The Multiplasmid Approach: A New Perspective for Combinatorial Biosynthesis

Christian Hertweck*[a]

KEYWORDS:

biotransformations · epothilones · gene technology · medicinal chemistry · polyketides

The investigation and manipulation of nature's vast biosynthetic potential with the tools of molecular biology, genetics, and chemistry has initiated an exciting new era for drug discovery.[1] At present, the prime targets for combinatorial biosynthesis are polyketides, due to their structural complexity and their large variety of pharmacological properties.[2] As a consequence of fundamental advances in genetic engineering of bacterial polyketide biosynthetic genes, it is now possible to generate structurally diverse libraries that are virtually unavailable by conventional synthetic methods.[3] Polyketides are constructed by repetitive Claisen condensations of (methyl)malonyl-coenzyme A [methyl(malonyl)-CoA] units with an acyl-CoA starter unit, in a manner that closely parallels fatty acid biosynthesis. On the basis of both genetic studies and architecture of the enzymes, bacterial polyketide synthases (PKSs) are generally classified into two types.[4] The aromatic type II system comprises a set of iteratively used individual enzymes that generate poly(β-keto) intermediates and catalyze cyclodehydratations to yield aromatic compounds, for example actinorhodin. In contrast, the modular type I system consists of large multifunctional proteins that accommodate distinct active sites for each step of catalysis. An example of a type I PKS product is 6-deoxyerythronolide B (6-dEB, **1**), the parent, nonglycosylated macrolide core of the antibiotic erythromycin. The linear, modular organization of the type I PKS catalytic domains programs the structure of the resulting polyketide and thus makes it predictable. Structural diversity of the polyketide family originates from a variety of biosynthetic building blocks and catalytic domains. First, the choice of starter unit, the number of C_2 extensions, and the mode of cyclization define the nature of the polyketide backbone. Second, the nature of the acyltransferase and the degree of β-oxo group processing (ketoreduction, dehydratation, and enoyl reduction) give rise to alkyl branches, keto, hydroxy, methenyl, or methylene functionalities (Figure 1). Finally, after the polyketide has been formed, tailoring enzymes may catalyze post-PKS transformations of the polyketide such as glycosylations, (ep)oxidations, acylations, and methylations.[1–5]

The tremendous variety of polyketides in nature[2a] is governed by many possible combinations and permutations of functional subunits, as well as by the sequence and number of modules. In a similar fashion, genetic engineering allows modifications of polyketide structures by a variety of targeted genetic

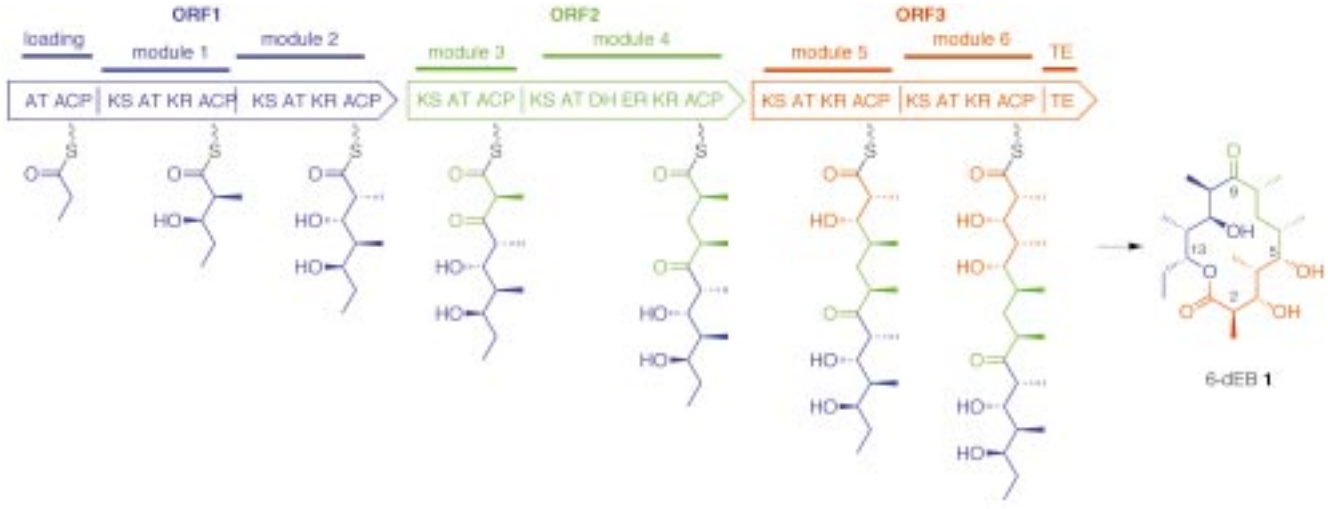

Figure 1. *Modular arrangement of a type I PKS (DEBS) that produces 6-deoxyerythronolide B (6-dEB, **1**). ACP = acyl carrier protein, AT = acyltransferase, DEBS = 6-deoxyerythronolide B synthase, DH = dehydratase, ER = enoylreductase, KR = ketoreductase, KS = ketosynthase, TE = thioesterase.*

[a] *Dr. C. Hertweck*
University of Washington, Seattle
Department of Chemistry, Box 351700
Seattle, WA 98185-1700 (USA)
Fax: (+ 1) 206-543-8318
E-mail: hertweck@u.washington.edu

  1439-4227/00/01/02 $ 17.50+.50/0

mutations, such as inactivation, substitution, and addition of catalytic domains or even entire modules.[6] As numerous studies revealed, modular PKSs are in fact amenable to a variety of modifications and display tolerance towards nonnatural PKS substrates.[7] Recently, it has become possible to conveniently express functional hybrid PKSs in genetically engineered bacterial hosts (i.e. *Streptomyces coelicolor* CH999[8] or *S. lividans* K4-114[9]), which lack a background of polyketide biosynthetic genes. Genetic manipulations of polyketide synthases are thus considerably promising for the biosynthesis of novel hybrid polyketides. On the basis of a single-plasmid expression system, however, for each new polyketide, specifically targeted mutagenesis experiments have to be performed prior to expression. The number of novel engineered polyketides thus does not exceed the number of separate engineering steps. In addition, considering the large size of the expression plasmids (ca. 50 kb for 6-dEB synthase, DEBS), and the limited number of unique restriction sites along the gene cluster, manipulation of the PKS genes becomes a cumbersome task.

To overcome these limitations, a multiple-plasmid system for the generation of polyketide libraries has been developed in the laboratories of KOSAN Biosciences, Inc. (Figure 2).[10] In lieu of constructing one single-expression plasmid containing the entire set of PKS genes, individual open reading frames (ORFs) or single modules of PKSs are cloned on separate, compatible expression plasmids.[11] To co-express the individual plasmids in a heterologous host, the vectors are equipped with mutually selectable resistance markers. Two vectors are autonomously replicating plasmids on the basis of the pRM1 shuttle vector with SCP2* and *Col*E1 origins of replication in *Streptomyces spp.* and *E. coli*, respectively. For the third plasmid, an integrating vector (pSET152) was chosen, which inserts the PKS genes site-specifically at the actinophage ϕC31 attachment site. In all vectors, the PKS genes (ORF1, ORF2, and ORF3) are cloned downstream of a *Streptomyces* promoter (P*act*I), which is activated by the product of the activator gene, *act*II-ORF4. Although the concomitant expression of vectors in a host with multiple antibiotic selection is an established method in molecular biology,[12] the multiplasmid approach has a particularly high impact on combinatorial biosynthesis. As soon as functional subunits are available, they can be readily combined in all possible permutations and variations with little effort. For example, with x versions of ORF1, y versions of ORF2, and z versions of ORF3, a combinatorial library of $x \times y \times z$ mutants will be rapidly accessible. In comparison, with the same number of mutagenesis experiments, a single-expression-plasmid system would only provide a maximum of $x+y+z$ polyketides. In theory, two possible AT domains and four possible β-carbon atom modifications contribute to eight possible mutations per module, which results in $8 \times 8 = 64$ variations in each DEBS ORF. Concomitant expression of ORF1 – ORF3 in a combinatorial fashion would provide $64^3 = 262\,144$ mutants and consequently, at least hypothetically, a library of 262 144 polyketides. In comparison, with the same experimental effort, a single-expression-plasmid system would only provide a maximum of $64 \times 3 = 192$ polyketides.[10]

Scientists at KOSAN Biosciences, Inc., have engineered altered DEBS subunits by exchanging DEBS genes with rapamycin PKS genes.[13] They engineered eight ORF3 variants into the pSET integration site of *S. lividans*, then individually co-transformed the resulting strains with each of four ORF1 variants on pRM1-tsr vectors and two ORF2 variants on pRM1-hyg-derived vectors.[10] From 14 plasmids prepared, $8 \times 4 \times 2 = 64$ triple transformants were obtained, out of which 46 contributed to a library of 43 hybrid polyketides (Figure 3). Novel compounds have been produced at a level comparable to expression studies with a single plasmid carrying all three ORFs. Considering the immense synthetic effort for the chemical de novo synthesis of erythromycin by R. B. Woodward and 48 colleagues in 1981,[14] the multiplasmid approach is clearly a powerful technique for generating novel erythromycin analogues. Due to the feasible co-expression of the modified DEBS subunits observed, combining entire protein subunits from different modular PKSs has been anticipated to result in hybrid PKS complexes. By means of the multiplasmid approach, natural and altered subunits from DEBS with related picromycin synthase (PikPKS) and oleandomycin synthase (OlePKS) subunits have been successfully co-expressed.[15] In fact, the in vivo assembly of heterologous PKS complexes from naturally occurring subunits resulted in an increased production of hybrid polyketides compared to the DEBS mutagenesis studies.

In the same context, the possibility to heterologously co-express fragments of very large (> 50 kb) PKS gene clusters, such as for rapamycin, rifamycin, or epothilone, is an important feature of the multiplasmid approach. If the native producer is a slowly growing microorganism, the use of a Streptomycete-based multiplasmid expression system may significantly enhance productivity. In addition, by dissection of the gene cluster, the biosynthetic becomes readily amenable to genetic manipulations. This has been impressively demonstrated for the biosynthetic genes of the myxobacterial

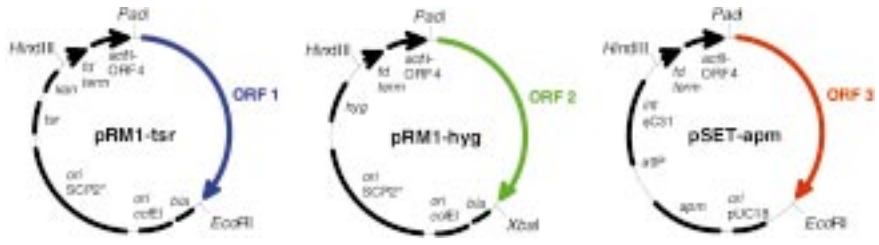

Figure 2. *Three-plasmid expression system for PKS genes. Shown are derivatives of pRM1-tsr, pRM1-hyg, and pSET-apm. pRM1-tsr and pRM1-hyg are bifunctional vectors for replication in both* E. coli *and* Streptomyces spp. *(*oriCol*E1 = origin of replication and* bla = β*-lactamase resistance gene for selection in* E. coli*;* ori*SCP2* = origin of replication;* tsr, kan, hyg = *thiostrepton, kanamycin, or hygramycin resistance genes, respectively, for selection in* Streptomyces ssp. *pSET-apm is an integrating plasmid (pSET152) with apramycin resistance gene (*apm*) and ϕC31-int attP loci for attachment and chromosomal integration of ORF3. The drawing is not to scale.*

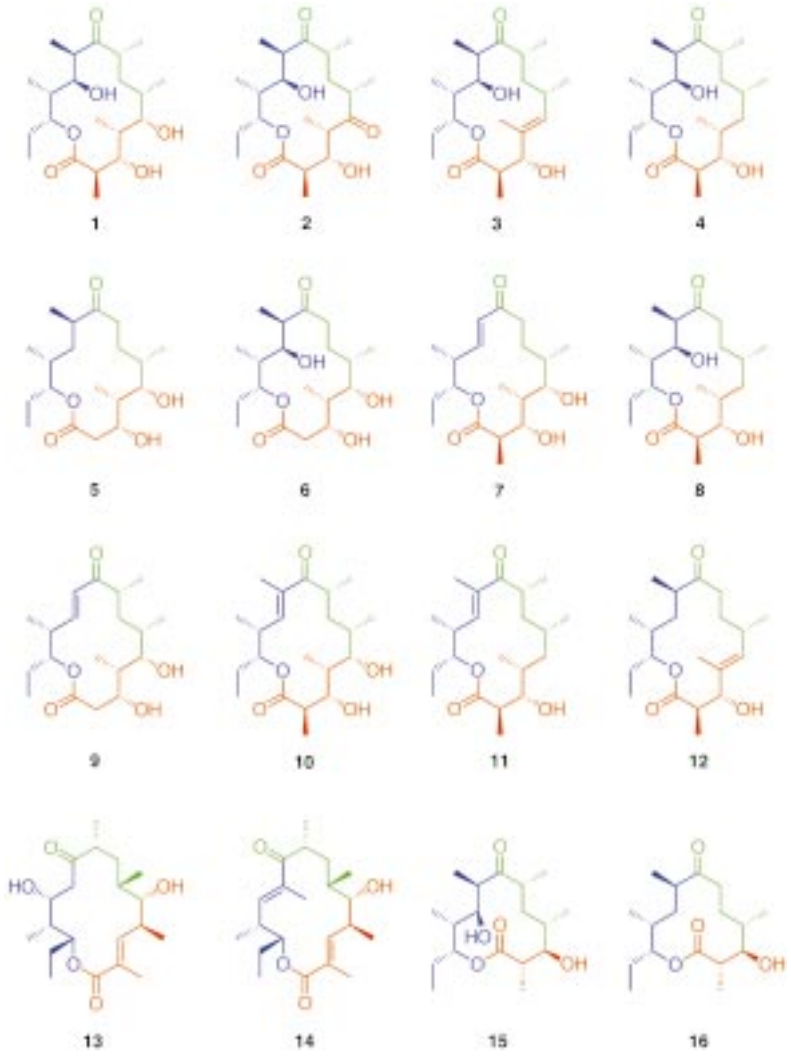

Figure 3. *Combinatorial library production. Shown are selected structures (**1** – **16**) out of 43 macrolactones produced by* S. lividans *K4 – 114 strains transformed with pRM1-tsr, pRM1-hyg, and pSET-apm derivatives.*

secondary metabolite epothilone.[16, 17] Epothilone is a promising anticancer agent against taxol-resistant tumors.[18] Since classical chemical approaches require over 20 synthetic steps, extraction of the natural compound from culture broths is favorable. Using the natural producer, on the other hand, does not appear to be economical. *Sorangium cellulosum* has a doubling time of 16 h, and 12,13-deoxyepothilone D (**18**), which is the most attractive clinical candidate, is formed only in trace amounts. To overcome these limitations, the large epothilone biosynthetic gene cluster was cloned into two compatible plasmids and coexpressed in a host (*Streptomyces coelicolor* CH999) with a doubling time of 2 h. The loading domain, the non-ribosomal peptide synthetase (NRPS), and modules 1 – 5 were cloned on a pRM1-tsr derivative, whereas modules 6 – 8, a gene for a P450 oxidase, and some post-PKS genes were cloned on an integrating pSET-apm derivative (Figure 4). In fact, sequential transformation of the plasmids and cultivation of the host – vector system resulted in successful production of epothilones A (**19**) and B (**20**). Although initial results showed a reduced in vivo productivity, with genetic and strain improvement, such as implementation of the epothilone resistance gene, the plasmid-borne expression system is likely to become a favorable producer of epothilones. Additionally, introduction of a methylmalonyl-specific AT and deletion of the P450 gene is a promising method for the fermentation-based preparation of epothilone D (**18**).

The multiple-plasmid approach is a promising method for exploring additional combinatorial post-PKS tailoring transformations such as oxidations, methylations, acylations, and glycosylations. Many of these tailoring reactions are critical for biological activity and for molecular recognition between the drug and its cellular target. A variety of substitutions have already been successfully carried out with a single-plasmid system by using heterologous hosts or knock-out mutants.[19] A two-plasmid system for the bioconversion of ε-rhodomycinone to rhodomycin D by concomitant expression of the carbohydrate biosynthetic and glycosyltransferase genes has recently been demonstrated,[20] but numberless possible combinations wait to be exploited. Likewise, the palette of type II PKS products may be extended by shuffling distinct PKS genes encoding minimal PKSs, acyltransferases, cyclases, and a variety of tailoring gene products. The multiplasmid approach is also full of promise for other biosynthetic systems such as non-ribosomal polypeptide synthetases, which have a structure similar to that of type I PKSs.[21]

To date, numerous modular and aromatic PKS biosynthetic gene clusters have been sequenced, which provide a versatile "toolbox" of building blocks for combinatorial biosynthesis. The multiplasmid approach is the method of choice to investigate the scope and limitations of creating novel compounds by fusing unique features from two or more biosynthetic pathways in a "mix-and-match" fashion. The technique permits both heterologous expression of very large biosynthetic genes by coexpression of subunits as well as generation of large numbers of structurally diverse compounds for high-throughput screening. Excitingly, the whole extent of possibilities for combinatorial biosynthesis remains to be explored.

I would like to thank Prof. Dr. Bradley S. Moore, University of Arizona, and Prof. Dr. H. G. Floss and Dr. Linquan Bai, University of Washington, for strong support and many helpful discussions. I am grateful to the Alexander von Humboldt Foundation for a Feodor Lynen fellowship.

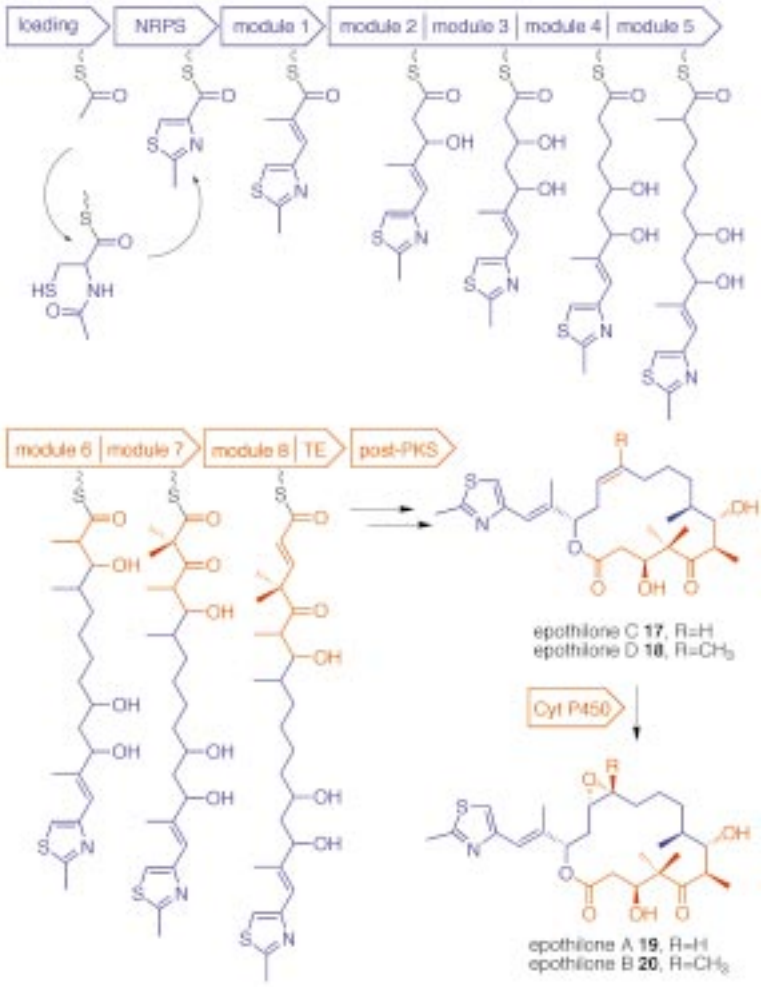

Figure 4. *Proposed epothilone biosynthesis. Heterologous expression of the myxobacterial metabolites epothilone A – D in an* S. coelicolor *CH999 host. Genes cloned on pRM1-tsr are drawn in blue, genes cloned on pSET-apm derivative in red. NRPS = non-ribosomal peptide synthase.*

[1] a) J. Rohr, *Angew. Chem.* **1995**, *107*, 963 – 967; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 881 – 885; b) C. A. Roessner, A. I. Scott, *Annu. Rev. Microbiol.* **1996**, *50*, 467 – 490; c) C. R. Hutchinson, *Curr. Opin. Microbiol.* **1998**, *1*, 319 – 329; d) D. H. Altreuter, D. S. Clark, *Curr. Opin. Biotechnol.* **1999**, *10*, 130 – 136; e) R. Bentley, J. W. Bennett, *Annu. Rev. Microbiol.* **1999**, *53*, 411 – 446.

[2] a) D. O'Hagan, *The Polyketide Metabolites*, Ellis Horwood, Chichester, **1991**; b) L. Katz, S. Donadio, *Annu. Rev. Microbiol.* **1993**, *47*, 875 – 912.

[3] a) C. Khosla, *Chem. Rev.* **1997**, *97*, 2577 – 2590; b) D. E. Cane, C. T. Walsh, C. Khosla, *Science* **1998**, *282*, 63 – 68.

[4] D. A. Hopwood, *Chem. Rev.* **1997**, *97*, 2465 – 2497.

[5] B. J. Rawlings, *Nat. Prod. Rep.* **1999**, *16*, 425 – 484.

[6] a) L. Katz, *Chem. Rev.* **1997**, *97*, 2557 – 2576; b) C. W. Carreras, D. V. Santi, *Curr. Opin. Biotechnol.* **1998**, *9*, 403 – 411.

[7] C. Khosla, R. S. Gokhale, J. R. Jacobsen, D. E. Cane, *Annu. Rev. Biochem.* **1999**, *68*, 219 – 253.

[8] a) R. McDaniel, S. Ebert-Khosla, D. Hopwood, C. Khosla, *Science* **1993**, *262*, 1546 – 1557; b) C. M. Kao, L. Katz, C. Khosla, *Science* **1994**, *265*, 509 – 512.

[9] R. Ziermann, M. C. Betlach, *BioTechniques* **1999**, *26*, 106 – 110.

[10] Q. Xue, G. Ashley, C. R. Hutchinson, D. V. Santi, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11 740 – 11 745.

[11] R. Ziermann, M. C. Betlach, *J. Ind. Microbiol. Biotechnol.* **2000**, *24*, 46 – 50.

[12] a) S. Kuhstoss, M. A. Richardson, R. N. Rao, *Gene* **1991**, *97*, 143 – 146; b) M. Bierman, R. Logan, K. Obrien, E. T. Seno, R. N. Rao, B. E. Schoner, *Gene* **1992**, *116*, 43 – 49; c) E. Guenzi, G. Galli, I. Grgurina, E. Pace, P. Ferranti, G. Grandi, *J. Biol. Chem.* **1998**, *273*, 14 403 – 14 410.

[13] R. McDaniel, A. Thamchaipenet, C. Gustafson, H. Fu, M. Betlach, G. Ashley, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1846 – 1851, and references therein.

[14] R. B. Woodward, E. Logusch, K. P. Nambiar, K. Sakan, D. E. Ward, B.-W. Au-Yeung, P. Balaram, L. J. Browne, P. J. Card, C. H. Chen, R. B. Chênevert, A. Fliri, K. Frobel, H.-J. Gais, D. G. Garratt, K. Hayakawa, W. Heggie, D. P. Hesson, D. Hoppe, I. Hoppe, J. A. Hyatt, D. Ikeda, P. A. Jacobi, K. S. Kim, Y. Kobuke, K. Kojima, K. Krowicki, V. J. Lee, T. Leutert, S. Malchenko, J. Martens, R. S. Matthews, B. S. Ong, J. B. Press, T. V. Rajan Babu, G. Rousseau, H. M. Sauter, M. Suzuki, K. Tatsuta, L. M. Tolbert, E. A. Truesdale, I. Uchida, Y. Ueda, T. Uyehara, A. T. Vasella, W. C. Vladuchick, P. A. Wade, R. M. Williams, H. N.-C. Wong, *J. Am. Chem. Soc.* **1981**, *103*, 3215 – 3217.

[15] L. Tang, H. Fu, R. McDaniel, *Chem. Biol.* **2000**, *7*, 77 – 84.

[16] I. Molnár, T. Schupp, M. Ono, R. E. Zirkle, M. Milnamow, B. Nowak-Thompson, N. Engel, C. Toupet, A. Stratmann, D. D. Cyr, J. Gorlach, J. M. Mayo, A. Hu, S. Goff, J. Schmid, J. M. Ligon, *Chem. Biol.* **2000**, *7*, 97 – 109.

[17] a) L. Tang, S. Shah, L. Chung, J. Carney, L. Katz, C. Khosla, B. Julien, *Science* **2000**, *287*, 640 – 642, and references therein.

[18] K. Gerth, N. Bedorf, G. Höfle, H. Irschik, H. Reichenbach, *J. Antibiot.* **1996**, *49*, 560 – 563.

[19] See for example: a) A. Trefzer, D. Hoffmeister, E. Künzel, S. Stockert, G. Weitnauer, L. Westrich, U. Rix, J. Fuchser, K. U. Bindseil, J. Rohr, A. Bechthold, *Chem. Biol.* **2000**, *7*, 133 – 142; b) Y. Q. Xue, D. H. Sherman, *Nature* **2000**, *403*, 571 – 575, and references therein; c) L. Zhao, J. Ahlert, Y. Xue, J. S. Thorson, D. H. Sherman, H.-w. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 9881 – 9882.

[20] C. Olano, N. Lomovskaya, L. Fonstein, J. T. Roll, C. R. Hutchinson, *Chem. Biol.* **1999**, *6*, 845 – 855.

[21] a) H. D. Mootz, M. R. Marahiel, *Curr. Opin. Biotechnol.* **1999**, *10*, 341 – 348; b) D. Konz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, R39-R48.