Biosynthesis

International Edition: DOI: 10.1002/anie.201502902
German Edition: DOI: 10.1002/ange.201502902

Discovery of a Single Monooxygenase that Catalyzes Carbamate Formation and Ring Contraction in the Biosynthesis of the Legonmycins

Sheng Huang, Jioji Tabudravu, Somayah S. Elsayed, Jeanne Travert, Doe Peace, Ming Him Tong, Kwaku Kyeremeh, Sharon M. Kelly, Laurent Trembleau, Rainer Ebel, Marcel Jaspars, Yi Yu, and Hai Deng**

Abstract: *Pyrrolizidine alkaloids (PAs) are a group of natural products with important biological activities. The discovery and characterization of the multifunctional FAD-dependent enzyme LgnC is now described. The enzyme is shown to convert indolizidine intermediates into pyrrolizidines through an unusual ring expansion/contraction mechanism, and catalyze the biosynthesis of new bacterial PAs, the so-called legonmycins. By genome-driven analysis, heterologous expression, and gene inactivation, the legonmycins were also shown to originate from non-ribosomal peptide synthetases (NRPSs). The biosynthetic origin of bacterial PAs has thus been disclosed for the first time.*

Pyrrolizidines are a group of heterocyclic compounds that consist of two fused five-membered rings with a nitrogen atom at the bridgehead position.[1] Naturally occurring pyrrolizidine alkaloids are mainly produced by plants as a defense mechanism against insect herbivores.[2] More than 660 PAs and derivatives have been found in over 6000 plants worldwide, and half of the PAs are hepatotoxic and carcinogenic.[3] Labeling studies indicated that putrescine is the common precursor of plant PAs.[2] In contrast to the widely distributed plant PAs, only 10 bacterial PAs have been discovered thus far, including the clazamycins,[4] bohemamines,[5] and jenamidines (Supporting Information, Figure S1).[6] Bacterial PAs possess promising antibacterial and antitumor activities[6] and have attracted considerable interest from academic research groups and the pharmaceutical industry.[7] However, the biosynthetic origin of bacterial PAs has not been disclosed thus far.

Herein, we report the identification of two new bacterial PAs, legonmycin A (**1**) and B (**2**; Figure 1 A), in the soil bacterium *Streptomyces* sp. MA37, a Ghanaian isolate that has the unique capacity of producing a range of fluorinated metabolites.[8] We demonstrate that only four genes, namely three non-ribosomal peptide synthetase (NRPS) genes and one that encodes for oxidative functions, are required for the biosynthesis of legonmycins by means of bioinformatic

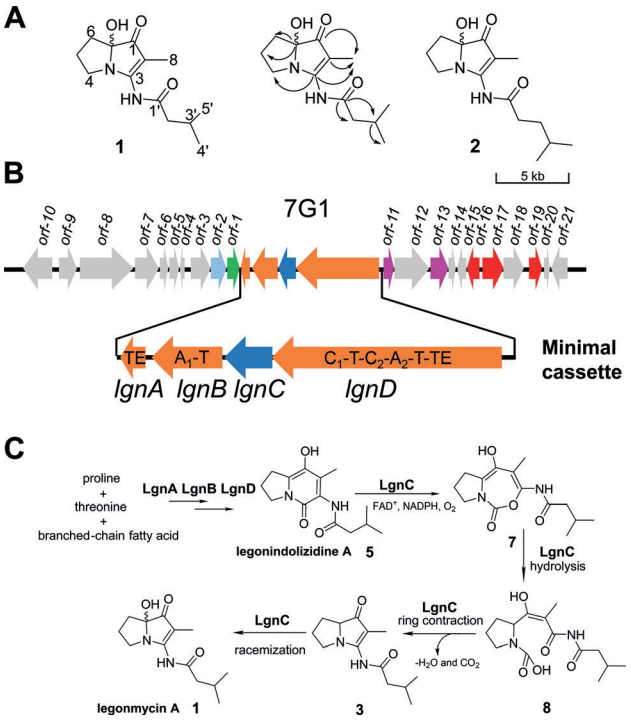


Figure 1. A) The structures of two new pyrrolizidine alkaloids, legonmycin A (**1**) and B (**2**). B) Genetic organization in the cosmid 7G1. The minimal cassette of the *lgn* biosynthetic gene cluster responsible for the production of **1** and **2** is highlighted. C) Proposed biosynthetic scheme for the transformation of indolizidine intermediate **5** into legonmycin A (**1**) catalyzed by LgnC.

[*] S. Huang,[+] Dr. Y. Yu
Key Laboratory of Combinatory Biosynthesis and Drug Discovery (Ministry of Education), School of Pharmaceutical Sciences
Wuhan University
185 East Lake Road, Wuhan 430071 (P.R. China)
E-mail: yu_yi@whu.edu.cn
Dr. J. Tabudravu,[+] S. S. Elsayed,[+] J. Travert, D. Peace, M. H. Tong, Dr. L. Trembleau, Dr. R. Ebel, Prof. Dr. M. Jaspars, Dr. H. Deng
Department of Chemistry, University of Aberdeen
Aberdeen (UK)
E-mail: h.deng@abdn.ac.uk
Dr. K. Kyeremeh
Department of Chemistry, University of Ghana
P.O. Box LG56, Legon-Accra (Ghana)
Dr. S. M. Kelly
Institute of Molecular, Cell and Systems Biology
College of Medical, Veterinary and Life Sciences
University of Glasgow
Glasgow, G12 8QQ (UK)
[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502902.

analysis, whole pathway expression, and gene deletion (Figure 1B). We further propose that a pyrrolizidine ring is formed during legonmycin biosynthesis.

A chemical investigation using AntiBase[9] as the dereplication method revealed that MA37 produces two new metabolites. Fermentation and isolation afforded two pure compounds, namely **1** (5 mg) and **2** (0.3 mg). The structure of **1** was deduced by HR-ESI mass spectrometry, 1D and 2D NMR spectroscopy, and theoretical calculations using the ACD/Labs Structural Elucidator software (Figures S2A–D, S3, and S4 and Tables S1, S2).[10] CD spectroscopy and HPLC analyses on a chiral stationary phase indicated that **1** was isolated as a racemate (Figure S5, see the Supporting Information for further details).[11] Comparison of the molecular formulas, UV maxima, and HR-ESI-MS, ^{1}H, and ^{13}C NMR spectra of **1** and **2** indicated that **2** contained one extra methylene group to form the isocaproic acid moiety (Figure S3E, F and Table S3).[9] These results confirmed **1** and **2** to be new PAs (see the Supporting Information), which were named legonmycin A and B, respectively, because of their association with Legon, Ghana.

Inspection of the structures of **1** and **2** led to speculation that the precursors of the pyrrolizidine core of the legonmycins are proline and dehydrobutyrine, a dehydrated form of threonine. In silico analysis of the draft genome of the MA37 strain using antiSMASH[12] allowed the identification of a putative gene cluster (*lgn*). The center of this cluster encodes two multidomain NRPS (LgnB and LgnD, respectively; Figure 1B). LgnD possesses an unusual arrangement of C-T-C-A-T-TE domains (Figure 1B). The A domains of LgnB and LgnD were predicted to activate threonine and proline,[12] respectively, which is consistent with the predicted precursors (Table S4).

To confirm the identity of the *lgn* gene cluster, we performed heterologous expression to produce **1** and **2** in a surrogate host. To this end, a genomic cosmid library of MA37 was constructed using an integrative *Streptomyces–E. coli* shuttle vector, pJTU2554.[13] PCR-based screening allowed for identification of one cosmid, 7G1, that contained the full length of the *lgn* cluster, which was then introduced into *Streptomyces albus*. The resultant transformant, *S. albus*:7G1 (WDY601), was cultivated. Comparative metabolic profiling of the extracts of WDY601 and the control strain *S. albus*:pJTU2554 (WDY502) allowed for the identification of two compounds that are present in the extract of WDY601 (Figure 2A), but not in that of WDY502 (Figure 2C). HR-ESI-MS analysis revealed two ions with the same molecular weights, which correspond to protonated **1** and **2** ($[M+H]^+$, Figure S6). We also identified two minor metabolites, **3** and **4** (Figure 2A), which share distinguishable UV profiles with the legonmycins. The molecular formulas of **3** and **4** were established as $C_{13}H_{21}O_2N_2^+$ and $C_{14}H_{23}O_2N_2^+$, respectively. Although these two compounds could not be isolated, their masses and MS^2 fragmentation patterns are consistent with the structures of **3** and **4**, the predicted derivatives of **1** and **2**, respectively (Scheme 1), which lack a hydroxy group at the C7 position of the pyrrolizidine ring (Figure S7). Inactivation of the *lgnD* gene completely abolished the production of **1** and **2** (Figure 2B), suggesting that the cosmid 7G1 contained the intact *lgn* cluster.

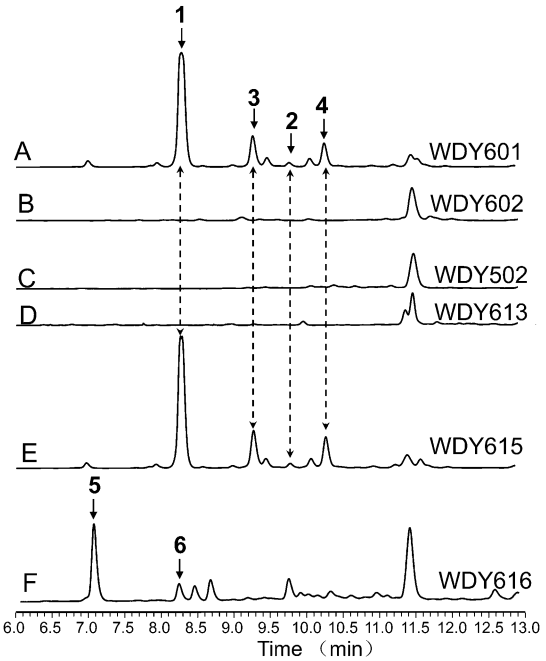
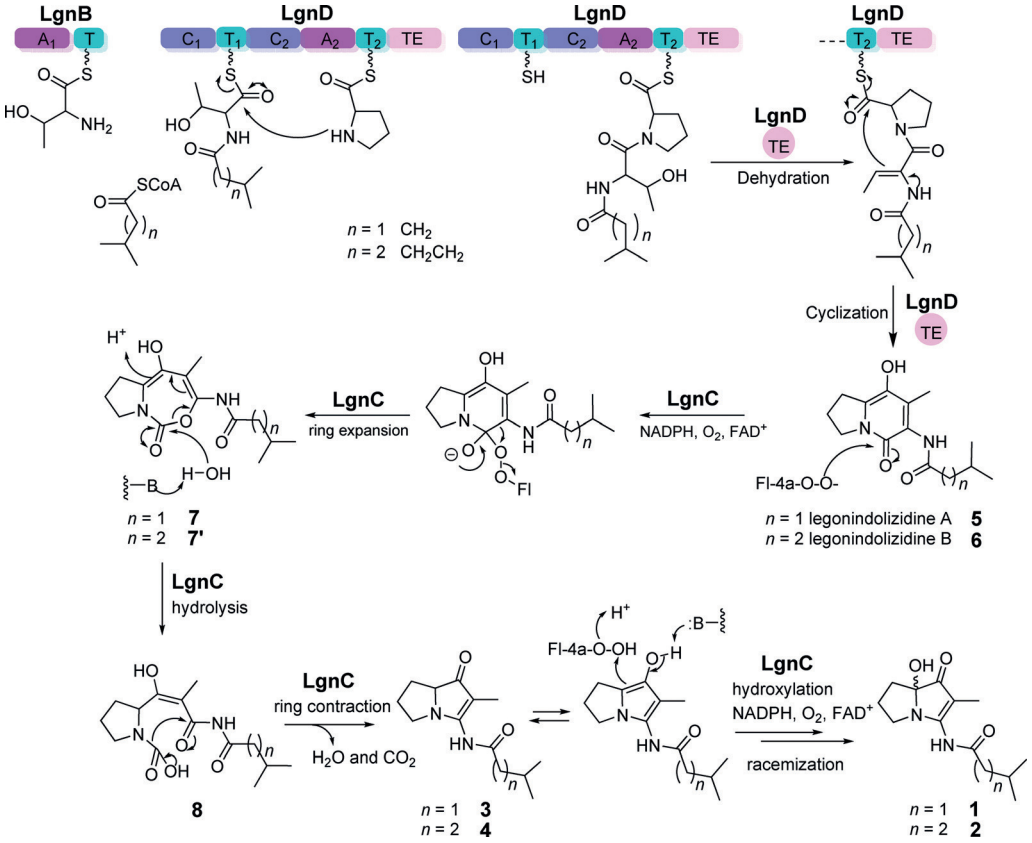


Figure 2. HPLC analysis (UV at 340 nm) of the production of compounds **1**–**6** from cultures of mutant strains of *S. albus*. A) *S. albus*:7G1 (WDY601), B) *S. albus*:7G1 *ΔlgnD* (WDY602), C) the control strain, *S. albus*:empty vector pJTU2554 (WDY502), D) *S. albus*:7G1 *ΔlgnA* (WDY613), E) *S. albus*:minimal cassette (WDY615); F) *S. albus*:minimal cassette *ΔlgnC* (WDY616).

To define the boundary of the gene cluster, in-frame deletion experiments were performed. To our surprise, knockout of most of the *orf* genes in 7G1 did not perturb production of **1** and **2** (Figure S8), whereas a *lgnA* knockout completely abolished the production of **1** and **2** (Figure 2D). Therefore, we could show that only four genes (*lgnA* to *lgnD*) are essential for the biosynthesis. Subsequently, we designed experiments to delete all of the nonessential genes on both sides of the minimal cassette, resulting in the cosmid 7G1 derived plasmid pWDY615. This construct was then introduced into *S. albus*, generating the WDY615 strain, and compound **1** was purified from this culture as confirmed by HR-ESI-MS and ^{1}H NMR spectroscopy (Figures S9 and S10). The identity of **2** in the WDY615 extract was also confirmed by comparison of the HR-ESI-MS and MS^n patterns with those of authentic **2** (Figure S11). These results clearly suggest that the minimal cassette (Figure 1B), which encodes one standalone thioesterase (TE, LgnA), one flavin-dependent monooxygenase (LgnC), and two multidomain NRPSs (LgnB and LgnD), is responsible for the production of **1** and **2** (Table S5).

Gene inactivation of *lgnC* completely abolished the production of **1** and **2** and resulted in the accumulation of two new metabolites, **5** and **6**, in the culture of the Δ*lgnC* mutant (Figure 2F). Fermentation and isolation afforded pure **5** (2 mg) and **6** (1 mg). They were established to be indolizidine alkaloids by analysis of their HR-ESI-MS and 1D and 2D NMR spectra and theoretical calculations[10] (Figur-



Scheme 1. Proposed biosynthetic pathway for legonmycin A (**1**) and B (**2**) via the intermediary legonindolizidines **5** and **6**, respectively.

es S12–S15, Tables S6 and S7). Compounds **5** and **6** were named legonindolizidine A and B, respectively.

Given that the legonindolizidines are not further processed in the Δ*lgnC* mutant, it was hypothesized that they might be immediate substrates for LgnC. To validate this hypothesis, we set out to investigate the role of LgnC during the biosynthesis of the legonmycins. Overexpression of *lgnC* in *E. coli* (DE3) allowed for the purification of its encoded protein to near homogeneity. The resultant recombinant LgnC with a trigger factor (TF) tag was a brown-colored protein with an estimated molecular weight of 94.94 kDa as determined by SDS-PAGE analysis (Figure S16A). The stoichiometry of bound FAD^+ and LgnC was estimated to be 1:1 (Figure S16B).

Upon incubation of **5** (0.5 mM) with the recombinant enzyme (10 μM) and FAD (0.1 mM), NADPH (1 mM), and a NADPH-regenerating system (see the Supporting Information), **3** and **1** were formed, as confirmed by the determination of their exact masses and co-elution experiments with standard samples of **3** and **1** (Figure 3; see also Figure S16C–I). Product formation increased with reaction time (Figure 3). Interestingly, the formation of two intermediates, **7** and **8**, with masses ($[M+H]^+$) of 281.1487 ($\Delta = -2.5$ ppm) and 299.1590 ($\Delta = -3.8$ ppm), respectively, was observed (Figure 3 and Figure S17). The corresponding masses and MS^2 fragmentation patterns are in good agreement with the structures of the predicted intermediates **7** and **8** (Figure S17). In a control assay with boiled LgnC and in other control assays lacking either FAD^+ or NADPH, the formation of **7**, **8**, **3** or **1** was not observed. These experiments confirmed that LgnC is responsible for converting **5** into **1** in a four-step enzymatic process: 1) ring expansion by oxygen insertion to yield **7**, 2) hydrolysis of the formed carbamate to generate **8**, 3) a decarboxylation-driven ring contraction to form **3** with a new nitrogen–carbon bond, and 4) hydroxylation to generate **1** (Figure 1C). Based on bioinformatic and experimental evidence, a proposed mechanism for the biosynthesis of **1** and **2** is shown in Scheme 1.

The exact role of LgnA in the legonmycin biosynthesis, a putative standalone TEII protein, remains elusive. When *lgnA* was disrupted, the mutant strain failed to produce legonmycins (Figure 2C). Disruption of both *lgnA* and *lgnC* in a double mutant abolished the production of both the legonmycins and legonindolizidines (Figure S8M), suggesting that LgnA is involved in the early stage of legonmycin biosynthesis. LgnA may have an editing role during NRPS assembly to remove amino acids and peptides that block the NRPS.[14]

We propose that the pyrrolizidine core of **1** and **2** is assembled by two NRPS enzymes, LgnB and LgnD. The biosynthesis of **1** and **2** is initiated by the condensation of the branched odd-chain fatty acids, isovaleryl and isocaproic CoA, respectively, with threonine loaded onto the thiolation (T) domain of LgnB in a reaction catalyzed by the first condensation (C1) domain of LgnD (Figure S18). The aminoacylated product tethered to the first T domain of LgnD would then be condensed with the proline of the second C2 domain of LgnD (Scheme 1 and Figure S18). Interestingly, a closely related pathway was proposed for the biosynthesis of the lipocyclocarbamates SB-253514, SB-311009, and SB315021 in *Pseudomonas fluorescens* DSM11579.[15] In the lipocarbamate pathway (Figure S19), the serine-acylated product was proposed to undergo condensation with the tethered proline, followed by serine dehydration and likely TE-based ring closure, leading to the putative indolizidine species.[15]

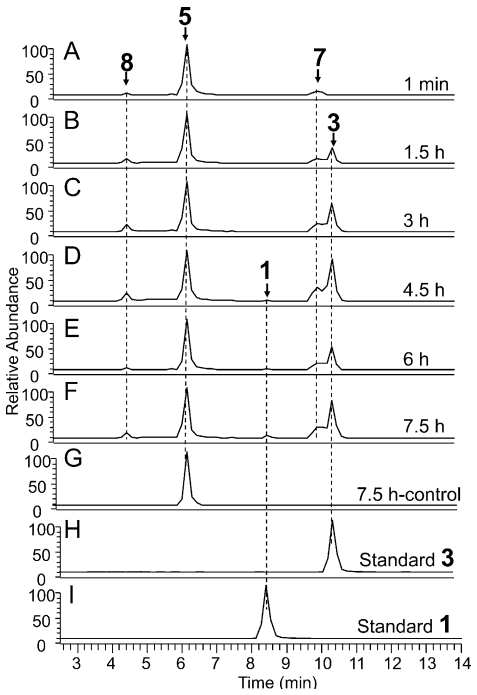


Figure 3. LC-HR-ESI-MS analysis of the products of the LgnC catalyzed reaction. The reaction mixture contains FAD^+ (0.1 mM), NADPH (1 mM), **5** (0.5 mM), LgnC (10 μM), and an NADPH regenerating system. A–F) Time-course analysis of the conversion of indolizidine **5** into carbamate **7**, carboxylate intermediate **8**, prelegonmycin A (**3**), and legonmycin A (**1**). G) Analysis of the control reaction with inactivated LgnC. H, I) Standard compounds **3** and **1**, respectively. Selective ion monitoring was carried out for $[M+H]^+$ (253.1547) of **1**, $[M+H]^+$ (237.1598) of **3**, $[M+H]^+$ (265.1547) of **5**, $[M+H]^+$ (281.1496) of **7**, and $[M+H]^+$ (299.1602) of **8**.

In the case of the legonmycin pathway, the TE domain of LgnD shares a high sequence identity (37.5%) with LpiB-TE (Figure S20). Given the similarity of the initial steps between these two pathways, LgnD-TE may catalyze the ring closures to form **5** and **6**. Phylogenetic analysis indicated that LgnD-TE was also grouped into unusual type I TE proteins (TEI; Figure S21 and Table S8), including the bifunctional domain GrsB-TE, which is involved in the biosynthesis of gramicidin S, the EntF-TE domain for the biosynthesis of enterobactin, and the Frc9-TE2 domain for the biosynthesis of FR901464, a potent anticancer agent from *Pseudomonas* sp. 2663.[16] In particular, biochemical evidence demonstrated that Frc9-TE2 catalyzes the dehydration of a D-hydroxy group in the growing polyketide chain to form a *cis* double bond during FR901464 biosynthesis.[16] Therefore, LgnD-TE may also enable the dehydration of a tethered threonine moiety.

The flavin enzyme LgnC acts on **5** and **6** and catalyzes a Baeyer–Villiger (BV) ring expansion to afford carbamate intermediates with a 5,7-fused ring system. This is followed by hydrolysis and decarboxylation-driven cyclization to generate **3** and **4**, respectively (Scheme 1). Finally, LgnC may mediate a stereospecific hydroxylation at the C7 position of **3** and **4**, but the corresponding products could readily undergo racemization at neutral pH,[4b,d] leading to **1** and **2** (Figure S5C). Some BV enzymes are known to mediate three-step cascade reactions consisting of oxygen insertion, hydrolysis, and decarboxylation. This is exemplified by the maturation of the tetracyclic intermediate premithramycin B into the tricyclic framework of mithramycin DK.[17] In that case, a biocatalytic BV reaction is used for the cleavage of the cyclohexadione ring in premithramycin B by oxygen insertion, followed by hydrolysis and facile decarboxylation.[17]

In conclusion, we have discovered two new PAs, legonmycin A and B, from *Streptomyces* sp. MA37 and shown that they are biosynthesized by an NRPS machinery. The Baeyer–Villiger enzyme LgnC was shown to catalyze a multistep process that is unprecedented in natural product biosynthesis. A plausible biosynthetic pathway for the legonmycins was thus proposed.

Acknowledgements

Y.Y. acknowledges financial support from the "973" Program (2012CB721006) and the National Natural Science Foundation of China (81102357). M.J., R.E., K.K., and H.D. acknowledge the Leverhulme Trust-Royal Society Africa (AA090088). J.T., R.E., M.J., Y.Y., and H.D. are grateful for financial support through the EU Seventh Framework Programme (312184). S.S.E. thanks the Egyptian Government for financial support of the PhD studies. We thank Dr. Richard Hodgson, Phenomenex, UK for HPLC analysis.

Keywords: biosynthesis · legonmycins · multifunctional enzymes · non-ribosomal peptide synthetases · pyrrolizidine alkaloids

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 12697–12701
Angew. Chem. **2015**, *127*, 12888–12892

[1] a) L. Smith, C. Culvenor, *J. Nat. Prod.* **1981**, *44*, 129–152; b) J. Robertson, K. Stevens, *Nat. Prod. Rep.* **2014**, *31*, 1721–1788.
[2] a) D. J. Robins, J. R. Sweeney, *J. Chem. Soc. Perkin Trans. 1* **1981**, 3083–3086; b) D. Ober, E. Kaltenegger, *Phytochemistry* **2009**, *70*, 1687–1695.
[3] a) H. Wiedenfeld, A. Andrade-Cetto, *Phytochemistry* **2001**, *57*, 1269–1271; b) R. Schoental, *Cancer Res.* **1968**, *28*, 2237–2246.
[4] a) L. Dolak, C. DeBoer, *J. Antibiot.* **1980**, *33*, 83; b) Y. Horiuchi, S. Kondo, T. Ikeda, D. Ikeda, K. Miura, M. Hamada, T. Takeuchi, H. Umezawa, *J. Antibiot.* **1979**, *32*, 762; c) H. Nakamura, Y. Iitaka, H. Umezawa, *J. Antibiot.* **1979**, *32*, 765; d) D. D. Buechter, D. E. Thurston, *J. Nat. Prod.* **1987**, *50*, 360–367.
[5] a) T. W. Doyle, D. E. Nettleton, D. M. Balitz, J. E. Moseley, R. E. Grulich, T. McCabe, J. Clardy, *J. Org. Chem.* **1980**, *45*, 1324–1326; b) Q. Zhang, K. K. Schrader, H. N. ElSohly, S. Takamatsu, *J. Antibiot.* **2003**, *56*, 673–681; c) T. S. Bugni, M. Woolery, C. A. Kauffman, P. R. Jensen, W. Fenical, *J. Nat. Prod.* **2006**, *69*, 1626–1628.
[6] a) J. F. Hu, D. Wunderlich, R. Thiericke, H. M. Dahse, S. Grabley, X. Z. Feng, L. Sattler, *J. Antibiot.* **2003**, *56*, 747–754; b) J. R. Duvall, F. Wu, B. B. Snider, *J. Org. Chem.* **2006**, *71*, 8579–8590.
[7] K. Stevens, A. J. Tyrrell, S. Skerratt, J. Robertson, *Org. Lett.* **2011**, *13*, 5964–5967.
[8] a) D. O'Hagan, H. Deng, *Chem. Rev.* **2015**, *115*, 634–649; b) H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith, D. O'Hagan,

ChemBioChem **2014**, *15*, 364 – 368; c) L. Ma, A. Bartholome, M. H. Tong, Y. Yu, T. Shepherd, K. Kyeremeh, H. Deng, D. O'Hagan, *Chem. Sci.* **2015**, *6*, 1414 – 1419.
[9] H. Laatsch, *AntiBase 2013: The Natural Compound Identifier*, Wiley-VCH, Weinheim, **2013**.
[10] Structure ACD/Structure Elucidator, **2014**, http://www.acdlabs.com.
[11] A. Rodger, B. Nordén, *Circular Dichroism and Linear Dichroism*, **1997**, p. 69.
[12] K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, T. Weber, *Nucleic Acids Res.* **2013**, W204 – 12.
[13] L. Li, Z. Xu, X. Xu, J. Wu, Y. Zhang, X. He, T. M. Zabriskie, Z. Deng, *ChemBioChem* **2008**, *9*, 1286 – 1294.
[14] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255.
[15] C. W. Johnston, R. Zvanych, N. Khyzha, N. A. Magarvey, *ChemBioChem* **2013**, *14*, 431 – 435.
[16] a) K. M. Hoyer, C. Mahlert, M. A. Marahiel, *Chem. Biol.* **2007**, *14*, 13 – 22; b) Z. Zhou, J. R. Lai, C. T. Walsh, *Chem. Biol.* **2006**, *13*, 869 – 879; c) H.-Y. He, M.-C. Tang, F. Zhang, G.-L. Tang, *J. Am. Chem. Soc.* **2014**, *136*, 4488 – 4491.
[17] a) C. T. Walsh, T. A. Wencewicz, *Nat. Prod. Rep.* **2013**, *30*, 175 – 200; b) M. P. Beam, M. A. Bosserman, N. Noinaj, M. Wehenkel, J. Rohr, *Biochemistry* **2009**, *48*, 4476 – 4487.

Received: March 30, 2015
Revised: June 9, 2015
Published online: July 17, 2015

