Organic LETTERS

Letter

pubs.acs.org/OrgLett

Cystomanamides: Structure and Biosynthetic Pathway of a Family of Glycosylated Lipopeptides from Myxobacteria

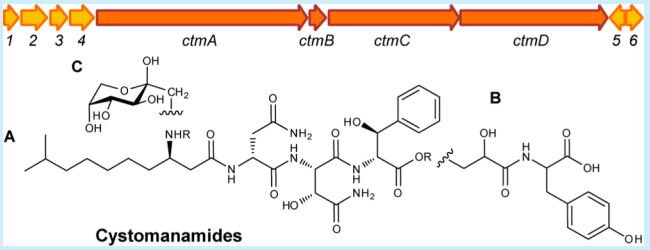
Lena Etzbach, Alberto Plaza, Ronald Garcia, Sascha Baumann, and Rolf Müller*

Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Pharmaceutical Biotechnology, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany

German Center for Infection Research (DZIF), Partner site Hannover-Braunschweig 38124, Germany

Supporting Information

ABSTRACT: Cystomanamides A–D were isolated as novel natural product scaffolds from *Cystobacter fuscus* MCy9118, and their structures were established by spectroscopic techniques including 2D NMR, LC-SPE-NMR/-MS, and HR-MS. The cystomanamides contain β-hydroxy amino acids along with 3-amino-9-methyldecanoic acid that is *N*-glycosylated in cystomanamide C and D. The gene cluster for cystomanamide biosynthesis was identified by gene disruption as PKS/NRPS hybrid incorporating an iso-fatty acid as starter unit and including a reductive amination step at the interface of the PKS and NRPS modules.



Myxobacteria have proven to exhibit a fascinating capacity to produce chemically intriguing natural products which often show unique structural elements rarely produced by other sources.[1] Myxobacterial secondary metabolites belong to multiple structural classes, and many of these diverse compounds originate from mixed polyketide-nonribosomal peptide biosynthetic pathways.[2] Complex multimodular enzymes involving various catalytic and structural domains are responsible for the biosynthesis of these versatile structures.[3] The ability to generate a wide range of complex natural products is expanded by enzymes introducing β-branching and post-PKS and NRPS reactions such as hydroxylation, glycosylation, and epimerization.[4] Our discovery strategy to determine natural products with novel structural frameworks includes UHPLC/HRMS-based metabolomics for strain selection and dereplication as well as hyphenated chromatographic methods such as LC-SPE-NMR/-MS for isolation driven by chemical and structural features. This approach indicated the presence of unusual peptides in an extract of *Cystobacter fuscus* MCy9118 and finally led to the isolation of a new family of lipopeptides, the cystomanamides (ctm). These new linear peptides comprise exclusively nonproteinogenic amino acids and bear an unusual 3-amino-9-methyldecanoic acid residue at the N-terminus. Cystomanamides C and D contain an *N*-linked glycosylation, which is one of the rare examples of late-stage modification reactions of myxobacterial metabolites.[5] In addition to these four new myxobacterial natural products, the strain was also found to produce the known antibiotics althiomycin,[6,7] roimatacene,[8] and myxochelin A[9] and B.[10]

HRESIMS of cystomanamide A (**1**) displayed an $[M + H]^+$ peak at *m/z* 609.3248 (calcd for $C_{28}H_{45}N_6O_9$, 609.3243), consistent with the molecular formula $C_{28}H_{44}N_6O_9$ containing 10 double-bond equivalents (DBE). The ^1H NMR spectrum of cystomanamide B in CD_3OD exhibited signals characteristic of a peptide including three α-proton signals at δ 4.77 (1H, dd, *J* = 7.4, 6.3 Hz), 4.80 (1H, d, *J* = 3.75 Hz), and 4.49 (1H, d, *J* = 3.25 Hz). Moreover, a downfield pair of triplets at δ 7.28 (2H, t, *J* = 7.45 Hz) and 7.20 (1H, t, *J* = 7.45 Hz) and a doublet at δ 7.34 (2H, d, *J* = 7.45 Hz) were observed (see Table S4, Supporting Information). The HSQC spectrum revealed the presence of several methylenes between δ_H 1.2 and 2.9 and two oxygenated methines at δ_C 72.8/δ_H 4.29 (β-OH-Asn) and δ_C 74.9/δ_H 5.22 (β-OH-Phe). A detailed analysis of the 2D NMR data obtained from HSQC, HMBC, COSY, and TOCSY experiments indicated the presence of asparagine, β-hydroxyasparagine (β-OH-Asn), and β-hydroxyphenylalanine (β-OH-Phe). Additionally, a 3-amino-9-methyldecanoic acid residue (AMDA) was identified, and its partial structure was deduced as follows. A sequential spin system starting from two methyl groups at δ 0.89 (Me-10 and Me-11_{AMDA}) attached to a methine at δ 1.55 (H-9_{AMDA}) and comprising five sequential methylene groups H_2-8 to H_2-4 was obtained from COSY and TOCSY correlations. COSY correlations extended this fragment by an additional aminomethine group at δ 3.54 (H-3_{AMDA}) followed by a methylene at δ 2.68, 2.50 (H_2-2_{AMDA}). Finally, an HMBC correlation from H_2-2_{AMDA} to a carbonyl resonance at δ 173.2 (C-1_{AMDA}) clearly indicated that AMDA represents a β-amino acid. The sequence was established through HMBC correlations from α-protons to carbonyl carbons of adjacent residues to result in the sequence: AMDA-Asn-(β-OH-Asn)-(β-OH-Phe). Long-range HBMC correlations from the α-proton at δ 4.77 (H-2_{Asn}) to C-1_{AMDA}

Received: March 14, 2014
Published: April 15, 2014

   dx.doi.org/10.1021/ol500779s | *Org. Lett.* 2014, 16, 2414–2417

connected AMDA to the N-terminus of asparagine, thereby completing the structure of **1** as linear tetrapeptide.

HRESIMS of cystomanamide B (**2**) displayed an $[M + H]^+$ peak at m/z 860.4039 (calcd for $C_{40}H_{58}N_7O_{14}$, 860.4036), consistent with the molecular formula $C_{40}H_{57}N_7O_{14}$ containing 16 DBE. The ^{1}H NMR spectra of **2** in comparison to that of **1** exhibited additional signals for one α-proton at δ 4.43 (1H, dd, J = 7.5, 5.5 Hz) and two downfield doublets at δ 7.05 (2H, d, J = 8.4 Hz) and δ 6.68 (2H, d, J = 8.4 Hz). The HSQC spectrum revealed the additional presence of one methylene at δ_C 38.2/δ_H 3.12, 2.97 (Tyr), one oxygenated methylene at δ_C 67.9/δ_H 4.32, 4.16 (glyceric acid GA), and one oxygenated methine at δ_C 71.0/δ_H 4.24 (GA). This evidence in combination with the DQF-COSY and HMBC correlations indicated that in comparison to **1**, cystomanamide B bears an additional glyceric acid (GA) and tyrosine residue. HMBC long-range correlations from α-protons to carbonyl carbons of adjacent residues and from the methylene protons H_2-3_{GA} to C-$1_{\beta\text{-OH-Phe}}$ resulted in the linear depsipeptide **2** with the sequence AMDA-Asn-(β-OH-Asn)-(β-OH-Phe)-GA-Tyr.

The ESI-MS/MS fragmentation patterns of cystomanamide C (**3**) and D (**4**) indicated a glycosylation. More precisely, **3** showed the loss of a 162 mass unit fragment compared to **1**, suggesting the presence of a hexose residue. The ^{1}H NMR and HSQC spectra in comparison with those of **1** showed various additional signals between 3.1 and 4.2 ppm with different intensities belonging to three rotamers of a sugar moiety. Based on HMBC and ROESY NMR data, the sugar residue was identified as fructose with β-D-fructopyranose as the most abundant rotamer, a sugar moiety also found in kwansonine A and B.[11] The HMBC spectrum showed a key correlation from the methylene at δ 3.27 (H-1_{FRU}) to the aminomethine at δ 3.59 (H-3_{AMDA}) indicating an *N*-linked glycosylation of the β-amino fatty acid chain. Brabantamides are another example for compounds showing a sugar attached to a fatty acid derived moiety.[12] Cystomanamide D showed similar NMR and MS data with a mass difference of 16 Da compared to **3** that was explained by the absence of the β-hydroxylation of the phenylalanine residue. The absolute configuration of the amino acid residues was elucidated by MS detected chromatographic analysis of the L- and D-FDLA (1-fluoro-2,4-dinitrophenyl-5-L/D-leucinamide) derivatives of the acid hydrolysate of cystomanamides A−C and comparison with respective standards.[13] The amino acid residues were assigned as D-asparagine, L-*erythro*-β-OH-asparagine, D-*threo*-β-OH-phenylalanine and D-tyrosine. The *R* configuration of C-3 of the AMDA residue was also determined by this method. β-Amino fatty acids derivatized with Marfey's reagent show a behavior analogue to α-amino acids where the L series elutes earlier than the D series.[14] The D configurations of glyceric acid in **2** and the fructose in **3** were both established by chiral HPLC of the acid hydrolysate and comparison with respective standards.

On the basis of the chemical structures of the cystomanamides (see Figure 1), it seemed likely that these compounds are products of a PKS/NRPS hybrid megasynthetase. Genome sequence data of *C. fuscus* MCy9118 were generated using Illumina sequencing. A retrobiosynthetic approach in combination with antiSMASH 2.0[15] analysis of the draft genome sequence led to the identification of the ctm biosynthetic gene cluster. The predicted gene cluster consists of 10 open reading frames (ORFs) and has an overall GC content of 69.8%.

To further analyze the catalytic domains and the A domain substrate specificity, the open reading frames were translated

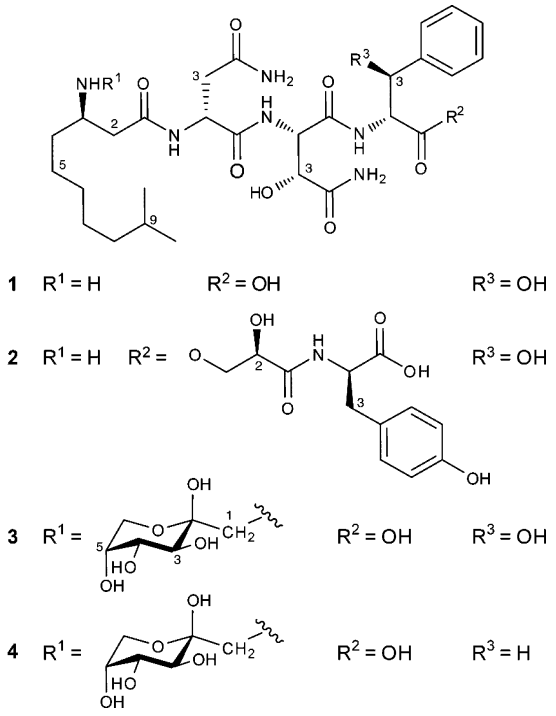


Figure 1. Structures of cystomanamide A (**1**), B (**2**), C (**3**), and D (**4**).

and analyzed using Pfam,[16] NRPS predictor2,[17] and PKS/NRPS analysis.[18] This analysis led to the prediction of one loading module, six elongation modules, and one termination module. Genes not resulting in predictions related to PKS or NRPS domains were analyzed via the BLAST algorithm[19] using the nonredundant sequence database at the National Center for Biotechnology Information (NCBI). *CtmA* encodes for a complex protein that contains functional domains belonging to fatty acid synthases, polyketide synthases, amino transferases, and nonribosomal peptide synthetases. The gene shows similarity to *mycA* encoding for mycosubtilin synthase subunit A.[20] Feeding experiments with L-[methyl-$^2H_3$]leucine indicated a leucine-derived branched chain carboxylic acid starter unit.[21] The assembly line starts with a CoA ligase domain responsible for recognition and activation of the starter molecule 9-methyldecanoic acid, an iso-odd fatty acid (see Figure 2). The activated substrate is then transferred to the first acyl carrier protein in the loading module.[22] It undergoes one elongation step using malonate as a substrate to form the β-ketothioester. The amino transferase domain (AMT) located in module 1 at the interface of the PKS and NRPS modules next reductively aminates the β-ketothioester which is then passed on to the NRPS module as shown for mycosubtilin biosynthesis.[23] This AMT shows significant (57%) similarity to the AMT found in MycA. It was proven *in vitro* that the AMT catalyzes amine transfer from an amino acid to a protein-bound β-ketothioester to generate the corresponding protein-bound β-aminothioester dependent on pyridoxal 5′-phosphate (PLP).[23] Biosynthesis continues with three NRPS based reaction cycles. *In silico* analysis of the A domain specificities is consistent with the incorporated amino acids. Modules 2, 4, and 6 contain epimerization domains that are responsible for the transformation of the L- into the respective D-amino acid. The incorporation of D-Asn, D-Phe, and D-Tyr are in accordance with the structure and the absolute configuration of **2**. The second NRPS module, module 3, is split into two proteins,

dx.doi.org/10.1021/ol500779s | *Org. Lett.* 2014, 16, 2414−2417

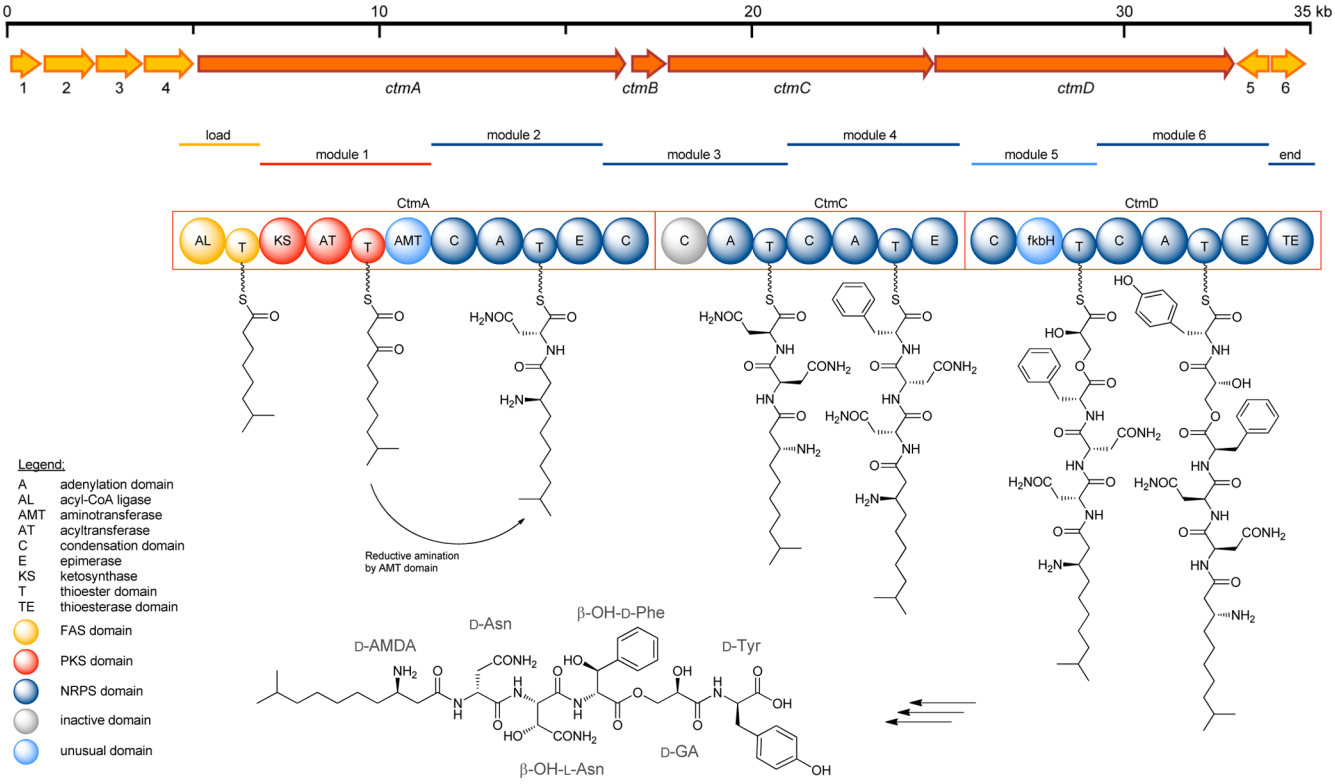


Figure 2. Gene cluster of cystomanamides in *Cystobacter fuscus* MCy9118 and the biosynthetic pathway of **2**.

CtmA and CtmC, which can be regarded as unusual but is not unprecedented.[22] The condensation domain is encoded by *ctmA* whereas the adenylation domain and the peptidyl carrier protein are encoded by *ctmC*. The module bears an additional condensation domain encoded by *ctmC*, which seems to be inactive due to the missing catalytic histidine residues in the active site.[24] Genetically, the module is separated by *ctmB* whose product is homologous to TauD from *Streptomyces auratus* (44.9%), a well-studied nonheme iron hydroxylase. CtmB contains the conserved 2-His-1-carboxylate facial triad responsible for iron binding[25,26] as well as the conserved Arg residue that ligates α-ketoglutarate.[27] It also shows similarity (35.1%) to SyrP from *Pseudomonas syringae*, which is responsible for the β-hydroxylation of an aspartyl residue in syringomycin E biosynthesis.[28] This indicates that CtmB is most likely responsible for the hydroxylation of the second asparagine residue to form L-*erythro*-β-OH-asparagine.

Module 5 has a similar organization as NRPS modules but instead of an adenylation domain it contains an unusual domain that exhibits the three conserved motifs specific for the HAD superfamily.[29] BLAST results show similarity to FkbH[30] from *Streptomyces hygroscopicus* subsp. *ascomyceticus* (29.5%) and OzmB[31] from *Streptomyces albus* (30.6%) which are responsible for the formation of glyceryl-ACP in the biosynthesis of the polyketide natural products FK520 and oxazolomycin. For OzmB it was demonstrated that it acts as a bifunctional glyceryl transferase/phosphatase that first binds D-1,3-bisphosphoglycerate from the glycolytic pool to form the D-3-phosphoglyceryl-S-OzmB intermediate. In the next step, it removes the phosphate group to receive the D-3-glyceryl-S-OzmB species (acting as a phosphatase), and finally it acts as glyceryl transferase by transferring the glyceryl group to an acyl carrier protein (ACP).[31] To the best of our knowledge, CtmD is the first protein containing such a domain integrated into an NRPS module underlining the enormous potential of these modular megaenzymes for combinatorial biosynthesis. Although the condensation domain of module 5 is homologous to the typical amide-forming condensation domains, it most likely catalyzes the formation of an ester bond instead of a peptide bond. The biochemical evidence for condensation domains being able to catalyze ester bond formation was given in studies concerning the mycotoxins fumonisins and the antitumor antibiotic C-1027.[32,33] The same mechanism with C domains embedded in elongation modules that are responsible for chain extension via ester bond formation has been proposed for the biosynthesis of some other nonribosomal peptides.[34−36] The last step in the biosynthesis of the cystomanamides is the incorporation and epimerization of tyrosine in module 6. The assembly line is terminated by a thioesterase domain that releases the product to receive the linear PKS/NRPS product **2**. The glycosylation of **3** and **4** requires enzymes encoded outside the aglycon cluster which could not be identified as 41 genes were annotated as glycosyltransferase in the genome of *Cystobacter fuscus* MCy9118.

To further confirm that the candidate gene cluster is responsible for biosynthesis of the cystomanamides, the PKS/NRPS gene *ctmA* was inactivated using a single crossover homologous recombination knockout strategy. The resulting *ctmA* mutant lost the ability to produce **1** as well as its analogues (see Figures S4 and S5, Supporting Information), therefore validating that the proposed gene cluster is indeed responsible for cystomanamide biosynthesis. To explore the relation of the surrounding open reading frames in the biosynthetic pathway, we inactivated *orf2*, *orf4* and *orf5*. Cultivation of the resulting *orf2* and *orf4* mutants and subsequent HPLC analysis revealed that inactivation of both genes abolished production of **1** and its analogues (see Figures S4 and S5, Supporting Information). The product of *orf1* shows

dx.doi.org/10.1021/ol500779s | *Org. Lett.* 2014, 16, 2414−2417

similarity to the JmjC domain containing proteins that are connected to transcription factors,[37] whereas the products of *orf2*, *orf3*, and *orf4* show similarity to the very heterogeneous group of β-lactamases. *Orf1* to *orf4* are most likely involved in the regulation of gene expression. At the same time, HPLC analysis of a cultivation of *orf5* mutants showed a cystomanamide production comparable to the wild type strain, indicating no or only a minor role of *orf5* in the biosynthesis.

The compounds were tested in various bioactivity assays including cytotoxicity against HCT-116 and CHO-K1 cells, antibacterial tests against various Gram-negative and Gram-positive bacterial strains, and antifungal assays against *Candida albicans* and *Mucor hiemalis* and HIV-1 inhibition. Until now, they have not shown biological activity. We continue functional testing to find the often very specific biological activity of these natural products and to further evaluate their biological function.

In summary, we discovered a new family of glycosylated lipopeptides using a structure-guided approach by LC-SPE-NMR. The compounds were fully characterized, and a gene cluster responsible for cystomanamide biosynthesis was identified. Inactivation of three independent genes in this cluster completely abolished cystomanamide production in the mutants, verifying their essential role during biosynthesis.

■ ASSOCIATED CONTENT

Supporting Information

Figures and tables giving configuration analysis, feeding studies, and inactivation of the ctm cluster in MCy9118 as well as experimental details, ^{1}H and ^{13}C NMR assignments, and 1D and 2D NMR spectra for **1**−**4**. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rolf.mueller@helmholtz-hzi.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Research in R.M.'s laboratory was funded by the Bundesministerium für Bildung and Forschung. We thank Kirsten Harmrolfs for support with LC-SPE-NMR and Thomas and Michael Hoffmann (all from R.M.'s laboratory) for MS measurements.

■ REFERENCES

(1) Wenzel, S. C.; Müller, R. *Curr. Opin. Drug Discovery Devel.* **2009**, *12*, 220−30.
(2) Weissman, K. J.; Müller, R. *Nat. Prod. Rep.* **2010**, *27*, 1276−95.
(3) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468−96.
(4) Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G.; Miller, D. A.; Patel, H. M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 525−34.
(5) Weissman, K. J.; Müller, R. *Bioorg. Med. Chem.* **2009**, *17*, 2121−36.
(6) Yamaguchi, H.; Nakayama, Y.; Takeda, K.; Tawara, K.; Maeda, K.; Takeuchi, T.; Umezawa, H. *J. Antibiot.* **1957**, *10*, 195−200.
(7) Bycroft, B. W.; Pinchin, R. *J. Chem. Soc., Chem. Commun.* **1975**, 121.
(8) Zander, W.; Gerth, K.; Mohr, K. I.; Kessler, W.; Jansen, R.; Müller, R. *Chem.—Eur. J.* **2011**, *17*, 7875−81.
(9) Kunze, B.; Bedorf, N.; Kohl, W.; Höfle, G.; Reichenbach, H. *J. Antibiot.* **1989**, *42*, 14−7.
(10) Ambrosi, H.-D.; Hartmann, V.; Pistorius, D.; Reissbrodt, R.; Trowitzsch-Kienast, W. *Eur. J. Org. Chem.* **1998**, *1998*, 541−551.
(11) Ogawa, Y.; Konishi, T. *Chem. Pharm. Bull.* **2009**, *57*, 1110−2.
(12) Schmidt, Y.; van der Voort, M.; Crüsemann, M.; Piel, J.; Josten, M.; Sahl, H.-G.; Miess, H.; Raaijmakers, J. M.; Gross, H. *ChemBioChem* **2014**, *15*, 259−66.
(13) Harada, K.; Fujii, K.; Mayumi, T.; Hibino, Y.; Suzuki, M.; Ikai, Y.; Oka, H. *Tetrahedron Lett.* **1995**, *36*, 1515−1518.
(14) Gerwick, W. H.; Jiang, Z. D.; Agarwal, S. K.; Farmer, B. T. *Tetrahedron* **1992**, *48*, 2313−2324.
(15) Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T. *Nucleic Acids Res.* **2013**, *41*, W204−12.
(16) Punta, M.; Coggill, P. C.; Eberhardt, R. Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; Heger, A.; Holm, L.; Sonnhammer, E. L. L.; Eddy, S. R.; Bateman, A.; Finn, R. D. *Nucleic Acids Res.* **2012**, *40*, D290−301.
(17) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. *Nucleic Acids Res.* **2011**, *39*, W362−W367.
(18) Bachmann, B. O.; Ravel, J. *Methods Enzymol.* **2009**, *458*, 181−217.
(19) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−10.
(20) Duitman, E. H.; Hamoen, L. W.; Rembold, M.; Venema, G.; Seitz, H.; Saenger, W.; Bernhard, F.; Reinhardt, R.; Schmidt, M.; Ullrich, C.; Stein, T.; Leenders, F.; Vater, J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13294−13299.
(21) Bode, H. B.; Dickschat, J. S.; Kroppenstedt, R. M.; Schulz, S.; Müller, R. *J. Am. Chem. Soc.* **2005**, *127*, 532−3.
(22) Silakowski, B.; Nordsiek, G.; Kunze, B.; Blöcker, H.; Müller, R. *Chem. Biol.* **2001**, *8*, 59−69.
(23) Aron, Z. D.; Dorrestein, P. C.; Blackhall, J. R.; Kelleher, N. L.; Walsh, C. T. *J. Am. Chem. Soc.* **2005**, *127*, 14986−7.
(24) Stachelhaus, T. *J. Biol. Chem.* **1998**, *273*, 22773−22781.
(25) Hegg, E. L.; Que, L. *Eur. J. Biochem.* **1997**, *250*, 625−9.
(26) Ryle, M. J.; Koehntop, K. D.; Liu, A.; Que, L.; Hausinger, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3790−5.
(27) Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21−68.
(28) Singh, G. M.; Fortin, P. D.; Koglin, A.; Walsh, C. T. *Biochemistry* **2008**, *47*, 11310−20.
(29) Koonin, E. V.; Tatusov, R. L. *J. Mol. Biol.* **1994**, *244*, 125−32.
(30) Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D. *Gene* **2000**, *251*, 81−90.
(31) Dorrestein, P. C.; Van Lanen, S. G.; Li, W.; Zhao, C.; Deng, Z.; Shen, B.; Kelleher, N. L. *J. Am. Chem. Soc.* **2006**, *128*, 10386−7.
(32) Zaleta-Rivera, K.; Xu, C.; Yu, F.; Butchko, R. a E.; Proctor, R. H.; Hidalgo-Lara, M. E.; Raza, A.; Dussault, P. H.; Du, L. *Biochemistry* **2006**, *45*, 2561−9.
(33) Lin, S.; Van Lanen, S. G.; Shen, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4183−8.
(34) Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. *ACS Chem. Biol.* **2006**, *1*, 766−79.
(35) Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16498−503.
(36) Xu, Y.; Orozco, R.; Wijeratne, E. M. K.; Gunatilaka, A. A.; Stock, S. P.; Molnár, I. *Chem. Biol.* **2008**, *15*, 898−907.
(37) Clissold, P. M.; Ponting, C. P. *Trends Biochem. Sci.* **2001**, *26*, 7−9.