Polyketide Biosynthesis

DOI: 10.1002/anie.201207690

Substrate Specificity in Ketosynthase Domains from *trans*-AT Polyketide Synthases**

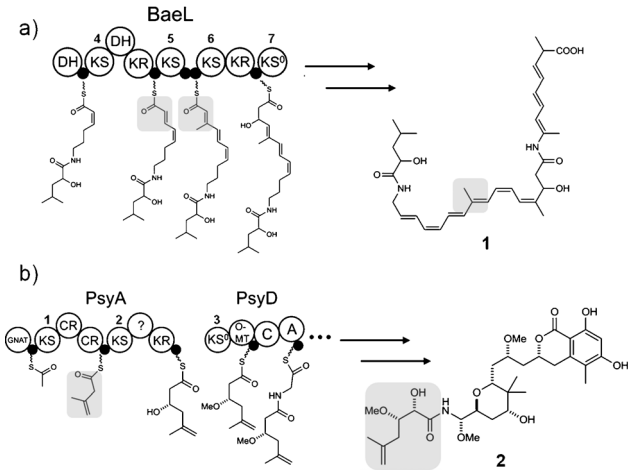
Matthew Jenner, Sarah Frank, Annette Kampa, Christoph Kohlhaas, Petra Pöplau, Geoff S. Briggs, Jörn Piel, and Neil J. Oldham**

Type I modular polyketide synthases (PKSs) catalyze the production of a remarkable array of biologically active natural products, with many employed as antibiotics, immunosuppressants, anti-parasitics, and anti-tumor agents.[1] Each PKS module possesses the enzymatic machinery for a single cycle of chain elongation and modification. Within a module, the acyltransferase domain (AT) loads an acyl-CoA derivative onto the phosphopantetheinyl (PPant) chain of the acyl carrier protein (ACP). Claisen thioester condensation, catalyzed by the ketosynthase (KS) domain, then results in chain elongation. The extent of β-keto processing is dictated by the presence of ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains allowing production of hydroxyl, olefinic, and fully saturated intermediates.[2]

Post-genomic techniques, and the propensity for type I PKS genes to arrange into distinct modules, have allowed the prediction, with fairly high confidence, of biosynthetic schemes from genetic sequences alone.[3] However, recently a novel group of PKSs lacking integral AT domains has been reported that deviates from standard colinearity rules. In these systems, the AT activity is instead supplied by free-standing enzymes acting in-trans.[4] Such *trans*-AT PKSs exhibit aberrant architecture and often incorporate novel enzymatic domains, resulting in poor biosynthetic assignments.[4b] Recent phylogenetic work established that KS domains from *trans*-AT systems correlate, at the sequence level, with their predicted biosynthetic intermediates. This finding offers the exciting potential for KS-based PKS-to-product assignment for *trans*-AT systems.[5]

The extent of evolution-based substrate correlation is predicted to extend as far as the β-position for each biosynthetic intermediate. Although the KS-specificity method works remarkably well computationally, it lacks functional testing. Such knowledge is important because it could provide rules for biosynthetic engineering and may allow for single-domain-based biosynthetic characterization of *trans*-AT PKS megaenzymes. To address substrate specificity in *trans*-AT PKSs, we have developed and applied a mass spectrometry (MS)-based method to assign the specificity of KS domains, thereby testing the predicted assignment of *trans*-AT PKS clusters. Previous methods to determine the specificity of the more familiar *cis*-AT PKSs have included radioactivity assays and MS analysis of trypsin-digested proteins.[6]

Herein, we report the substrate specificity of KS5 located on BaeL of the bacillaene **1** PKS,[7] and KS1, KS2, KS3^0 on PsyA/PsyD of the psymberin **2** PKS (Scheme 1).[8] Bacillaene is an antibiotic isolated from *Bacillus* strains,[9] and psymberin is produced by an uncultivated symbiont of a sponge.[10] Using a rapid, MS-based method to analyze intact domains we show that BaeL KS5 and PsyD KS3^0 do not accept an acyl chain with methyl branching at the β-position, a common modification of *trans*-AT PKS products.[4b,11] We rationalize this specificity using homology modeling of the enzyme active



Scheme 1. Partial proposed biosynthetic schemes for a) bacillaene **1** and b) psymberin **2**. KS = ketosynthase; KS^0 = non-elongating ketosynthase; KR = ketoreductase; DH = dehydratase; CR = crotonase; O-MT = *O*-methyl transferase; C = NRPS condensation domain; A = NRPS adenylation; GNAT = acetyl-loading AT; ACP domains (●). The sequential KS numbering is denoted above the KS domains, and relevant carbon-branched regions of the intermediates are highlighted in gray.

[*] M. Jenner, Dr. G. S. Briggs, Dr. N. J. Oldham
School of Chemistry, University of Nottingham
University Park, Nottingham NG7 2RD (UK)
E-mail: neil.oldham@nottingham.ac.uk
S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, Prof. Dr. J. Piel
Kekulé Institute of Organic Chemistry and Biochemistry
University of Bonn
Gerhard-Domagk-Strasse 1, 53121 Bonn (Germany)
and
Institute of Microbiology,
Eidgenössische Technische Hochschule (ETH) Zurich
Wolfgang-Pauli-Strasse 10, 8083 Zurich (Switzerland)
E-mail: joern.piel@uni-bonn.de
jpiel@ethz.ch

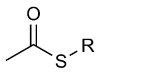
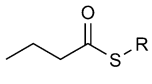
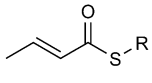
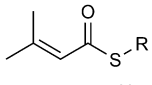
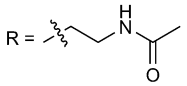
[**] We are grateful for financial support by the Biotechnology and Biological Science Research Council (BBSRC) and the University of Nottingham to N.J.O. and the DFG (SFB 624 and PI 430/8-1) to J.P.

Supporting information for this article (experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201207690.

sites, modulate the tolerance of BaeL KS5 with a single site-directed mutation, and propose a general rule for the tolerance of β-methyl-branching by KS domains.

Previous sequence analysis places BaeL KS5 in a clade with other KS domains predicted to accept an unbranched α,β-unsaturated intermediate.[5a] KS5 is located before a key β-branching step in the bacillaene biosynthetic pathway (Scheme 1 a), and would be expected to be specific for an unbranched substrate.[11,12] To probe the specificity of this KS domain, a range of simple *N*-acyl cysteamine (SNAC) thioesters was synthesized (Table 1); these thioesters are known to be suitable mimics of the acyl-phosphopantetheine chain attached to the ACP.[6]

Table 1: Estimated initial acylation rates[a] for KS domains with SNAC-thioesters **3**–**6**.

Substrate		Rate [$\times 10^{-6}$ mol dm^{-3} min^{-1}]				
		BaeL KS5	PsyA KS1	PsyA KS2	PsyD KS3^0	BaeL KS5 (M237A)
	**3**	0.07	3[b]	0.3	0.3	0.06
	**4**	0.06	2	0.3	0.1	0.05
	**5**	0.07[a]	2	0.3	0.2	0.05
	**6**	ND	0.04	0.03	ND	0.01
						

[a] Initial rate was estimated from a plot of ln[KS]/[KS$_0$] versus *t*, given to one significant figure. [b] The SNAC-mimic of the predicted substrate for each KS domain. ND = No acylation was detected over the time scale of the incubation. Estimated error in measurements: $\pm 0.005 \times 10^{-6}$ mol dm^{-3} s^{-1}.

Following incubation of KS5 with each SNAC derivative, the reactions were quenched at set time points by addition of 0.1% TFA. After desalting (see Supporting Information), electrospray ionization mass spectra of KS5 were obtained for each time point. The mass of the non-acylated KS5 was determined to be 78684 Da, in excellent agreement with the predicted value of 78686 Da (Supporting Information, Figure S1). KS5 acylation was detected based on the appearance of an additional signal corresponding to acyl-KS5 (Figure 1 a). Time-dependent acylation of the KS5 domain was measured from the relative intensities of the non-acylated and acyl-KS5 peaks for each SNAC derivative and plotted (Figure 1 b).

The *E*-2-butenoyl **5** and a 3-methyl-2-butenoyl-SNAC **6** thioesters were prepared as simplified analogues of the native substrates of BaeL KS5 and KS6, respectively (Scheme 1 and Table 1). MS-based kinetic analysis of KS5 with **5** and **6** revealed that this domain was highly selective for the unbranched substrate, with no acylation observed after 40 minutes with SNAC derivative **6**. This result provided strong evidence that KS5 can not catalyze acylation by a β-methyl-branched substrate. In contrast, the unbranched SNACs **3**–**5** successfully acylated KS5, reaching saturation after 20 minutes. *O*-branched SNACs were found to acylate KS5, but at a slower rate than SNACS **3**–**5** (Table S2). MS data obtained from incubation with SNAC derivative **5** resulted in additional signals that corresponded to the mass of the attachment of the intact SNAC derivative (Figure S6). This is explained by Michael addition to **5** by nucleophilic sites on the protein surface. Conjugate addition did not appear to interfere with the ability of KS5 to be acylated by **5**, and the two products were resolved by MS (something that would not be possible with a radioactivity-based assay).

These results demonstrate that simple, short-chain SNAC analogues of acyl-ACP intermediates are able to reveal substrate specificity in KS5 of the bacillaene PKS. Notably, this domain, which is positioned immediately before a β-branching step, is highly specific for an unbranched substrate, providing evidence that specificity studies of KS domains may aid biosynthetic assignment.

The first three KS domains from the psymberin cluster were then examined. Each of these KSs is predicted to use different substrate intermediates (Scheme 1 b). Substrate profiling of PsyA KS1 revealed remarkably relaxed specificity, with both branched and unbranched substrates accepted. Kinetic analysis showed that the native acetyl substrate **3** was readily incorporated (Table 1, Figure S7). However, the butyryl derivative **4** was found to acylate at a comparable rate, suggesting that chain length was not a dictating factor. In contrast, the β-methyl-branched SNAC **6** was a worse substrate, with rates approximately 30-fold slower than the unbranched intermediates. However, unlike BaeL KS5, carbon-branched SNACs were able to acylate PsyA KS1 to some extent. Although initially hypothesized to be a highly specific KS domain, owing to its position at the start of a biosynthetic cluster and to its small substrate, functionally KS1 is capable of accepting a range of substrates other than the acetyl-SNAC **3**. We postulate that there may be little or no evolutionary pressure for KS1 to acquire specificity, as the preceding GNAT domain supplies the correct acetyl starter unit from malonyl-CoA.[13] The promiscuity of this KS domain, together with its relatively high turnover, suggests that it may be an attractive choice for biosynthetic engineering.

Interestingly, PsyA KS2 produced a similar tolerance profile for the SNAC-thioesters as KS1 (Figure S8). KS2 is predicted to accept a β-branched intermediate, and **6** was indeed found to be a viable substrate. KS2 was also acylated by straight-chain acyl SNACs **3**–**5**, as well as by *O*-branched substrates (Table S2).

PsyD KS3^0 is a non-elongating KS, and is believed to accept an (*S*)-β-hydroxyl intermediate (Scheme 1 b). The corresponding SNAC was found to be a viable substrate (Figure S9, Table S2). The β-methyl-branched SNAC **6** was not found to acylate KS3^0 after 40 minutes. Thus, with respect to this substrate the specificities of BaeL KS5 and PsyD KS3^0 were rather similar, but different from those of PsyA KS1 and KS2, for which tolerance for a β-carbon branch was seen.

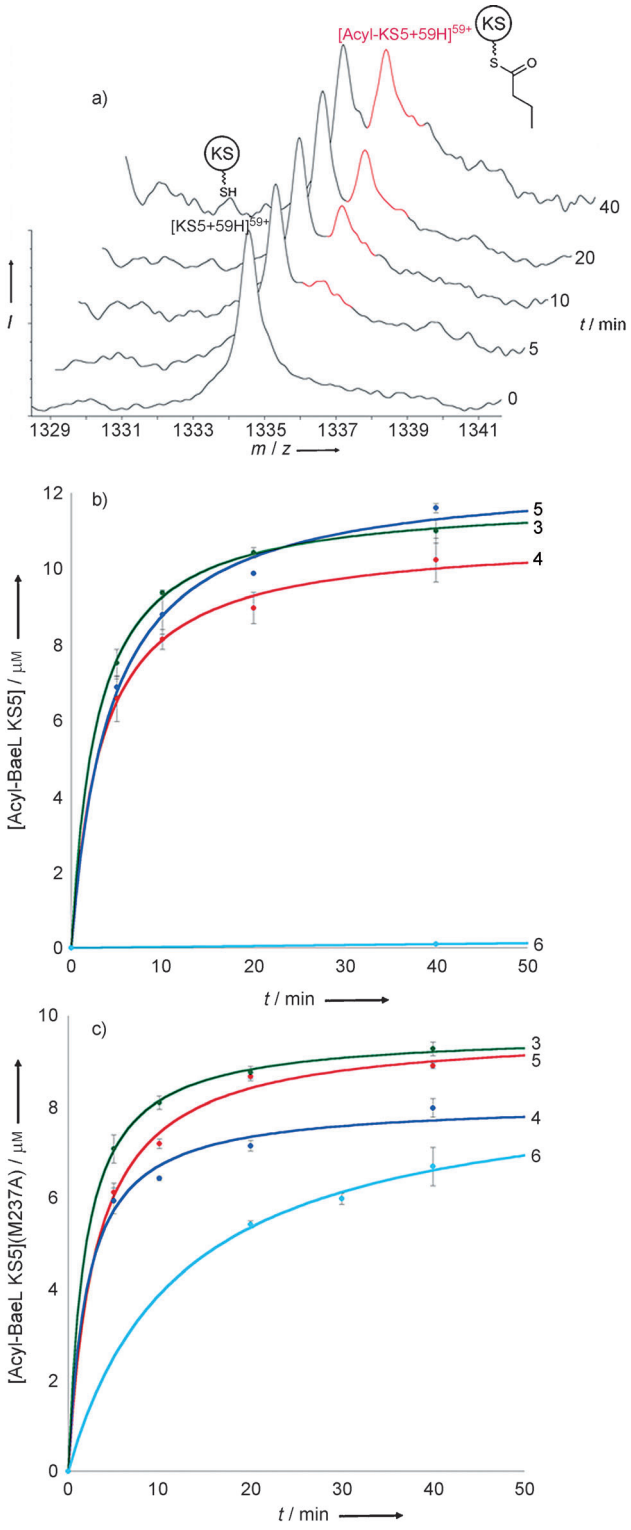


Figure 1. Acylation of BaeL KS5 using SNAC thioesters. a) Stacked mass spectrum of the 59^{+} charge state of BaeL KS5 showing acylation by SNAC **4**. Kinetic plots of b) WT BaeL KS5 and c) BaeL KS5(M237A) with SNAC thioesters **3**–**6**.

A combination of homology modeling and sequence analysis was employed to rationalize the observed differences in KS specificity. Initial focus was directed on BaeL KS5, as this domain appeared to display the highest levels of specificity towards unbranched substrates. Using the CPHmodel server,[14] homology models of BaeL KS5 and the downstream, β-branch accepting, BaeL KS6 were successfully constructed. Upon examination of the binding sites, a subtle, but potentially important difference between KS5 and KS6 was observed. The nature of the residue immediately preceding the active site cysteine (X-Cys) appeared to affect the size of the binding pocket. X was identified as Met in BaeL KS5 and Ala for KS6. To examine the effect of this position on substrate binding, the unbranched biosynthetic intermediate of KS5 was manually docked into the binding site of KS5 (Figure 2a). Interestingly, the X residue was found to be positioned adjacent to the β-position of the substrate, with the potential to dictate the ability of the KS domain to accept a β-branched substrate.

Applying this same method to the psymberin KS domains revealed that PsyA KS1 and KS2 both had Ala residues at position X, producing a binding pocket similar to BaeL KS6 (Figure S12). Interestingly, in this study, KS1 and KS2 have been shown experimentally to accept the carbon branched substrate **6**. In contrast PsyD $KS3^{0}$ and BaeL KS5, which do not accept this test substrate, possess a Met at the X position. These results provide strong evidence for the role of a Met residue in excluding bulky carbon β-branches. A sequence analysis of the 138 KS domains used in the initial phylogenetic study,[5a] together with 12 subsequently identified in the psymberin PKS,[8] focused on the nature of the X-Cys residues. Results from this analysis showed that for a large proportion of known KS domains X = Ala. However, there are some clades that deviate markedly from this norm. The clade containing BaeL KS5, for example, comprises KS domains that are believed to accept unbranched enoyl substrates. Here, 94% of the KS domains in this clade feature a Met residue at the X position (Table S3). This observation added further evidence to the hypothesis that X = Met was required to confer specificity towards unbranched substrates.

To test the hypothesis that the bulkier Met residue of BaeL KS5 prevents acylation by carbon-branched substrates, a BaeL KS5 (M237A) mutant was designed. The Met to Ala mutation was predicted to provide the necessary space in the binding pocket for a sterically demanding β-methyl branch (see Figure 2B for homology model and branched substrate docking). Incubation of the KS5 (M237A) mutant with the unbranched SNACs **3**–**5** resulted in rates of acylation similar to those of WT BaeL KS5. This was expected, as increasing the space inside the binding cleft was unlikely to affect the ability of an unbranched chain to acylate the KS. However, there was a notable change in the tolerance towards the β-branched test substrate. The carbon-branched SNAC derivative **6** was found to acylate BaeL KS5(M237A) at a reasonable rate, demonstrating that the mutation had altered the specificity of the KS domain. These observations suggest that KS domains from *trans*-AT PKSs have the ability to differentiate substrates by the extent of β-branching, but can be engineered to accept non-natural acyl units.

Although the use of short SNAC thioesters has allowed elucidation of a certain level of specificity in the KS domains we studied, the specificity is notably lower than that predicted

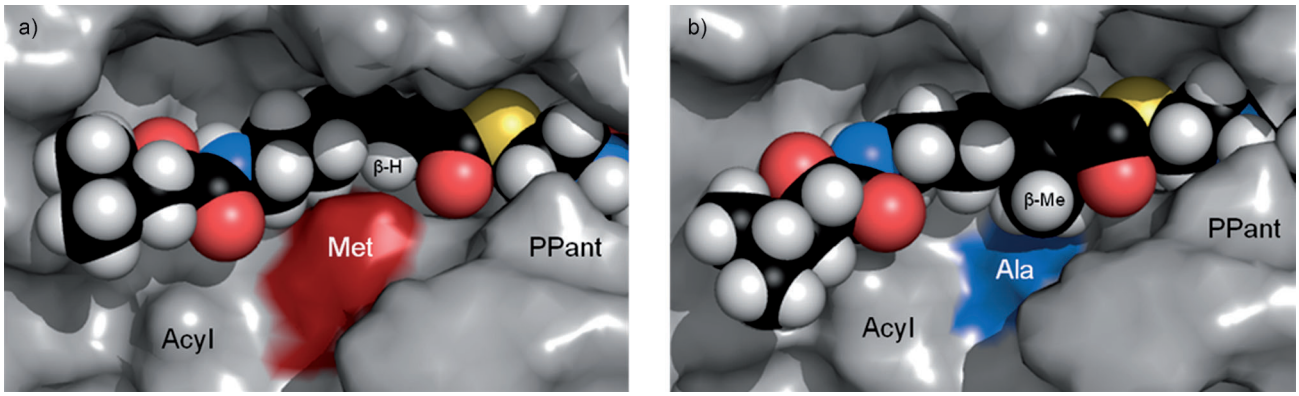


Figure 2. Homology models of KS domains. a) WT BaeL KS5 with a predicted, unbranched biosynthetic intermediate docked. The methionine residue highlighted in red is postulated to provide the necessary steric bulk to prevent β-branched substrates entering the binding pocket. b) BaeL KS5 (M237A) with the β-methyl analogue docked. In this case, the alanine residue highlighted in blue is thought to reduce the steric crowding in the binding pocket, allowing a β-branched substrate to bind. PPant represents the phosphopantetheinyl chain of the preceding ACP domain.

by phylogenetic analysis.[5a] This may reflect low specificity at the functional level, or might be an artifact from the use of simplified substrates. Extension of the acyl chain to better mimic that of the natural substrate, or use of the full pantetheinyl chain, instead of the simple SNAC thioester, may allow higher-resolution insight into KS specificity.

The current hypothesis concerning the evolutionary origin of *cis*-AT and *trans*-AT PKSs is that they have evolved by different routes.[5a] It has been suggested that the KS domains from *trans*-AT PKSs developed through horizontal gene transfer, by assembly of substrate-specific KS domains.[5a,15] This is in contrast to *cis*-AT KS domains, which have evolved as a result of duplication of entire modules in which KS domains are embedded.[16] A prerequisite for this evolutionary-duplication mode is relatively low substrate specificity, which is indeed observed for many *cis*-AT KSs.[17] Examination of the X-Cys active site motif in KS domains from *cis*- and *trans*-AT PKSs revealed interesting differences between the two types (Figure S13). For *trans*-AT PKSs, which accept α- and β-carbon-branched substrates, X is always a less sterically demanding residue Ala or Gly. Also, X is exclusively Asn for KSs that use amide-containing substrates. *cis*-AT PKSs, in contrast, might not be expected to have evolved the variety of amino acids at the crucial X position to develop this mechanism of specificity. Examination of the available sequence data reveals that X is Ala for all cases of fully assigned *cis*-AT PKSs in the literature (Figure S13). This observation adds further evidence to the theory that different modes of evolution have occurred in *cis* and *trans*-AT PKSs.

In summary, by using a novel MS-based assay we have identified substrate specificity associated with carbon β-branching in *trans*-AT PKS KS domains. We have characterized a key residue in BaeL KS5 that dictates tolerance for this branching. Examination of the equivalent residue in the sequences of 150 KS domains from *trans*-AT PKSs suggests that the presence of Gly or Ala is required for acceptance of a carbon branch in the acyl chain. This general rule provides insight for successful *trans*-AT PKS engineering and may aid functional assignment of natural product biosynthesis in favorable cases; specifically in highlighting the incompatibility of proposed β-methyl intermediates with X = Met KS domains. Additionally, these results indicate the potential to tune the specificity of KS domains towards desired biosynthetic products.

Received: September 24, 2012
Published online: December 5, 2012

Keywords: biosynthesis · electrospray ionization · ketosynthases · mass spectrometry · polyketides

[1] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2007**, *70*, 461 – 477.
[2] C. Khosla, S. Kapur, D. E. Cane, *Curr. Opin. Chem. Biol.* **2009**, *13*, 135 – 143.
[3] C. Hertweck, *Angew. Chem.* **2009**, *121*, 4782 – 4811; *Angew. Chem. Int. Ed.* **2009**, *48*, 4688 – 4716.
[4] a) J. Piel, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14002 – 14007; J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16222 – 16227; Y. Q. Cheng, G. L. Tang, B. Shen, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3149 – 3154; A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419 – 430; b) J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996 – 1047.
[5] a) T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225 – 233; b) R. Teta, M. Gurgui, E. J. N. Helfrich, S. Künne, A. Schneider, G. Van Echten-Deckert, A. Mangoni, J. Piel, *ChemBioChem* **2010**, *11*, 2506 – 2512.
[6] J. Q. Wu, K. Kinoshita, C. Khosla, D. E. Cane, *Biochemistry* **2004**, *43*, 16301 – 16310; N. A. Schnarr, A. Y. Chen, D. E. Cane, C. Khosla, *Biochemistry* **2005**, *44*, 11836 – 11842.
[7] a) A. M. Albertini, T. Caramori, F. Scoffone, C. Scotti, A. Galizzi, *Microbiology* **1995**, *141*, 299 – 309; b) X.-H. Chen, J. Vater, J. Piel, P. Franke, R. Scholz, K. Schneider, A. Koumoutsi, G. Hitzeroth, N. Grammel, A. W. Strittmatter, G. Gottschalk, R. D. Süssmuth, R. Borriss, *J. Bacteriol.* **2006**, *188*, 4024 – 4036.
[8] 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.
[9] P. S. Patel, S. Huang, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes, F. Mayer, *J. Antibiot.* **1995**, *48*, 997 – 1003;

R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1506 – 1509.
[10] R. H. Cichewicz, F. A. Valeriote, P. Crews, *Org. Lett.* **2004**, *6*, 1951 – 1954.
[11] C. T. Calderone, W. K. Kowtoniuk, N. L. Kelleher, P. C. Dorrestein, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8977 – 8982.
[12] J. Moldenhauer, X.-H. Chen, R. Borriss, J. Piel, *Angew. Chem.* **2007**, *119*, 8343 – 8345; *Angew. Chem. Int. Ed.* **2007**, *46*, 8195 – 8197.
[13] L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Hakansson, J. L. Smith, D. H. Sherman, *Science* **2007**, *318*, 970 – 974.
[14] M. Nielsen, C. Lundegaard, O. Lund, T. N. Petersen, *Nucleic Acids Res.* **2010**, *38*, W576 – W581.
[15] T. Hochmuth, J. Piel, *Phytochemistry* **2009**, *70*, 1841 – 1849.
[16] H. Jenke-Kodama, A. Sandmann, R. Müller, E. Dittmann, *Mol. Biol. Evol.* **2005**, *22*, 2027 – 2039; A. Ginolhac, C. Jarrin, P. Robe, G. Perriere, T. Vogel, P. Simonet, R. Nalin, *J. Mol. Evol.* **2005**, *60*, 716 – 725.
[17] C. Khosla, R. S. Gokhale, J. R. Jacobsen, D. E. Cane, *Annu. Rev. Biochem.* **1999**, *68*, 219 – 253.

