Genomics-inspired discovery of massiliachelin, an agrochelin epimer from *Massilia* sp. NR 4-1

Jan Diettrich, Hirokazu Kage and Markus Nett[*]

Full Research Paper

Address:
Department of Biochemical and Chemical Engineering, Laboratory of Technical Biology, TU Dortmund University, Emil-Figge-Strasse 66, 44227 Dortmund, Germany

Email:
Markus Nett[*] - markus.nett@tu-dortmund.de

* Corresponding author

Abstract

A putative siderophore locus was detected in the genome of the violacein-producing bacterium *Massilia* sp. NR 4-1 and predicted to direct the biosynthesis of a molecule that is structurally related to the thiazoline-containing siderophore micacocidin. In order to track this compound, we analyzed the metabolic profiles of *Massilia* cultures grown under different iron concentrations. A compound which was found to be predominantly produced under iron deficiency was subsequently isolated. Its structural characterization by spectroscopic and bioinformatic analyses revealed a previously not known diastereomer of the cytotoxic alkaloid agrochelin. The structure of this natural product, which was named massiliachelin, corresponds to the assembly line encoded by the identified siderophore locus.

Introduction

In recent years, chemical investigations as well as genomics led to the recognition of a far greater taxonomic diversity of microbes that can produce bioactive compounds [1-3]. The exploration of previously neglected taxa has been demonstrated to bear significant potential for finding new natural products and is thus highly promising from a drug discovery perspective [4,5]. In this study, we analyzed the β-proteobacterium *Massilia* sp. NR 4-1, which had been isolated from a soil sample collected under a nutmeg tree [6]. Although, this strain had already been identified as producer of the antibiotic violacein [6], still not much is known about its secondary metabolism or

the chemistry of the genus *Massilia*, in general. Because natural product-competent microorganisms typically synthesize multiple compounds [7], strain NR 4-1 appeared as a promising candidate to find further secondary metabolites. Further incentive for the chemical analysis of this bacterium came from the discovery of novel natural products in the genera *Janthinobacterium* [8] and *Collimonas* [9,10], which are taxonomically closely related to *Massilia*.

Bioinformatic analysis of the 6.36 Mbp-sized genome of strain NR 4-1 using antiSMASH 4.0 [11,12] revealed a total of

16 biosynthesis gene clusters (BGCs), of which one (RS19155-RS19175) could be linked with the production of violacein. Another BGC bears notable similarities to a locus from the plant pathogenic bacterium *Ralstonia solanacearum* GMI1000, which is involved in the biosynthesis of the siderophore micacocidin [13,14]. Differences in the domain organization of the two corresponding biosynthetic assembly lines further indicated that *Massilia* sp. NR 4-1 does not produce micacocidin, but a derivative of this secondary metabolite. The isolation of siderophores from microorganisms is usually straightforward due to their iron-dependent production and complexing properties [15]. Therefore, we decided to initially focus our genome mining efforts on the micacocidin-type cluster in *Massilia* sp. NR 4-1. Here, we report the outcome of this study, which led to the identification of a previously unrecognized agrochelin epimer and, furthermore, unveiled the genetic basis of its biosynthesis.

Results and Discussion
The micacocidin-type BGC from *Massilia* sp. NR 4-1 and its enzymatic assembly line are depicted in Figure 1A and 1B. The gene cluster covers 34.7 kbp of contiguous DNA and includes ten genes (RS02190-RS02235), of which seven have homologs in the *mic* cluster from *R. solanacearum* GMI1000 [13]. A closer inspection of the loci shows a strong conservation of two core biosynthesis genes, namely *micC* and *micG*. On the other hand, the nonribosomal peptide synthetase (NRPS) gene *micH* is missing in the *Massilia* locus. As evidenced by biosynthetic precedence, MicH is responsible for the assembly of a thiazoline ring through condensation of a cysteine residue and subsequent cyclization [16]. The lack of MicH would hence indicate the absence of the corresponding thiazoline motif in the corresponding natural product.

Based upon the assumption that the production of the micacocidin-like compound in *Massilia* sp. NR 4-1 is iron-dependent, we analyzed the HPLC UV profiles of cultures grown under iron-deficient and iron-replete conditions. A comparison of the respective metabolic profiles revealed a distinctive peak, which was massively increased in the extract from the iron-deficient culture (Figure 1C). The corresponding metabolite, which is in the following referred to as massiliachelin (**1**), was subsequently isolated by HPLC.

High-resolution ESIMS of **1** yielded a pseudomolecular ion peak at *m/z* 467.2033 [M + H]$^+$, which indicates a molecular formula of $C_{23}H_{34}N_2O_4S_2$ and is consistent with eight degrees of unsaturation. NMR measurements confirmed the number of carbon atoms and, furthermore, revealed the presence of 31 non-exchangeable protons (Table 1). Eight carbon atoms of **1** are sp^2-hybridized according to their chemical shifts. Of these,

six could be attributed to a 2,3-substituted phenol moiety (C-1 to C-6), whereas the other two carbons exhibited resonances at 181.3 ppm (C-23) and 182.3 ppm (C-12) characteristic of carbon–heteroatom double bonds. This left two degrees of unsaturation for additional ring structures. HMBC and COSY data indicated that the phenol moiety of **1** bears an *n*-pentyl side chain in *meta*-position to its hydroxy group. Long-range correlations of H-4 and H-6 further established the linkage between C-2 and C-12. In addition, the HMBC experiment detected correlations from H-13 and H-14 to C-12, which, in combination with characteristic chemical shift values [13,17], confirmed the presence of a thiazoline substituent at C-2. Likewise, the adjacent thiazolidine moiety was determined. The ^{1}H NMR spectrum of **1** features three methyl singlets, of which one could be assigned to an *N*-methyl group of the thiazolidine by HMBC data. The remaining two methyl groups (CH$_3$-21 and CH$_3$-22) are connected to C-20. Long-range correlations of their protons also established the position of C-19 and C-23 next to C-20. The hydroxy group at C-19 and the carboxyl group at C-23 were deduced from the carbon chemical shifts of these atoms. In this way, the full planar structure of **1** was elucidated.

According to a literature search, **1** possesses the same chemical constitution as the alkaloid agrochelin (**2**, Figure 2), which was previously reported from a marine bacterium of the genus *Agrobacterium* [18]. However, we also noted some discrepancies in the observed chemical shifts, which suggested that the isolated natural product from *Massilia* sp. is not identical to agrochelin. Instead **1** and **2** are assumed to represent diastereomers. The occurrence of diastereomers is known for natural products, such as pyochelin and yersiniabactin, which are structurally very similar to agrochelin. In both, pyochelin and yersiniabactin, the stereochemical variability was attributed to the thiazoline–thiazolidine motive, which is also present in **1** and **2** [19-22]. In order to deduce the configuration of **1** we used an approach integrating bioinformatics as well as spectroscopy. First, we analyzed whether the isolated natural product possesses a D- or L-configured thiazoline ring. Previous studies had revealed that the D-thiazoline ring in pyochelin is due to an unusual methyltransferase-like epimerization domain in the biosynthesis protein PchE [23]. An analysis of the analogous enzymes in micacocidin [13], yersiniabactin [16], and enantio-pyochelin [24] biosynthesis confirmed that the presence or absence of this feature allows a reliable prediction of the stereochemistry in this position (Table S1, Supporting Information File 1). In every single case, the domain-based configurational prediction matched the experimental assignment [24-28]. By applying this method to the MicC homolog from *Massilia* sp. NR 4-1, the presence of a D-thiazoline ring in **1** can be deduced. Moreover, the thiazolidine ring in **1** must have the stereochemistry at the C2 position derived from L-cysteine. This is because
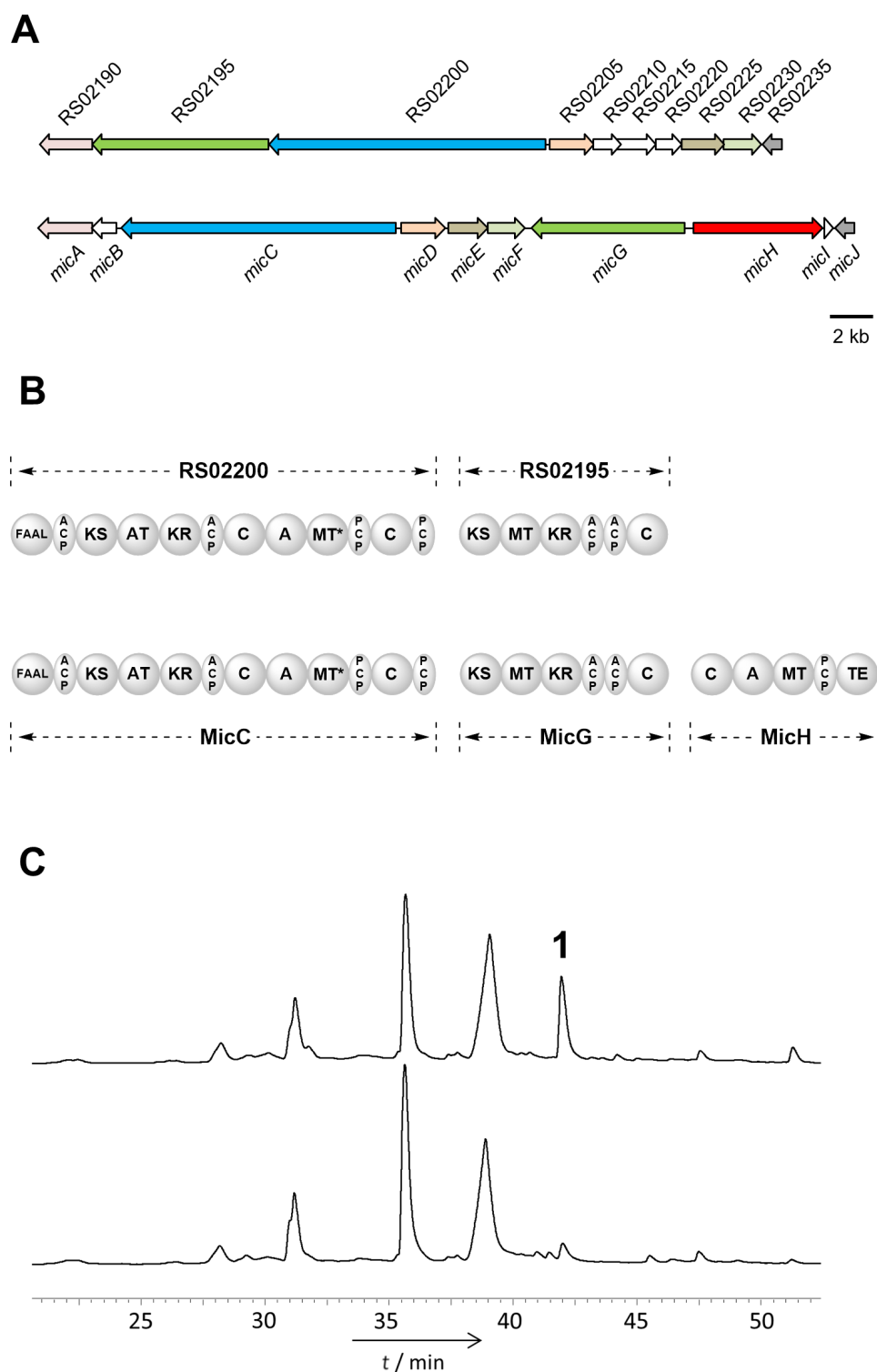
Figure 1: A) Organization of the micacocidin-type gene cluster from *Massilia* sp. NR 4-1 (top) and of the *mic* gene cluster from *R. solanacearum* GMI1000 (bottom). The open reading frames are color-coded according to their predicted function. B) Domain architecture of the massiliachelin (top) and the micacocidin (bottom) assembly lines. Domain notation: FAAL, fatty acyl-AMP ligase; ACP, acyl carrier protein; KS, β-ketoacyl synthase; AT, acyltransferase; KR, ketoreductase; C, condensation; A, adenylation; MT, methyltransferase; PCP, peptidyl carrier protein; TE, thioesterase. The asterisk indicates a methyltransferase-like epimerization domain. C) UV chromatograms of crude culture extracts from *Massilia* sp. NR 4-1 grown under iron-deficient (top) and iron-replete (bottom) conditions. The chromatograms were recorded at a wavelength of 274 nm.

Table 1: NMR spectroscopic data of **1** in CDCl$_3$.

pos.	δ_C, type	δ_H, mult. (*J* in Hz)	COSY	HMBC	NOESY
C1	158.8, C				
C2	112.6, C				
C3	144.7, C				
C4	122.2, CH	6.76, d (7.4)	5	1, 2, 3, 6, 7, 12	5, 7a, 7b, 8
C5	135.9, CH	7.33, t (8.2, 7.8)	4, 6	1, 2, 3	4, 6
C6	116.1, CH	7.22, d (8.5)	5	2, 4, 12	5
C7	35.1, CH$_2$	a: 2.92, ddd (14.4, 9.8, 6.4)	7b, 8	2, 3, 4, 8	4, 7b, 8, 9
		b: 2.76, ddd (14.4, 9.8, 6.3)	7a, 8	2, 3, 4, 8	4, 7a, 8, 9
C8	31.8, CH$_2$	1.58, m	7a, 7b, 9	3, 7, 9, 10	7a, 7b, 9, 10
C9	31.7, CH$_2$	1.33, m	8, 10	7, 11	7a, 7b, 8, 10, 11
C10	22.4, CH$_2$	1.33, m	9, 11	8	9, 11
C11	14.0, CH$_3$	0.89, t (7.1)	10	9, 10	9, 10
C12	182.3, C				
C13	33.5, CH$_2$	a: 3.72, dd (12.1, 9.9)	13b, 14	12, 14, 15	13b, 14
		b: 3.46, dd (12.1, 6.4)	13a, 14	12, 14, 15	13a, 14, 15
C14	70.7, CH	5.09, td (9.9, 6.4)	13a, 13b, 15	12, 15	13a, 13b, (15), 16
C15	80.4, CH	4.20, d (9.9)	14	13, 14, 16, 17, 18	13b, (14), 18
C16	36.1, CH$_2$	3.23, d (7.9)	17	15, 17	14, 17, 19
C17	73.1, CH	3.49, dt (7.9, 2.6)	16, 19	15, 16, 18	16, 18, 19, 22
C18	47.3, CH$_3$	2.66, s		15, 17	15, 17, 22
C19	78.6, CH	3.31, d (2.6)	17	16, 20, 21, 22, 23	16, 17, 21, 22
C20	44.3, C				
C21	26.5, CH$_3$	1.38, s		19, 20, 22, 23	19
C22	22.1, CH$_3$	1.25, s		19, 20, 21, 23	17, 18, 19
C23	181.3, C				
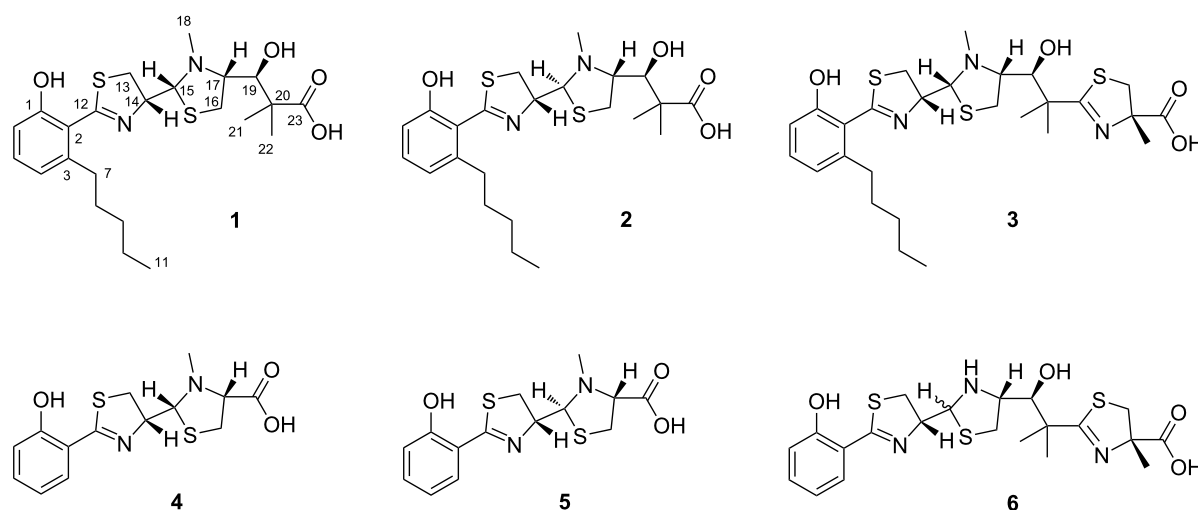
Figure 2: Structures of massiliachelin (**1**), agrochelin (**2**), micacocidin (**3**), pyochelin I (**4**), pyochelin II (**5**), and yersiniabactin (**6**).

the MicC homolog from *Massilia* sp. NR 4-1 features only a single adenylation domain for the activation of L-cysteine, which corresponds to the micacocidin and yersiniabactin assembly lines [13,16]. An inspection of the ketoreductase (KR) domain in the MicG homolog from *Massilia* sp. NR 4-1 revealed an aspartic acid residue at position 95 and a proline residue at position 144, which are both indicative for the formation of B-type alcohol stereochemistry [29]. The two motifs are

also conserved in MicG and HMWP1 from micacocidin and yersiniabactin biosynthesis (Table S2, Supporting Information File 1), which led us to infer that the absolute configuration at C-19 is *S*. It was hence possible to predict the configuration of all stereocenters in **1** except C-15 by bioinformatics. To conclude the stereochemical analysis, we resorted to the NOESY spectrum of **1**. Cañedo and co-workers had previously reported a (14*R*,15*S*,17*R*,19*S*) relative configuration for **2** [18]. As opposed to the NMR analysis of **2**, we detected only a weak NOE correlation between H-14 and H-15. Furthermore, strong NOEs were observed from the *N*-methyl protons (H$_3$-18) to both, H-17 and H-15, which is only possible if the latter two protons are *syn*-oriented. We thus propose (14*R*,15*R*,17*R*,19*S*) configuration for **1**. It is noteworthy that the same conclusion can also be drawn from a comparison of the ^{1}H NMR data of the thiazoline-thiazolidine moieties in pyochelin and agrochelin (Table S3, Supporting Information File 1).

Conclusion

In summary, *Massilia* sp. NR 4-1 was found to synthesize an epimer of the alkaloid agrochelin under iron-deficient conditions. The structure of massiliachelin is consistent with the architecture of a biosynthetic assembly line, which is encoded in the genome of this bacterium. Bioinformatic analyses greatly facilitated the stereochemical analysis and also demonstrated the usefulness of computational methods in the configurational assignment of this class of natural products.

Experimental
Analytical methods

LC–MS analyses were performed with a Nucleoshell RP18 column (150 × 2.0 mm, Macherey-Nagel) using an Agilent 1260 Infinity LC system combined with a Compact quadrupole-time of flight (Q-TOF) mass spectrometer (Bruker Daltonics). The Q-TOF mass spectrometer was interfaced with an electrospray ionization source. NMR spectra were recorded on a Bruker AV 600 MHz Avance III HD system with chloroform-*d* as solvent and internal standard. The solvent signals were referenced to δ_H 7.24 ppm and δ_C 77.0 ppm.

Cultivation and extraction of *Massilia* sp. NR 4-1

Strain NR 4-1 was cultured in three 5 L Simax flasks each containing two liters of R2A medium (0.5 g/L yeast extract, 0.5 g/L proteose peptone, 0.5 g/L casamino acids, 0.5 g/L glucose, 0.5 g/L soluble starch, 0.3 g/L sodium pyruvate, 0.3 g/L K$_2$HPO$_4$ and 0.05 g/L MgSO$_4$·7H$_2$O, pH 7.2). The cultures were shaken at 160 rpm and 30 °C. After seven days of cultivation the fermentation broth was extracted three times with ethyl acetate and the solvent was removed in vacuo to give 356 mg of dried extract.

Isolation of massiliachelin (**1**)

The extract was dissolved in 3 mL methanol and purified by reversed-phase HPLC with a Nucleodur 18 PAH column (250 × 8.0 mm, 3 μm, Macherey Nagel) using a linear gradient of methanol in water supplemented with 0.1% (v/v) trifluoro-acetic acid. The gradient conditions were as follows: 10% methanol for 5 min, from 10% to 100% over 35 min, 100% for 10 min, followed by 10% for 10 min. The flow rate was set to 1 mL/min. The elution of compounds was monitored with a diode array detector. In total, 12.0 mg of **1** were isolated.

Massiliachelin (**1**)

pale yellow solid; UV (MeOH) λ_{max} (log ε): 356 (3.06), 280 (4.18), 228 (4.64) nm; ^{1}H and ^{13}C NMR, Table 1; HRESIMS *m/z*: [M + H]$^+$, calcd for C$_{23}$H$_{35}$N$_2$O$_4$S$_2$, 467.2033; found, 467.2033.

Supporting Information

Supporting Information File 1

Additional tables and copies of NMR spectra.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-128-S1.pdf]

ORCID® iDs

Markus Nett - https://orcid.org/0000-0003-0847-086X

References

1. Masschelein, J.; Jenner, M.; Challis, G. L. *Nat. Prod. Rep.* **2017**, *34*, 712–783. doi:10.1039/c7np00010c
2. Challinor, V. L.; Bode, H. B. *Ann. N. Y. Acad. Sci.* **2015**, *1354*, 82–97. doi:10.1111/nyas.12954
3. Molloy, E. M.; Hertweck, C. *Curr. Opin. Microbiol.* **2017**, *39*, 121–127. doi:10.1016/j.mib.2017.09.006
4. Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. *Nature* **2015**, *517*, 455–459. doi:10.1038/nature14098
5. Lincke, T.; Behnken, S.; Ishida, K.; Roth, M.; Hertweck, C. *Angew. Chem., Int. Ed.* **2010**, *49*, 2011–2013. doi:10.1002/anie.200906114
6. Myeong, N. R.; Seong, H. J.; Kim, H.-J.; Sul, W. J. *J. Biotechnol.* **2016**, *223*, 36–37. doi:10.1016/j.jbiotec.2016.02.027
7. Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. *ChemBioChem* **2002**, *3*, 619–627. doi:10.1002/1439-7633(20020703)3:7<619::aid-cbic619>3.0.co;2-9
8. Graupner, K.; Scherlach, K.; Bretschneider, T.; Lackner, G.; Roth, M.; Gross, H.; Hertweck, C. *Angew. Chem., Int. Ed.* **2012**, *51*, 13173–13177. doi:10.1002/anie.201206658

9. Fritsche, K.; van den Berg, M.; de Boer, W.; van Beek, T. A.;
    Raaijmakers, J. M.; van Veen, J. A.; Leveau, J. H. J.
    Environ. Microbiol. **2014,** *16,* 1334–1345.
    doi:10.1111/1462-2920.12440
10. Kai, K.; Sogame, M.; Sakurai, F.; Nasu, N.; Fujita, M. *Org. Lett.* **2018,**
    20, 3536–3540. doi:10.1021/acs.orglett.8b01311
11. Blin, K.; Pascal Andreu, V.; de los Santos, E. L. C.; Del Carratore, F.;
    Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019,** *47,*
    D625–D630. doi:10.1093/nar/gky1060
12. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.;
    Kautsar, S. A.; Suarez Duran, H. G.; de los Santos, E. L. C.;
    Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.;
    Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H.
    Nucleic Acids Res. **2017,** *45,* W36–W41. doi:10.1093/nar/gkx319
13. Kreutzer, M. F.; Kage, H.; Gebhardt, P.; Wackler, B.; Saluz, H. P.;
    Hoffmeister, D.; Nett, M. *Appl. Environ. Microbiol.* **2011,** *77,*
    6117–6124. doi:10.1128/aem.05198-11
14. Kage, H.; Kreutzer, M. F.; Wackler, B.; Hoffmeister, D.; Nett, M.
    Chem. Biol. **2013,** *20,* 764–771. doi:10.1016/j.chembiol.2013.04.010
15. Kreutzer, M. F.; Nett, M. *Org. Biomol. Chem.* **2012,** *10,* 9338–9343.
    doi:10.1039/c2ob26296g
16. Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T.
    Chem. Biol. **2002,** *9,* 333–344. doi:10.1016/s1074-5521(02)00115-1
17. Souto, A.; Montaos, M. A.; Rivas, A. J.; Balado, M.; Osorio, C. R.;
    Rodríguez, J.; Lemos, M. L.; Jiménez, C. *Eur. J. Org. Chem.* **2012,**
    5693–5700. doi:10.1002/ejoc.201200818
18. Cañedo, L. M.; de la Fuente, J. A.; Gesto, C.; Ferreiro, M. J.;
    Jiménez, C.; Riguera, R. *Tetrahedron Lett.* **1999,** *40,* 6841–6844.
    doi:10.1016/s0040-4039(99)01377-5
19. Ankenbauer, R. G.; Cox, C. D. *J. Bacteriol.* **1988,** *170,* 5364–5367.
    doi:10.1128/jb.170.11.5364-5367.1988
20. Rinehart, K. L.; Staley, A. L.; Wilson, S. R.; Ankenbauer, R. G.;
    Cox, C. D. *J. Org. Chem.* **1995,** *60,* 2786–2791.
    doi:10.1021/jo00114a029
21. Schlegel, K.; Taraz, K.; Budzikiewicz, H. *BioMetals* **2004,** *17,* 409–414.
    doi:10.1023/b:biom.0000029437.42633.73
22. Drechsel, H.; Stephan, H.; Lotz, R.; Haag, H.; Zähner, H.; Hantke, K.;
    Jung, G. *Liebigs Ann.* **1995,** 1727–1733.
    doi:10.1002/jlac.1995199510243
23. Patel, H. M.; Tao, J.; Walsh, C. T. *Biochemistry* **2003,** *42,*
    10514–10527. doi:10.1021/bi034840c
24. Youard, Z. A.; Mislin, G. L. A.; Majcherczyk, P. A.; Schalk, I. J.;
    Reimmann, C. *J. Biol. Chem.* **2007,** *282,* 35546–35553.
    doi:10.1074/jbc.m707039200
25. Kobayashi, S.; Nakai, H.; Ikenishi, Y.; Sun, W.-Y.; Ozaki, M.;
    Hayase, Y.; Takeda, R. *J. Antibiot.* **1998,** *51,* 328–332.
    doi:10.7164/antibiotics.51.328
26. Ino, A.; Hasegawa, Y.; Murabayashi, A. *Tetrahedron* **1999,** *55,*
    10283–10294. doi:10.1016/s0040-4020(99)00583-9
27. Ino, A.; Murabayashi, A. *Tetrahedron* **2001,** *57,* 1897–1902.
    doi:10.1016/s0040-4020(01)00012-6
28. Miller, M. C.; Parkin, S.; Fetherston, J. D.; Perry, R. D.; DeMoll, E.
    J. Inorg. Biochem. **2006,** *100,* 1495–1500.
    doi:10.1016/j.jinorgbio.2006.04.007
29. Caffrey, P. *ChemBioChem* **2003,** *4,* 654–657.
    doi:10.1002/cbic.200300581