Natural Products

DOI: 10.1002/anie.201103085

Transcriptome Mining of Active Biosynthetic Pathways and Their Associated Products in *Streptomyces flaveolus***

*Xudong Qu, Chun Lei, and Wen Liu**

Natural products (NPs) are small molecules of incredible structural diversity that have long been appreciated for their critical role in drug discovery and development.[1] Production of NPs from biosynthetic gene clusters depends on the coordinated transcription of relevant genes to messenger RNAs (mRNAs), the translation of these mRNAs to polypeptide chains, and the correct folding of these polypeptides to create functional biosynthetic machineries that cooperate to catalyze the formation of complex structures from simple precursor molecules (Figure 1). Sequencing of numerous actinomycete genomes during the past decade has revealed a stunning number of NP biosynthetic gene clusters, only about 10% of which have been linked to characterized NPs.[2] While bioinformatics-guided efforts have been successful in correlating some of these NP gene clusters to new metabolites,[3] which include those produced by polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), and terpene synthases,[4] it is clear that our ability to mine these new gene clusters to uncover the chemical potential hidden within them has not kept pace with DNA sequencing technology. In the postgenomic era, strategies that target different stages of NP production have been developed to validate the genetic basis that is relevant to chemical biosynthesis.[3–10] We used *S. flaveolus* DSM 9954 as a model system,[11] and report herein the mining of active biosynthetic pathways and their associated products at the transcriptional level, at which a key linkage of transforming the potential genotype to the practical chemotype exists. This effective strategy, which uses mRNA-based mining, has the potential to be generally applied to increase the efficiency of NP discovery after further development, even from strains for which the genome sequence is unknown.

We began by evaluating the chemical profile of *S. flaveolus* in different fermentation media. Six media, which have been used previously in our laboratory to produce known metabolites (see methods in the Supporting Information), were assessed individually in four-day fermentations. The culture broths were analyzed by HPLC, and showed remarkable differences in product profiles (see Figure S1 in the Supporting Information). Medium V, in which *S. flaveolus* produced the greatest variety of metabolites, was selected for a time-course analysis of the production (see Figure S2 in the Supporting Information). As the highest yields of most metabolites were observed on day 4, this point was selected for transcriptome analysis. Total RNA isolated from *S. flaveolus* mycelia on day 4 was converted to complementary DNA (cDNA), and thus constitutes a conditional transcriptome (see Figure S3a in the Supporting Information).

To explore the biosynthetic potential of *S. flaveolus* under these conditions, the resulting cDNAs were screened by PCR by using primers corresponding to a variety of biosynthetic genes that produce diverse structural motifs (see methods and Figure S4 in the Supporting Information). PCR products generated by degenerate primers varied in length from 280 to 550 bps and included those for ketoacylsynthase (KS) domain of modular type I PKS to assemble the polyketide carbon backbone,[12a] adenylation (A) domain of linear NRPS to activate the amino acid substrate,[12b] thiopeptide cyclodehydrase (TCDH) to afford thiazoline moieties on a ribosomal peptide,[12c] (d)NDP-D-glucose-4,6-dehydrase (NGDH) involved in the deoxysugar formation,[12d] and FAD-dependent halogenase (FDH) responsible for incorporation of a halogen atom.[12e] Whereas amplifications of *tcdh*, *dgdh*, and *fdh* failed to give products, distinct PCR products of *pks* and *nrps* were amplified from *S. flaveolus* cDNAs grown in medium V (see Figure S3b in the Supporting Information). PCR products were cloned into plasmids and transformed into *E. coli*. Fifteen colonies from both the *pks* and *nrps* populations were randomly selected for sequencing, and all sequenced products displayed homology to known genes of PKS or NRPS. The products were subsequently grouped, according to sequence identity, into seven *pks*s (*pcr1–7*) and three *nrps*s (*pcr8–10*), respectively. Among them, two *pks*s (*pcr6* and *pcr7*) and two *nrps*s (*pcr9* and *pcr10*) are identical to the counterparts within the newly identified sanglifehrin (SF) gene cluster,[11] which includes the sequences encoding KS domains of PKSs in module 3 of SfaF (for *pcr6*) and in module 6 of SfaG (for *pcr7*), and sequences encoding A domains in module 14 (for *pcr9*) and in module 15 (for *pcr10*) of NRPS SfaD. *S. flaveolus* DSM 9954 is a known producer of SFs (represented by SF-A, Figure 1),[13] which are a class of hybrid polyketide–polypeptide NPs with immunosuppressive activity, the biosynthesis of which involves a hybrid PKS–NRPS system to incorporate a number of unusual PKS and NRPS building blocks.[11] The presence of *sf* mRNAs in cultures based on medium V is supported by the

[*] Dr. X. Qu, Dr. C. Lei, Prof. Dr. W. Liu
State Key Laboratory of Bioorganic and Natural Products Chemistry
Shanghai Institute of Organic Chemistry
Chinese Academy of Sciences
345 Lingling Rd., Shanghai 200032 (China)
E-mail: wliu@mail.sioc.ac.cn

[**] This work was supported in part by grants from NNSF of China (30900021, 20902102, 20921091, and 20832009), SRF for 39th ROCS-SEM, "973 program" (2010CB833200, and 2012CB721100), CAS (KJCX2-YW-201), and STCSM (09QH1402700). We thank Dr. M. J. Smanski from the University of Wisconsin–Madison for discussion and critical reading.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201103085.

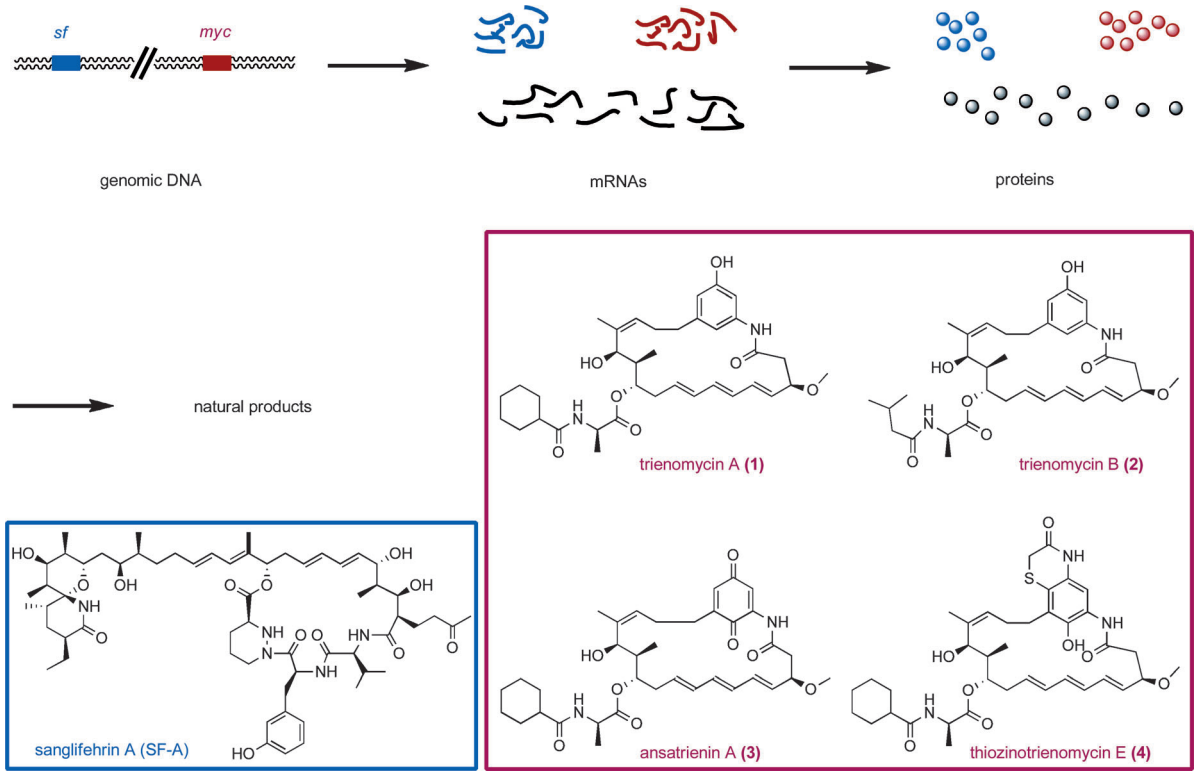

Figure 1. Multiple stages are required for converting genotype into chemotype in NP production. The example shown here is that from *S. flaveolus*, which produces sangifehrins (blue, shown by the representative SF-A), and MYCs such as compounds **1**, **2**, **3**, and **4** (red).

identification of SFs in the culture broth by HPLC–MS analysis (see Figure S5 in the Supporting Information). In contrast, the remaining five *pks*s (*pcr1–5*) and one *nrps* (*pcr8*) are new. Given that no counterparts were found in the published database, they may encode the biosynthesis of previously unknown polyketides, polypeptides, or their hybrids in *S. flaveolus* DSM 9954.

The effectiveness of this transcriptome-based approach was compared to traditional PCR-dependent genome screening methods by repeating the above-mentioned procedure with genomic DNA, which replaces the cDNA template (see Figure S3b in the Supporting Information). Similary, no product was obtained from the *tcdh* and *fdh* reactions, thus suggesting that these genes are either not present in the *S. flaveolus* genome or have diverged sufficiently to prevent annealing of the degenerate primers. Amplification of *ngdh* was successful, thus indicating that the gene exists but is not expressed in the fermentation conditions based on medium V. Lastly, consistent with our initial attempts to clone the SF gene cluster,[11] diverse genes of PKS (11) and NRPS (11) were observed from the PCR products by sequencing 15 random clones in each case, and only one (*pcr6* for *pks* also identified by cDNA amplification) was found within the SF biosynthetic gene cluster. None of the other genes (21) are similar to the new products from cDNA amplification, thus suggesting that the PKS and NRPS genes identified in the transcriptome produced in medium V were selectively enriched in cDNA compared to genomic DNA. These results demonstrate a significant advantage in efficiency that arises from mining biosynthetic genes at the transcriptional level as opposed to the genetic level.

To establish the relationship of *pks*s *pcr1–5* and *nrps pcr8* to unknown compounds, each gene was systematically inactivated in *S. flaveolus*, and the crude extracts generated from culturing the resulting mutant strains were analyzed by HPLC (see Figure S6 in the Supporting Informaton). Intriguingly, all mutant strains shared a common phenotype under the conditions based on medium V. A set of metabolites could be clearly identified in both wild-type *S. flaveolus* DSM 9954 and in the *sf* gene cluster deleted mutant TL3001;[11] crude extracts were entirely absent, while the production of SFs was maintained (Figure 2a). The results strongly suggest that these metabolites are a group of biosynthetically related, hybrid polyketide–polypeptide NPs. To confirm this hypothesis, we fermented *S. flaveolus* and selectively isolated the metabolites absent from the mutant strains for structural elucidation (see results in the Supporting Information). Consequently, the four major products (compounds **1**–**4**, Figure 1) were characterized on the basis of spectral analysis as the mycotrienin (MYC) members, trienomycin A (**1**), trienomycin B (**2**), ansatrinenin A (**3**), and thiozinotrienomycin E (**4**), all of which belong to a large ansamycin family, structurally feature a 3-amino-5-hydroxybenzoic acid (AHBA) derived moiety in an aminoacylated macrolactam ring, and possess potent antitumor and antibacterial properties.[14]

The biosynthetic gene cluster responsible for production of these MYCs was cloned from a genomic DNA library by

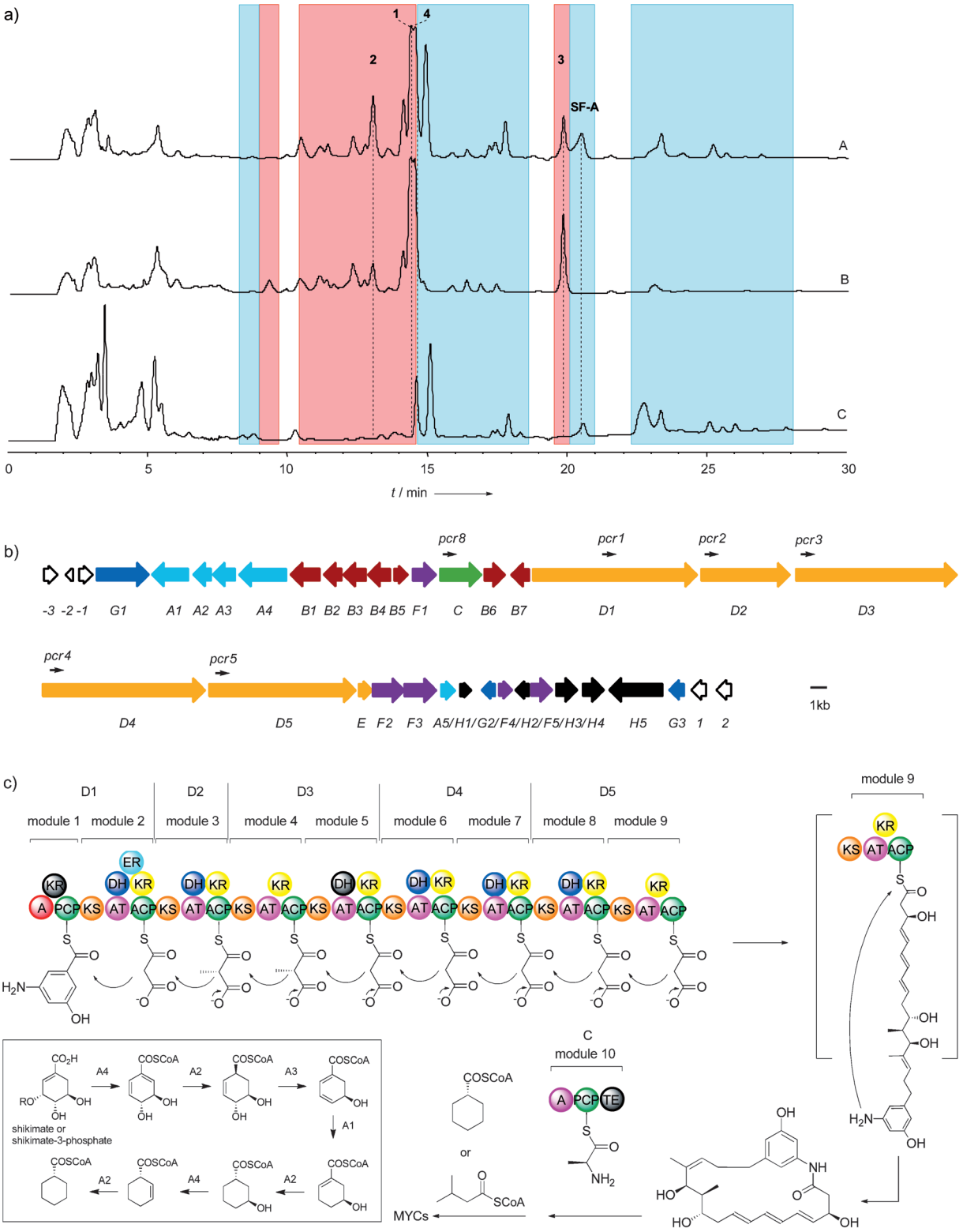

Figure 2. Discovery of MYCs along with the *myc* gene cluster (GenBank accession number JF803483) and proposal for the biosynthetic pathway. a) Production of SFs and MYCs in *S. flaveolus*. HPLC analysis of the culture broths of the wild-type strain based on medium V (A), SF (shaded in blue) nonproducing mutant strain *TL3001* (B), and MYC (shaded in purple) nonproducing mutant strain *TL3016* (C). b) Organization of the MYC biosynthetic genes. Their deduced functions include PKS (yellow), NRPS (green), AHBA (red), CHC (light blue), regulation (dark blue), tailoring/edition (purple), and unknown proteins (black). The cDNA-based PCR products (*pks1–5* and *nrps8*) are shown at the corresponding positions. c) Hybrid PKS–NRPS system for assembly of the macrolactam ring and incorporation of L-Ala during MYC biosynthesis. Domains shown in black are predicted to be inactive. A = adenylation, PCP = peptidyl carrier protein, KS = ketoacylsynthase, AT = acyltransferase, ACP = acyl carrier protein, DH = dehydrase, KR = ketoreductase, TE = thioesterase.

using *pcr1–5* and *pcr8* as probes. This approach led to identification of a contiguous sequence of 90709 bp that harbors the *myc* gene cluster comprised of 32 open reading frames (ORFs; Figure 2b; see also Table S3 and Figures S8 and S9 in the Supporting Information). Counterparts of the probes were identified and included sequences that encode KS domains of PKSs in module 2 of MycD1 (for *pcr1*), in module 3 of MycD2 (for *pcr2*), in module 4 of MycD3 (for *pcr3*), in module 6 of MycD4 (for *pcr4*), and in module 8 of MycD5 (for *pcr5*), and the sequence that encodes an A domain of NRPS MycC (for *pcr8*). Bioinformatics-based analysis confirmed that a PKS–NRPS hybrid system is indeed involved in the biosynthesis of these MYCs (see Figure 2c and the results in the Supporting Information). As shown for trienomycin A: 1) the modular type I PKSs MycD1–D5, which use AHBA (formed by MycB1–B7) as the starter unit, could catalyze eight decarboxylative condensations in a linear manner to assemble an octaketide intermediate, along with an intramolecular lactonization to close the macrocyclic ring; and 2) the NRPS MycC might be involved in incorporating an L-Ala residue at C-11, followed by attachment of a cyclohexanecarboxylic acid (CHC) moiety (catalyzed by MycA1–A4) to eventually afford **1**. Partial genes, including the cassette for the formation of CHC, which has been incorporated into the avermectin biosynthetic machinery to produce new antiparasitic agents doramectins by applying the principle of synthetic biology,[15] were previously found in *S. collinus*, a strain that produces compound **1**;[16] however, the entire gene cluster has not been reported. To further confirm the role of the *myc* gene cluster, an internal fragment that encompasses *mycA1–mycD2* in the cluster was deleted in *S. flaveolus*, thus completely abolishing MYC production in the resulting mutant (see Figure S6 in the Supporting Information). These results demonstrate a second advantage of the cDNA approach in expediting the identification of active biosynthetic pathways that correlate to unknown NPs.

In conclusion, we have developed a new transcriptome-screening strategy for exploring NP production as well as for accessing the biosynthetic pathways in NP producers for which the genome sequence is unknown. The approach is rapid and can be tailored to suit specific interests, such as altering culture conditions to activate the production of certain metabolites or designing new primers for PCR amplification of unique biosynthetic elements. This approach can be easily integrated into available genome mining strategies for awakening cryptic gene clusters,[7–10] and potentially facilitate the screening for NPs and their biosynthetic investigations.

Received: May 5, 2011
Revised: July 30, 2011

Keywords: biosynthesis · natural products · nonribosomal peptide synthetase · polyketide synthase · transcriptomes

[1] D. J. Newman, G. M. Cragg in *Natural Product Chemistry for Drug Discovery* (Eds.: A. D. Buss, M. S. Butler), Royal Society of Chemistry, Cambridge, **2010**, pp. 3–27.
[2] M. Nett, H. Ikeda, B. S. Moore, *Nat. Prod. Rep.* **2009**, *26*, 1362–1384.
[3] C. Corre, G. L. Challis, *Nat. Prod. Rep.* **2009**, *26*, 977–986.
[4] a) L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis, B. Aigle, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6258–6263; b) S. Bergmann, J. Schumann, K. Scherlach, C. Lange, A. A. Brakhage, C. Hertweck, *Nat. Chem. Biol.* **2007**, *3*, 213–217; c) C. Wang, D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 8908–8909.
[5] a) S. B. Bumpus, B. S. Evans, P. M. Thomas, I. Ntai, N. L. Kelleher, *Nat. Biotechnol.* **2009**, *27*, 951–956; b) B. S. Evans, I. Ntai, Y. Chen, A. J. Robinson, N. L. Kelleher, *J. Am. Chem. Soc.* **2011**, *133*, 7316–7319.
[6] C. I. Keeling, S. Weisshaar, S. G. Ralph, S. Jancsik, B. Hamberger, H. K. Dullat, J. Bohlmann, *BMC Plant Biol.* **2011**, *11*, 43.
[7] E. Zazopoulos, K. Huang, A. Staffa, W. Liu, B. O. Bachmann, K. Nonaka, J. Ahlert, J. S. Thorson, B. Shen, C. M. Farnet, *Nat. Biotechnol.* **2003**, *21*, 187–190.
[8] J. W. Bok, Y.-M. Chiang, E. Szewczyk, Y. Reyes-Dominguez, A. D. Davidson, J. F. Sanchez, H.-C. Lo, K. Watanabe, J. Strauss, B. R. Oakley, C. C. C. Wang, N. P. Keller, *Nat. Chem. Biol.* **2009**, *5*, 462–464.
[9] K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews, J. Piel, *Nat. Chem. Biol.* **2009**, *5*, 494–501.
[10] T. Hosaka, M. Ohnishi-Kameyama, H. Muramatsu, K. Murakami, Y. Tsurumi, S. Kodani, M. Yoshida, A. Fujie, K. Ochi, *Nat. Biotechnol.* **2009**, *27*, 462–464.
[11] X. Qu, N. Jiang, F. Xu, L. Shao, G. Tang, B. Wilkinson, W. Liu, *Mol. BioSyst.* **2011**, *7*, 852–861.
[12] a) J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380–416; b) M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651–2674; c) R. Liao, L. Duan, C. Lei, H. Pan, Y. Ding, Q. Zhang, D. Chen, B. Shen, Y. Yu, W. Liu, *Chem. Biol.* **2009**, *16*, 141–147; d) H. Decker, S. Gaisser, S. Pelzer, P. Schneider, L. Westrich, W. Wohlleben, A. Bechthold, *FEMS Microbiol. Lett.* **1996**, *141*, 195–201; e) A. Hornung, M. Bertazzo, A. Dziarnowski, K. Schneider, K. Welzel, S. Wohlert, M. Holzenkämpfer, G. J. Nicholson, A. Bechthold, R. D. Süssmuth, A. Vente, S. Pelzer, *ChemBioChem* **2007**, *8*, 757–766.
[13] J. J. Sanglier, V. Quesniaux, T. Fehr, H. Hofmann, M. Mahnke, K. Memmert, W. Schuler, G. Zenke, L. Gschwind, C. Maurer, W. Schilling, *J. Antibiot.* **1999**, *52*, 466–473.
[14] a) S. Funayama, K. Okada, K. Iwasaki, K. Komiyama, I. Umezawa, *J. Antibiot.* **1985**, *38*, 1677–1683; b) M. Damberg, P. Russ, A. Zeeck, *Tetrahedron Lett.* **1982**, *23*, 59–62; c) N. Hosokawa, H. Naganawa, H. Inuma, M. Hamada, T. Takeuchi, T. Kanbe, M. Hori, *J. Antibiot.* **1995**, *48*, 471–478.
[15] T. A. Cropp, D. J. Wilson, K. A. Reynolds, *Nat. Biotechnol.* **2000**, *18*, 980–983.
[16] S. Chen, D. v. Bamberg, V. Hale, M. Breuer, B. Hardt, R. Muller, H. G. Floss, K. A. Reynolds, E. Leistner, *Eur. J. Biochem.* **1999**, *261*, 98–107.

