**Reengineering Biosynthetic Pathways**

# Biosynthesis of Epothilone Intermediates with Alternate Starter Units: Engineering Polyketide–Nonribosomal Interfaces**

*Sarah E. O'Connor, Christopher T. Walsh,* and Fei Liu*

The epothilones, a class of antitumor macrolides,[1] are nonribosomal-peptide–polyketide natural products that are constructed by mixed enzyme assembly lines composed of nonribosomal-peptide synthetases (NRPS) and polyketide synthases (PKS).[2] In the early steps of the epothilone

[*] Prof. Dr. C. T. Walsh, Dr. S. E. O'Connor,+ Dr. F. Liu+
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, MA 02215 (USA)
Fax: (+1) 617-432-0438
E-mail: christopher_walsh@hms.harvard.edu

[+] These authors contributed equally to this work.

biosynthetic pathway, the incipient methylthiazolyl group is formed by acetyl transfer from the PKS subunit EpoA to the cysteine residue of the NRPS subunit EpoB.[3] Subsequent cyclization, dehydration, and oxidation by the cyclization (Cy) and oxidase (Ox) domains of EpoB results in the methylthiazolyl species, which is then transferred from EpoB to the downstream PKS acceptor subunit EpoC by the ketosynthase (KS) domain of EpoC (Figure 1a).[4] As both EpoB and EpoC have demonstrated tolerance for unnatural upstream acyl substrates in vitro,[5] it was hypothesized that proteins from other biosynthetic pathways with novel substrate specificities could be used in place of EpoA and EpoB within the epothilone assembly line (Figure 1b).

This study demonstrates that EpoA or EpoB can be replaced in vitro with proteins from the rapamycin,[6] enterobactin,[7] or yersiniabactin[8] biosynthetic pathways. Fusion of a protein-recognition element to the C terminus of the noncognate protein is critical for substrate transfer, as observed in studies with PKS enzymes.[9] Although similar recognition domains have been proposed for mixed PKS/NRPS assembly lines,[10] this hypothesis has not been validated experimentally. A putative recognition sequence at the C terminus of the PKS EpoA appears to facilitate substrate transfer to EpoB, whereas a sequence at the C terminus of the NRPS EpoB facilitates substrate transfer to EpoC (Figure 2a).
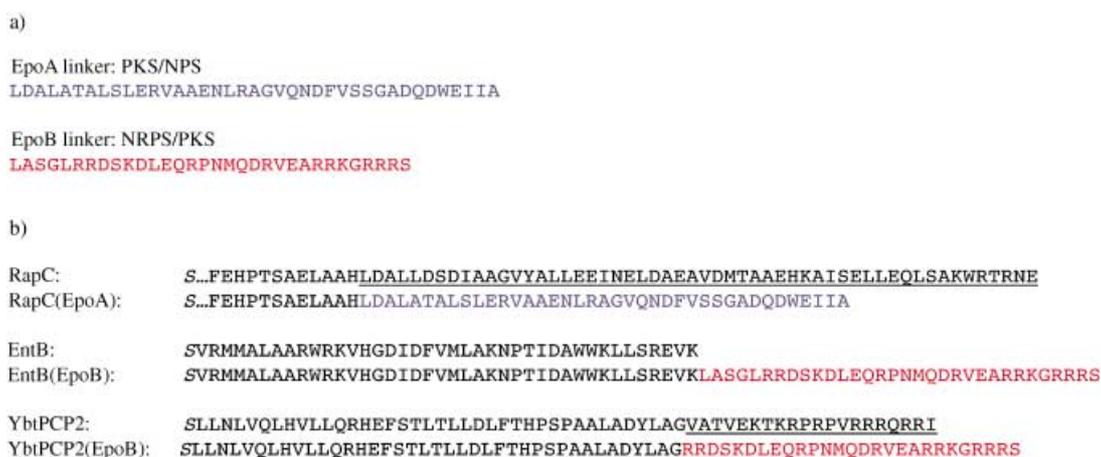
When the acyl carrier protein from RapC of the rapamycin biosynthetic pathway was loaded with [³H]acetyl-CoA (see Experimental Section) and incubated with cysteinyl-S-EpoB and methylmalonyl-S-EpoC, no substrate transfer occurred (Figure 3a, top). However, when the 39-residue EpoA recognition sequence was fused in place of the C terminus of RapC to yield RapC(EpoA) (Figure 2b) and this hybrid protein was used, formation of compound **1** was observed (Figure 3a, bottom).

The free-standing aryl carrier protein EntB from the enterobactin biosynthetic pathway,[7] primed with the methylthiazolyl-S-pantetheinyl moiety derived from methylthiazolyl-CoA, produced compound **1** when incubated with methylmalonyl-S-EpoC (Figure 3b, top). However, after fusion of the 32-residue EpoB linker to the enterobactin protein (EntB(EpoB), Figure 2b), the formation of product **1** was improved (Figure 3b, bottom). When Ybt-PCP2, a peptidyl carrier protein from the yersiniabactin biosynthetic pathway,[8] was primed with the methylthiazolyl-S-acyl group and incubated with methylmalonyl-S-EpoC, a small amount of product formation was again observed (Figure 4). However, replacement of the C-terminal sequence of Ybt-PCP2 by
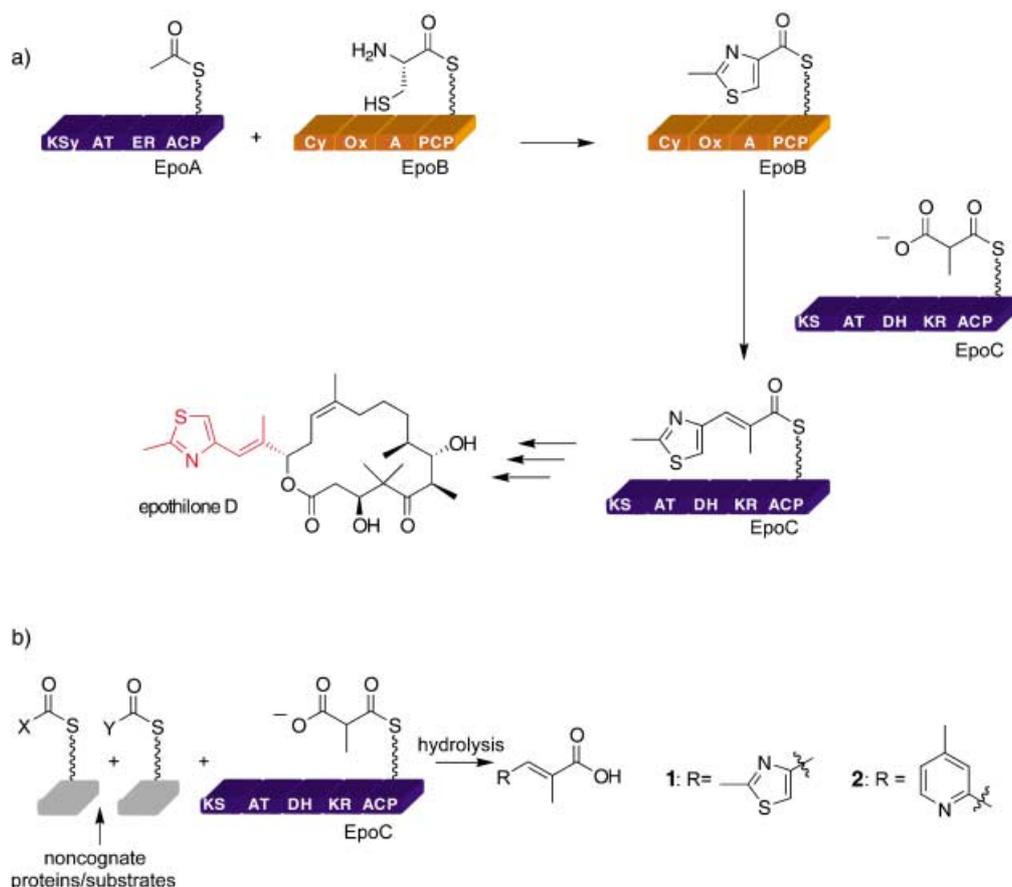


**Figure 1.** a) Early steps in the biosynthesis of epothilone. The EpoA, EpoB, and EpoC proteins are depicted with their catalytic modules. b) Integration of noncognate proteins into the epothilone assembly line. Use of noncognate proteins with different substrates. (X = methyl and Y = cysteine to make epothilone fragment **1**; alternatively the X-containing substrate was omitted from the enzymatic reaction, and the substrates Y = methylthiazole or methylpyridine were used to make epothilone fragments **1** or **2**, respectively.)

**a)**

EpoA linker: PKS/NPS
LDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIA

EpoB linker: NRPS/PKS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS

**b)**

RapC:        S...FEHPTSAELAAHLDALLDSDIAAGVYALLEEINELDAEAVDMTAAEHKAISELLEQLSAKWRTRNE
RapC(EpoA):  S...FEHPTSAELAAHLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIA

EntB:        SVRMMALAARWRKVHGDIDFVMLAKNPTIDAWWKLLSREVK
EntB(EpoB):  SVRMMALAARWRKVHGDIDFVMLAKNPTIDAWWKLLSREVKLASGLRRDSKDLEQRPNMQDRVEARRKGRRRS

YbtPCP2:     SLLNLVQLHVLLQRHEFSTLTLLDLFTHPSPAALADYLAGVATVEKTKRPRPVRRRQRRI
YbtPCP2(EpoB): SLLNLVQLHVLLQRHEFSTLTLLDLFTHPSPAALADYLAGRRDSKDLEQRPNMQDRVEARRKGRRRS

*Figure 2.* a) Proposed C-terminal recognition sequences of EpoA (in blue) and EpoB (in red). b) Amino acid sequences of the C termini of the proteins used in this study. Underlined sequences in RapC and YbtPCP2 were removed and replaced with the EpoA or EpoB recognition sequence. Italicized letters indicate the serine residue that is pantetheinylated.

that of the EpoB recognition sequence (Ybt-PCP2(EpoB), Figure 2b)[11] greatly improved product yield (Figure 4). The use of the protein Ybt-PCP2(NS), which lacks the C-terminal residues of the wild-type carrier protein, also led to diminished product formation (decreased by 60% relative when Ybt-PCP2(EpoB) was used). These studies indicate that the recognition sequences that assist protein recognition in the PKS pathways[9a] are also observed in both PKS/NRPS and NRPS/PKS mixed interfaces. A preliminary study revealed that product formation in the presence of methylthiazolyl-*S*-EntB(EpoB) and methylmalonyl-*S*-EpoC is much slower than the rate of product formation that is observed with the cognate EpoB protein (Figure 3c). This suggests that the engineered partner protein, although viable, is not as efficient as the wild-type EpoB donor. Further optimization of the recognition sequence may lead to improved acyl-group transfer.

The use of noncognate proteins to deliver tethered acyl groups to NRPS or PKS assembly lines for the synthesis of unnatural products is an important goal. Therefore, we examined the ability of EntE,[7] the partner protein of EntB, to load aromatic acyl substrates onto the phosphopantetheinyl group of EntB(EpoB). Although the interaction of EntE with EntB(EpoB) is adversely affected by the addition of the linker sequence, a substantial extent of [14C]salicylate-group loading (57%) was noted for the modified carrier protein (compared to 90% for the wild type). EntE can also activate and load the heteroaromatic substrate 4-methyl-2-pyridine carboxylic acid (picolinic acid), though at a significantly decreased rate (0.33 min⁻¹) to that of its activation and loading of the natural substrate 2,3-dihydroxybenzoate (330 min⁻¹).[7a] As a methylpyridine epothilone derivative has been shown to exhibit enhanced cytotoxic activity,[12] this substrate provides an interesting test case for further study. Acyl transfer, reduction, and dehydration on EpoC could be demonstrated with this novel substrate by incubating EntE, picolinic acid, EntB(EpoB), and EpoC, in the presence of [14C]methylmalonyl-CoA and NADPH (Figure 5). The recognition sequence from EpoB appears to dramatically improve product formation when this noncognate substrate is used, thus suggesting that improvement of the protein–protein interaction by the recognition sequence is critical when unusual substrates are incorporated into the enzymatic assembly line.

This study demonstrates that proteins from rapamycin, yersiniabactin, and enterobactin biosynthetic pathways can interact productively with the epothilone assembly line, which suggests that short recognition sequences in hybrid PKS/NRPS systems may be fused to multiple proteins. Thus, the "linker hypothesis", which has enabled the mixing and matching of proteins within PKS systems,[9a] can also be used to mix PKS and NRPS systems and enable acyl transfer and subsequent reactions such as cyclization, reduction, or dehydration to build novel assembly-line intermediates. The rates for product formation in the hybrid system are decreased relative to that observed in the natural system, thus suggesting that the terminal recognition sequence is not the only element required for productive acyl transfer. Consistent with this observation, studies with the erythromycin biosynthetic pathway suggest that the entire protein may contribute to the protein–protein interaction.[9d] Nevertheless, addition of the C-terminal recognition sequence greatly improved substrate transfer in the three examples described herein. Although a coiled-coil protein interaction has been proposed for the PKS recognition elements,[9e] it is not known whether PKS/NRPS sequences interact in a similar manner. More detailed studies designed to elucidate the mechanism by which these recognition elements foster protein–protein interactions are currently underway.

### Experimental Section

Carrier proteins were amplified by polymerase chain reaction (PCR) from existing constructs or from genomic DNA and ligated into pET28b (Novagen) to create constructs that contained an N-terminal His tag. The carrier protein of RapC consists of 151 amino acids (residues 6110–6260 of RapC); Ybt-PCP2 consists of the amino acids 1933–2041 of HMWP2. Carrier proteins with a linker sequence were
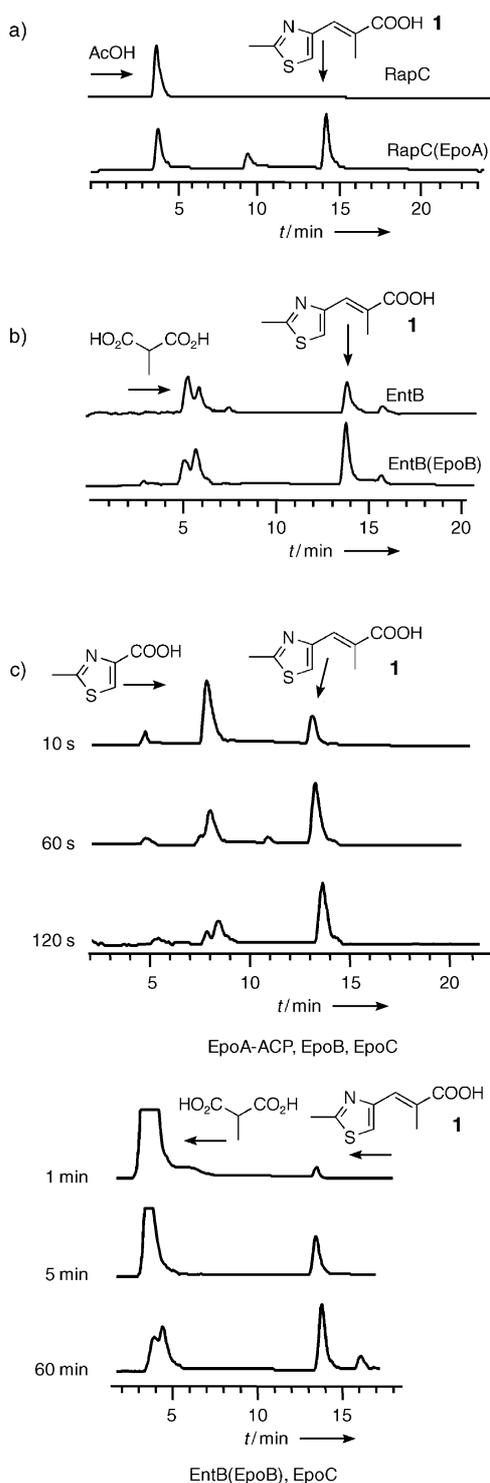
**Figure 3.** a) Radio-HPLC traces that indicate the formation of compound **1** from the incubation of cysteinyl-S-EpoB, methylmalonyl-S-EpoC, and either 1) [³H]acetyl-S-RapC or 2) [³H]acetyl-S-RapC(EpoB). b) Radio-HPLC traces that indicate the formation of compound **1** from the incubation of [¹⁴C]methylmalonyl-S-EpoC and 1) methylthiazolyl-S-EntB or 2) methylthiazolyl-S-EntB(EpoB). c) Comparison of product formation from wild-type epothilone biosynthetic enzymes (acetyl-S-EpoA-carrier protein, [³⁵S]cysteine-S-EpoB, and methylmalonyl-S-EpoC) with product formation from enterobactin and epothilone biosynthetic pathways (methyl-thiazolyl-S-EntB(Epo) and [¹⁴C]methylmalonyl-S-EpoC) over time. ACP = acyl carrier protein.
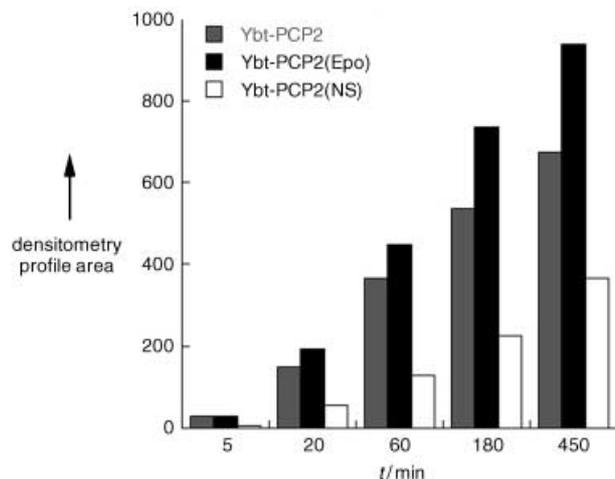


**Figure 4.** Comparison of the formation of compound **1** by the interfaces formed between [¹⁴C]methylmalonyl-S-EpoC and methyl-thiazolyl-S-Ybt-PCP2, Ybt-PCP2(EpoB), and Ybt-PCP2(NS) over time.
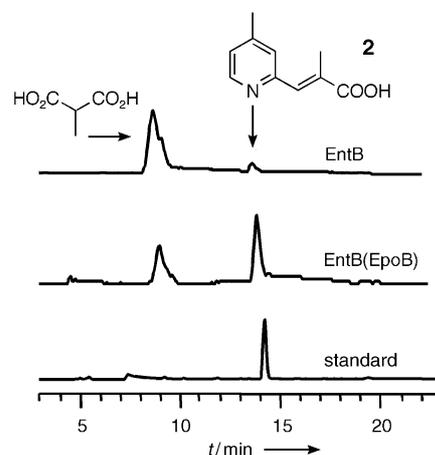


**Figure 5.** Radio-HPLC trace that indicates the product **2** formed upon incubation of [¹⁴C]methylmalonyl-S-EpoC and picolinyl-S-EntB or picolinyl-S-EntB(EpoB), against the UV-HPLC trace (254 nm) of compound **2** as a chemical standard.

constructed by splicing-overlap-extension PCR.[13] Carrier proteins were expressed in the apo (nonpantetheinylated) form in BL21(DE3) *E. coli* cells under control of a T7 promoter at 25 °C and purified by nickel chromatography. Apo carrier proteins were modified by incubation with the broad-specificity phosphopantetheinyl transferase Sfp[14] and acetyl-CoA or methylthiazolyl-CoA. The Ybt proteins exhibited a loading capacity of 65%, as judged by quantitation of covalent modification of the protein with the radiolabeled coenzyme [³H]acetyl-CoA. EntB(EpoB) and RapC(EpoA) both exhibited approximately 50% modification, whereas the wild-type counterparts RapC and EntB showed greater than 90% priming. EpoB and EpoC were expressed and purified as reported previously.[3,4] In a representative assay, apo-EntB(EpoB) (5 μM) in a total volume of 100 μL was incubated at pH 7 with Sfp (0.5 μM) and methylthiazolyl-CoA (200 μM) in the presence of MgCl₂ (5 mM). EpoC (5 μM), NADPH (3 mM), and [¹⁴C]methylmalonyl-CoA (10 μM) (total reaction volume 65 μL) were then added, and the reaction was left for 20 min at 25 °C. The reaction was quenched by the addition of trichloroacetic acid (10%), and the product was hydrolyzed from the precipitated protein

by heating with KOH (0.5 m, 60 °C) for 5 min. The methylthiazole methylacrylic acid **1** and picolinic methylacrylic acid **2** standards were made as previously reported and used to verify product formation in the reaction.[4, 15] Reversed-phase HPLC (C18 Vydac column) with a UV detector and an online radioisotope detector was used for the assays. (**1**: MeCN (0–70%) in aqueous TFA (0.1%), 25 min; **2**: MeCN (0–25%) in aqueous TFA (0.1%), 25 min.) The formation of **1** was also assayed by radio-TLC (silica; 90:5:5 chloroform/ethanol/acetic acid). Rates for EntE activation were determined by using a PPi (inorganic pyrophosphatase) exchange assay as previously described.[7a] Quantitation of the loading onto EntB by EntE was performed by loading with [$^{14}$C]salicylate.

[1] D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, C. M. Woods, *Cancer Res.* **1995**, *55*, 2325–2333.

[2] a) L. Tang, S. Shah, L. Chung, J. Carney, L. Katz, C. Khosla, B. Julien, *Science* **2000**, *287*, 640–642; b) I. Molnar, 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.

[3] H. Chen, S. O'Connor, D. E. Cane, C. T. Walsh, *Chem. Biol.* **2001**, *8*, 899–912.

[4] S. E. O'Connor, H. Chen, C. T. Walsh, *Biochemistry* **2002**, *41*, 5685–5694.

[5] T. L. Schneider, C. T. Walsh, S. E. O'Connor, *J. Am. Chem. Soc.* **2002**, *124*, 11272–11273.

[6] A. Konig, T. Schwecke, I. Molnar, G. A. Bohm, P. A. Lowden, J. Staunton, P. F. Leadley, *Eur. J. Biochem.* **1997**, *247*, 526–534.

[7] a) A. M. Gehring, K. A. Bradley, C. T. Walsh, *Biochemistry* **1997**, *36*, 8495–8503; b) F. Rusnak, W. S. Faraci, C. T. Walsh, *Biochemistry* **1989**, *28*, 6827–6835.

[8] a) D. A. Miller, L. Luo, N. Hillson, T. A. Keating, C. T. Walsh, *Chem. Biol.* **2002**, *9*, 333–344; b) H. Drechsel, H. Stephan, R. Lotz, H. Haag, H. Zaehner, K. Hantke, G. Jung, *Liebigs Ann.* **1995**, 1727–1733; c) A. M. Gehring, E. DeMoll, J. D. Fetherston, I. Mori, G. F. Mayhew, F. R. Blattner, C. T. Walsh, R. D. Perry, *Chem. Biol.* **1998**, *5*, 573–586.

[9] a) R. S. Gokhale, S. Y. Tsuji, D. E. Cane, C. Khosla, *Science* **1999**, *284*, 482–485; b) S. Y. Tsuji, D. E. Cane, C. Khosla, *Biochemistry* **2001**, *40*, 2326–2331; c) N. Wu, S. Y. Tsuji, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2001**, *123*, 6465–6474; d) N. Wu, D. E. Cane, C. Khosla, *Biochemistry* **2002**, *41*, 5056–5066; e) P. Kumar, Q. Li, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **2003**, *125*, 4097–4102.

[10] L. Du, C. Sanchez, B. Shen, *Metab. Eng.* **2001**, *3*, 78–95.

[11] A previously reported definition of a PCP domain was used in this study to separate the linker region (the sequence after the fourth helix) from the minimal PCP domain: T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407–418.

[12] K. C. Nicolaou, R. Scarpelli, B. Bollbuck, B. Werschkun, M. M. A. Pereira, M. Wartmann, K. H, Altmann, D. Zaharevitz, R. Gussio, P. Giannakakou, *Chem. Biol.* **2000**, *7*, 593–599.

[13] R. M. Horton, Z. Cai, S. N. Ho, L. R. Pease, *BioTechniques* **1990**, *8*, 528.

[14] a) L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595; b) P. J. Belshaw, C. T. Walsh, T. Stachelhaus, *Science* **1999**, *284*, 486–489.

[15] Characterization of **2**. Reversed-phase HPLC (MeCN (0–25%) in TFA (0.1%) over 25 min, C18 Vydac column): retention time = 14.6 min; MALDI-MS: calcd: 178.08, found: 178.06; $^{1}$H NMR (200.055 MHz, CDCl$_3$): $\delta$ = 8.63 (1 H, d, $J$ = 4.81 Hz; Ar), 7.76 (1 H, d, $J$ = 1.40 Hz; Ar), 7.09 (1 H, d, $J$ = 4.80 Hz; Ar), 2.40 (3 H, s; CH$_3$-Ar), 2.30 (3 H, d, $J$ = 1.60 Hz; CH=C(C*H*$_3$)), 2.16 ppm (1 H, d, $J$ = 0.80 Hz; C*H*=C); $^{13}$C NMR (50.31 MHz, CDCl$_3$): $\delta$ = 172.6, 154.5, 148.9, 148.5, 137.3, 132.7, 126.8, 124.1, 21.39, 14.35 ppm.