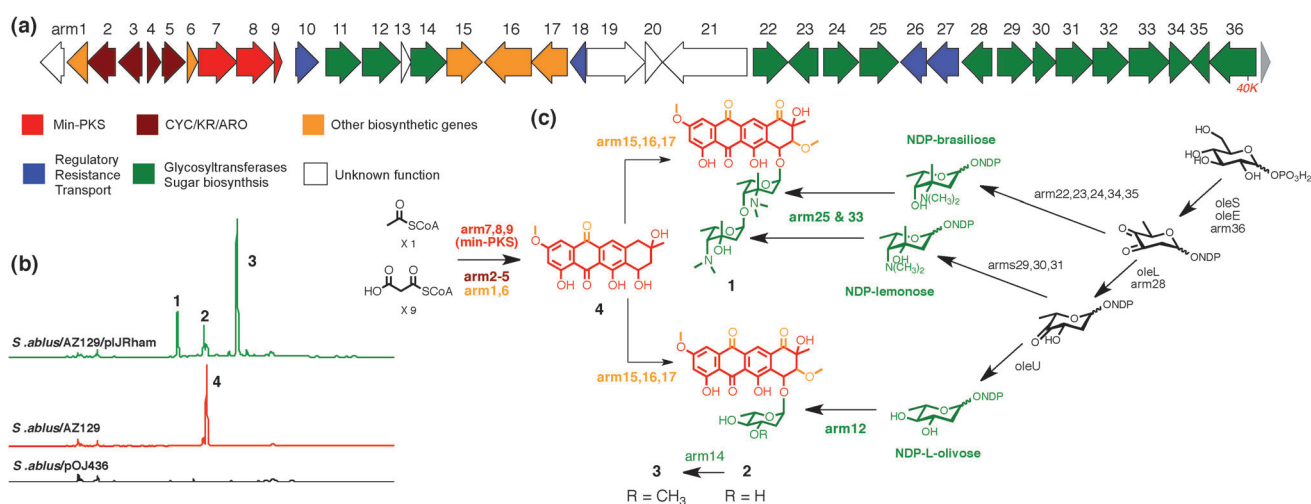


Figure 4. Heterologous production and proposed biosynthesis of the arimetamycins based on gene function prediction by sequence homology and domain analysis. a) The *arm* (arimetamycin) gene cluster spanning roughly 40 kb and containing 36 predicted ORFs (GenBank KF040454). b) HPLC analysis of fermentation broth extracts showing arimetamycins A–C production by *S. albus/AZ129/pIJRhams*. c) Proposed biosynthetic scheme for the arimetamycins based on gene prediction with the aid of Pfam domain analysis. See Figure 3 in the Supporting Information for an expanded biosynthesis scheme and Table 4 in the Supporting Information for a complete *arm* cluster gene table.

biosynthetic scheme suggest that the genetic instructions for the biosynthesis and addition of the two aminodeoxy sugars seen in arimetamycin A were likely acquired at some point by an existing monosaccharide-generating anthracycline biosynthetic gene cluster that is closely related to the known steffimycin gene cluster (Figure 4).^[12] Mimicking this natural evolutionary phenomenon recombinantly by transferring the *arm*-derived aminodeoxy sugar biosynthesis cassette to other anthracycline producers could prove useful for improving the activity of other compounds in this family as well.

Resistance mechanisms related to many of those found in human healthcare settings, like MDR efflux pumps, are also seen in lower eukaryotes and bacteria.^[23] The ultimate consequence of a natural microbial small-molecule “arms race” may be a situation where resistance and molecules capable of circumventing resistance evolve hand in hand. In this scenario, the *arm* gene cluster, in particular the aminodeoxy sugar biosynthesis cassette (*arm22-36*) would arise in response to naturally occurring MDR-like resistance mutations. The systematic screening of eDNA libraries for analogues of clinically relevant families of natural products should prove useful for identifying additional compounds capable of overcoming clinically relevant resistance mechanisms.

The work presented here demonstrates that the pool of biosynthetic gene clusters captured in saturating eDNA soil libraries is large enough to permit the directed discovery of gene clusters that encode for metabolites with improved activity compared to that of known biomedically relevant natural products. The sequence-tag-guided gene cluster discovery approach we outline enables the rapid profiling of chemical diversity encoded by gene clusters previously hidden in microbial communities and the subsequent prioritization of these gene clusters for heterologous expression efforts. The strong correlation observed between individual biosynthetic gene phylogeny and the chemical structures of natural

products associated with these sequences is not unique to aromatic polyketide biosynthesis. In fact, this phenomenon is likely to be seen universally throughout natural product biosynthesis as a result of functionally related gene clusters having common evolutionary ancestors. The approach outlined here should therefore be generically useful for rapidly assessing nature’s ability to encode for metabolites capable of improving the efficacy of diverse families of biomedically relevant natural products.

Received: June 13, 2013

Published online: September 3, 2013

Keywords: anthracyclines · antitumor agents · MDR · metagenomics · natural products

- [1] a) V. Torsvik, L. Ovreas, *Curr. Opin. Microbiol.* **2002**, *5*, 240–245; b) M. S. Rappé, S. J. Giovannoni, *Annu. Rev. Microbiol.* **2003**, *57*, 369–394.
- [2] K. Scherlach, C. Hertweck, *Org. Biomol. Chem.* **2009**, *7*, 1753–1760.
- [3] a) J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy, R. M. Goodman, *Chem. Biol.* **1998**, *5*, R245–249; b) A. Harvey, *Drug Discovery Today* **2000**, *5*, 294–300; c) R. Daniel, *Curr. Opin. Biotechnol.* **2004**, *15*, 199–204; d) C. Schmeisser, H. Steele, W. R. Streit, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 955–962; e) V. Miao, J. Davies in *Uncultivated Microorganisms*, Vol. 10 (Ed.: S. S. Epstein), Springer, Berlin, **2009**, pp. 217–236; f) H. A. Iqbal, Z. Feng, S. F. Brady, *Curr. Opin. Chem. Biol.* **2012**, *16*, 109–116.
- [4] a) J. Zhou, M. A. Bruns, J. M. Tiedje, *Appl. Environ. Microbiol.* **1996**, *62*, 316–322; b) S. Courtois, C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helyncck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osborne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, J. L. Pernodet, *Appl. Environ. Microbiol.* **2003**, *69*, 49–55; c) H. Bertrand, F. Poly, V. T. Van, N. Lombard, R. Nalin, T. M. Vogel, P. Simonet, *J. Microbiol.*

- Methods* **2005**, *62*, 1–11; d) S. F. Brady, *Nat. Protoc.* **2007**, *2*, 1297–1305.
- [5] a) 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*, 5522–5527; b) K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzter, B. K. Rubio, S. J. Robinson, P. Crews, J. Piel, *Nat. Chem. Biol.* **2009**, *5*, 494–501; c) R. W. King, J. D. Bauer, S. F. Brady, *Angew. Chem.* **2009**, *121*, 6375–6379; *Angew. Chem. Int. Ed.* **2009**, *48*, 6257–6261.
- [6] a) C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, *Nat. Prod. Rep.* **2007**, *24*, 162–190; b) C. Hertweck, *Angew. Chem.* **2009**, *121*, 4782–4811; *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–4716.
- [7] a) C. P. Ridley, H. Y. Lee, C. Khosla, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4595–4600; b) B. V. Reddy, D. Kallifidas, J. H. Kim, Z. Charlop-Powers, Z. Feng, S. F. Brady, *Appl. Environ. Microbiol.* **2012**, *78*, 3744–3752.
- [8] K. T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, J. Davies, *J. Bacteriol.* **1997**, *179*, 7360–7368.
- [9] a) G. Minotti, P. Menna, E. Salvatorelli, G. Cairo, L. Gianni, *Pharmacol. Rev.* **2004**, *56*, 185–229; b) C. Monneret, *Eur. J. Med. Chem.* **2001**, *36*, 483–493; c) A. Grimm, K. Madduri, A. Ali, C. R. Hutchinson, *Gene* **1994**, *151*, 1–10.
- [10] a) P. Vasey, *Int. J. Oncol.* **1993**, *2*, 677–681; b) R. B. Weiss, G. Sarosy, K. Clagett-Carr, M. Russo, B. Leyland-Jones, *Cancer Chemother. Pharmacol.* **1986**, *18*, 185–197.
- [11] S. Torkkell, T. Kunnari, K. Palmu, P. Mantsala, J. Hakala, K. Ylihonko, *Mol. Genet. Genomics* **2001**, *266*, 276–288.
- [12] S. Gullon, C. Olano, M. S. Abdelfattah, A. F. Brana, J. Rohr, C. Mendez, J. A. Salas, *Appl. Environ. Microbiol.* **2006**, *72*, 4172–4183.
- [13] a) P. F. Wiley, D. W. Elrod, V. P. Marshall, *J. Org. Chem.* **1978**, *43*, 3457–3461; b) Y. Matsuzawa, T. Oki, T. Takeuchi, H. Umezawa, *J. Antibiot.* **1981**, *34*, 1596–1607; c) H. Laatsch, S. Fotso, *Top. Curr. Chem.* **2008**, *282*, 3–74.
- [14] F. Zunino, G. Pratesi, P. Perego, *Biochem. Pharmacol.* **2001**, *61*, 933–938.
- [15] E. Rodriguez, H. G. Menzella, H. Gramajo, *Methods Enzymol.* **2009**, *459*, 339–365.
- [16] L. Rodriguez, C. Oelkers, I. Aguirrezabalaga, A. F. Brana, J. Rohr, C. Mendez, J. A. Salas, *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 271–276.
- [17] H. J. Hong, M. I. Hutchings, L. M. Hill, M. J. Buttner, *J. Biol. Chem.* **2005**, *280*, 13055–13061.
- [18] H. He, B. Shen, G. T. Carter, *Tetrahedron Lett.* **2000**, *41*, 2067–2071.
- [19] Y. Ding, Y. Yu, H. Pan, H. Guo, Y. Li, W. Liu, *Mol. Biosyst.* **2010**, *6*, 1180–1185.
- [20] S. E. Mirski, J. H. Gerlach, S. P. Cole, *Cancer Res.* **1987**, *47*, 2594–2598.
- [21] C. Olano, M. S. Abdelfattah, S. Gullon, A. F. Brana, J. Rohr, C. Mendez, J. A. Salas, *ChemBioChem* **2008**, *9*, 624–633.
- [22] a) G. Capranico, R. Supino, M. Binaschi, L. Capolongo, M. Grandi, A. Suarato, F. Zunino, *Mol. Pharmacol.* **1994**, *45*, 908–915; b) A. Schaefer, J. Westendorf, K. Lingelbach, C. A. Schmidt, D. L. Mihalache, A. Reymann, H. Marquardt, *Cancer Chemother. Pharmacol.* **1993**, *31*, 301–307; c) L. Fang, G. Zhang, C. Li, X. Zheng, L. Zhu, J. J. Xiao, G. Szakacs, J. Nadas, K. K. Chan, P. G. Wang, D. Sun, *J. Med. Chem.* **2006**, *49*, 932–941.
- [23] M. H. Saier, Jr., I. T. Paulsen, M. K. Sliwinski, S. S. Pao, R. A. Skurray, H. Nikaido, *FASEB J.* **1998**, *12*, 265–274.