



Biosynthesis *in vitro* of bacillamide intermediate-heterocyclic AlaCys_{thiazole} by heterologous expression of nonribosomal peptide synthetase (NRPS)



Fengli Zhang, Nayila Mulati, Yukun Wang, Yingxin Li, Sanqiang Gong, Loganathan Karthik, Wei Sun, Zhiyong Li*

Marine Biotechnology Laboratory, State Key Laboratory of Microbial Metabolism and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:

Bacillus atrophaeus
Bacillamides
Heterologous expression
Nonribosomal peptide synthetase (NRPS)
Thiazole

ABSTRACT

Bacillamide C, a potential natural anti-algae active compound, is produced by *Bacillus atrophaeus* C89 derived from marine sponge *Dysidea avara*. A nonribosomal peptide synthetase (NRPS) cluster is hypothesized to be involved in the biosynthesis of bacillamide C. The NRPS with a domain string of A1-PCP1-Cy-A2-PCP2-C can be divided into three functional modules. After heterologous expression and purification of module A1-PCP1 and module Cy-A2-PCP2, their catalytic activities were biochemically proven *in vitro* by the reaction with the apo-PCP domain transformed to the holo-PCP domain through a phosphopantetheinyl transferase, ATP, and substrate amino acids. Five-membered heterocyclic AlaCys_{thiazole} with molecular weight of 172.0389 was detected. This proved the formation of the heterocyclic dipeptide AlaCys_{thiazole}, which is considered to be a building block for the biosynthesis of bacillamide. This study provides a basis for further biosynthesis of bacillamides.

1. Introduction

Bacillamides, which are nonribosomal peptides, can be categorized as bacillamides A, B, C, D, and E (Fig. S1) (Bloudoff et al., 2017; Ivanova et al., 2007; Jeong et al., 2003; Omura et al., 1975; Socha et al., 2007). In our previous study, neobacillamide A and bacillamide C were isolated from marine *Bacillus atrophaeus* C89 (Yu et al., 2009). The tryptamide thiazole (bacillamides A–D) or tryptamide thiazoline (bacillamide E) motif as a typical characteristic of bacillamides (Fig. S1), is a building block present in many bioactive cyclic peptides, such as the antibiotic zerkovamycin produced by *Streptomyces* sp. (Tabata et al., 1999), the protein synthesis inhibitors A-21459A and B from *Actinoplanes* sp. (Selva et al., 1996), and the immunosuppressive argyriins from the myxobacterium *Archangium gephyra* (Sasse et al., 2002). Similarly, bacillamides have multiple biological activities, for example, bacillamide A and its derivatives possess substantial algicidal activities against a range of dinoflagellates and raphidophytes, and exhibit anti-biosis against a particular species of cyanobacteria (Jeong et al., 2003; Churro et al., 2009) and bacillamide analogs, thus rendering them novel cytotoxic and anti-inflammatory agents (Kumar et al., 2016).

Bacillamides have been generally synthesized chemically (Fache et al., 2012; Kumar et al., 2016; Li et al., 2009; Wang et al., 2010), only one study reported the biosynthetic pathway of bacillamide E (Bloudoff

et al., 2017). Even though the biosynthesis of bacillamide C was optimized, the yield was very low (Jin et al., 2011; Yu et al., 2015). Therefore, investigating the biosynthetic mechanisms of bacillamide C is essential for its mass production and for understanding the biosynthetic pathways of other bacillamides.

Nonribosomal peptides are produced by large multifunctional enzymes (Schwarzer and Marahiel, 2001). A nonribosomal peptide synthetase (NRPS) gene cluster was detected in the genome sequence of *B. atrophaeus* C89 (Genbank No. NZ_AJRJ01000045.1), which was predicted to synthesize bacillamide C (Liu et al., 2012). Nonribosomal peptide synthesis requires at least three functional domains: the adenylation (A) domain, which selects and activates the cognate amino acid; the peptidyl carrier protein (PCP) domain, which transports the activated intermediate; and the condensation (C) domain, which catalyzes peptide bond formation, or the cyclization (Cy) domain, which catalyzes heterocycle formation (Marshall et al., 2002). In our previous study, the biosynthetic mechanism of bacillamide C was predicted (Zhang et al., 2018). An NRPS with a domain string of A1-PCP1-Cy-A2-PCP2-C can be divided into three functional modules. The A1-PCP1 module consists of the first A (A1) and the first PCP domain and activates and transports L-alanine. The Cy-A2-PCP2 module consists of a Cy, the second A (A2), and the second PCP (PCP2) domains and activates and carries L-cysteine, catalyzes peptide bond formation, and

* Corresponding author.

E-mail address: zylis@sjtu.edu.cn (Z. Li).

heterocyclizes the product 4-carboxylic acid thiazoline derivative. Moreover, the flavin mononucleotide-containing oxidase (Ox) domain oxidizes the dehydro heterocycle by a two-electron transfer to produce heteroaromatic thiazole. The 3rd module consists of the C domain and catalyzes peptide bond formation between the intermediates synthesized by the A1-PCP1 and Cy-A2-PCP2 modules, and

mass spectrometry and tandem mass spectrometry using a Waters ACQUITY Ultra Performance Liquid Chromatography system coupled with a Waters Micromass Q-TOF Premier Mass Spectrometer equipped with an electrospray interface (Waters Corporation, Milford, MA). Samples were separated using an Acquity BEH C18 column (100 mm × 2.1 mm i.d., 1.7 μm; Waters, Milford, USA), with gradient solvent A (water with 0.1% (v/v) formic acid) and solvent B (acetonitrile with 0.1% (v/v) formic acid) serving as the mobile phase at a flow rate of 0.40 mL/min. The detailed gradient conditions of the mobile phase are outlined as follows: 0–0.50 min, 2% B; 0.50–6.00 min, 2%–15% B; 6.00–9.00 min, 15%–30% B; 9.00–10.00 min, 30%–50% B; 10.00–11.00 min, 50%–85% B; and 11.00–12.00 min, 85%–100% B. The mass spectrometer was operated in the positive ion electrospray ionization mode (capillary voltage: 3.0 kV; sampling cone: 35 V; collision energy: 4 eV; source temperature: 115 °C; desolvation temperature: 300 °C; desolvation gas: 600 L/h). Two-stage mass spectrometry was also conducted in the positive ion electrospray ionization mode with a collision energy ramp of 12–30 eV instead.

3. Results

3.1. Sequence analysis of NRPS

Recombinant plasmid harboring *nrps* gene was sequenced and the gene sequence is consistent with the sequence AFJ42298.1 (<https://www.ncbi.nlm.nih.gov/nucore/386833542>). Protein sequence of NRPS was compared with other species and the phylogenetic tree was constructed (Fig. 1). The NRPS shared maximum homology (99%) with *B. subtilis* NRPS.

3.2. Expression of A1-PCP1 and Cy-A2-PCP2 modules

The N-terminally His6-tagged A1-PCP1 and Cy-A2-PCP2 domains were expressed in *E. coli* BL21 (DE3) and purified using affinity chromatography. The molecular weights of A1-PCP1 and Cy-A2-PCP2 were 98 and 116 kDa, respectively (Fig. 2), and the protein purity was determined using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2). Sfp expressed in *E. coli* BL21 was purified to convert PCP domains from inactive apo- to cofactor-containing holoforms (Fig. 2). The purified soluble recombinant protein samples were determined by the Bradford assay using Bovine serum albumin (BSA). About 1.5 mg of A1-PCP1 module protein, 1.8 mg of Cy-A2-PCP2 module protein, and 2.1 mg of Sfp protein were obtained from 1-L culture *E. coli* BL21, respectively.

3.3. Homologous comparison of domains involved in the nonribosomal synthesis of bacillamide

PCP is the site of 4'-PP cofactor binding and substrate acylation. The

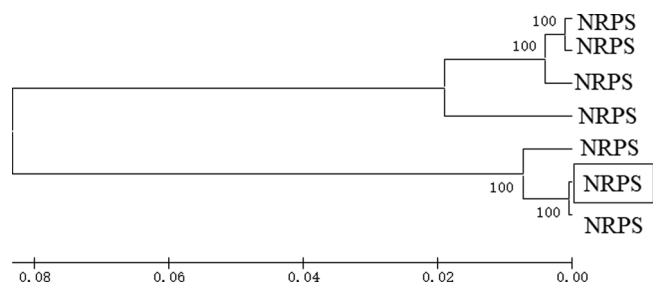


Fig. 1. Phylogenetic tree of NRPS from *B. atrophaeus* C89. Phylogenetic tree was drawn using UPGMA algorithm. *Bacillus mycooides* NRPS (WP_098632591.1), *Bacillus mycooides* NRPS (WP_088291524.1), *Bacillus thuringiensis* NRPS (WP_088033793.1), *Bacillus cereus* NRPS (WP_078175201.1), *Bacillus atrophaeus* NRPS (WP_094231268.1), *Bacillus atrophaeus* NRPS (WP_087942465.1), *Bacillus subtilis* NRPS WP_013390728.1.

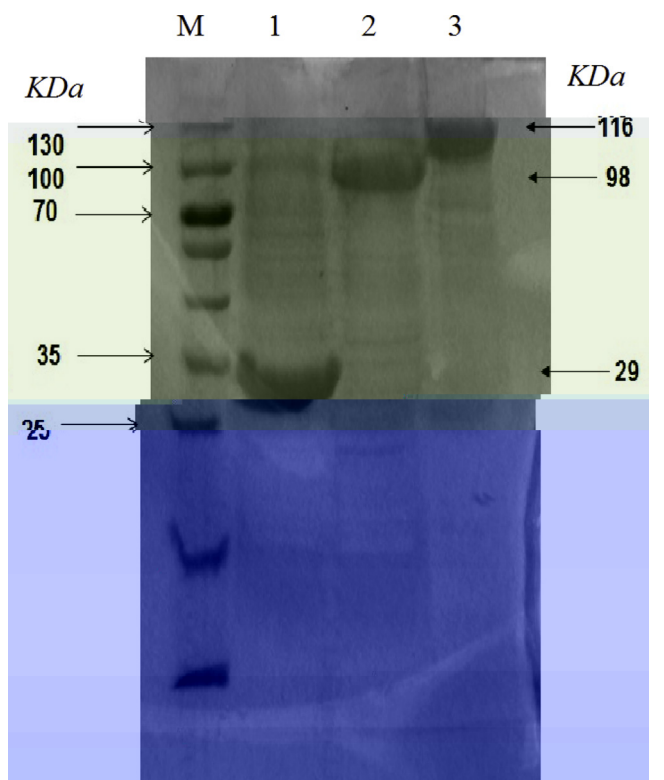


Fig. 2. SDS-PAGE of purified recombinant A1-PCP1 module, Cy-A2-PCP2 module, and Sfp. M: protein marker; Lane 1: purified Sfp (29 kDa); Lane 2: purified module A1-PCP1 (98 kDa); Lane 3: purified module Cy-A2-PCP2 (116 kDa).

functional unit of approximately 64 amino acid residues, to which aminoacyl substrates are bound as carboxy thioesters, was located in peptide synthetases directly downstream of the adenylation domains. PCP protein sequences revealed a signature sequence LGG (E/D) S around a highly conserved serine residue of the cofactor binding sequence (Table 1). The active site regions of the Cy and C domains were DXXXXD (XXS) and HHXXXX (G), respectively (Table 1).

3.4. Elucidation of bacillamide intermediate and functional prediction of the cyclization domain

Cy domain in NRPS was predicted to catalyze the heterocyclization of cysteine and alanine to thiazoline rings. To confirm this hypothesis, two modules of bacillamide synthetase, namely A1-PCP1 and Cy-A2-PCP2, were expressed and purified (Fig. 2). The NRPS requires post-translational modification of PCP domains with a 4'-phosphopantetheinyl (Ppant) prosthetic group, and this modification for covalent substrate tethering is conducted through a dedicated phosphopantetheinyl transferase. A biosynthetic experiment *in vitro* was performed, and the product AlaCys_{thiazole} was detected by mass spectrometry and two-stage mass spectrometry. One intermediate of bacillamide C—namely 4-carboxylic acid thiazole derivative (C₆H₈N₂O₂S, mass: 172.03), a cyclic dipeptide thiazole derivative was detected (Figs. 3 and 4). The sample from biosynthetic reaction mixture was confirmed by HR-ESI-MS, which showed an additional peak ([M + H] + 173.0379, identical to the predicted AlaCys_{thiazole}) compared with control groups (Fig S2). The biosynthetic process of NRPS domains involved in the synthesis of the intermediate product AlaCys_{thiazole} was predicted. We verified no product was formed in a control reaction in the absence of substrate amino acids or with the boiled protein (Fig. 4 and Fig. S2). The enzyme-bound amino acids are combined through a nucleophilic attack of either the cysteine amino group or the

Table 1

Amino acid alignment of the Peptidyl Carrier Proteins (PCP) domain core region LGGXS (A), the Cy domain core region DXXXXXXS (B), and C domain core HHXXXDG (C).

Domain	Consensus Sequence	Organism	Protein ID	Ref
A) peptidyl carrier proteins				
NrpsPCP1	LGGESL	<i>Bacillus atrophaeus</i>	EIM09914	In this study
NrpsPCP2	LGGDSI	<i>Bacillus atrophaeus</i>	EIM09914	In this study
TycAPCP	LGGDSI	<i>Bacillus brevis</i>	CAA31623.1	(Weckermann et al., 1988)
GrsBPCP	LGGHSL	<i>Bacillus brevis</i>	CAA43838.1	(Turgay et al., 1992)
SrfA-BPCP	IGGHSL	<i>Bacillus subtilis</i>	AIU78653.1	(Kesel et al., 2014)
B) Cylization domain				
NrpsCy	DALLMDGAS	<i>Bacillus atrophaeus</i>	EIM09914	In this study
VibFCy1	DMIACDAQS	<i>Vibrio cholerae</i>	AAF02102.1	(Marshall et al., 2001)
BacACy	DPLICDDSS	<i>Bacillus licheniformis</i>	AAC06346.1	(Duerfahrt et al., 2004)
HMWP2Cy1	DLIMDASS	<i>Yersinia pestis</i>	AAM85957.1	(Deng et al., 2002)
HMWP2Cy2	DNLLLDGLS	<i>Yersinia pestis</i>	AAM85957.1	(Duerfahrt et al., 2004)
HMWP1Cy3	DLLQFDVQS	<i>Yersinia pestis</i>	AAM85957.1	(Deng et al., 2002)
C) Condensation domain				
NrpsC	HHLLCDG	<i>Bacillus atrophaeus</i>	EIM09914	In this study
VibFC2	HHIVLDG	<i>Vibrio cholerae</i>	AAF02102.1	(Marshall et al., 2001)
BacAC1	HHIISDG	<i>Bacillus licheniformis</i>	AAC06346.1	(Duerfahrt et al., 2004)
EntFC	HHLLVDG	<i>Klebsiella pneumoniae</i>	CDO13417.1	(Bialek-Davenet et al., 2014)
GrsBC1	HHILMDG	<i>Bacillus brevis</i>	CAA43838.1	(Turgay et al., 1992)

heteroatom of the side chain onto the carbonyl C of alanine, and a proton is abstracted through base catalysis, thereby enabling the attack of either the cysteine side chain or the amino group onto the amide bond carbonyl. This results in the formation of a hydroxylated thiazolidine intermediate, which is subsequently dehydrated and dehydrogenated to provide the final thiazole-containing product AlaCys_{thiazole} (Fig. 5).

4. Discussion

Bacillamides are promising naturally occurring compounds for controlling harmful algal blooms (Churro et al., 2009; Shao et al., 2013). However, the low biosynthetic yield of these compounds has restricted not only their bioactivity studies but also their application in research. Therefore, to improve the potential for mass production of bacillamide C, investigating the biosynthetic mechanisms of bacillamide C is essential. Bloudoff et al. (2017) established a bacillamide synthesis assay by expression and purification of entire bacillamide synthetase with 265-kDa NRPS. Our prior study (Yuwen et al., 2013)

speculated that *B. atrophaeus* C89 can produce tryptamine as a building block of bacillamide C; and the study partially described AADC enzyme in the biosynthesis of bacillamide C. In our previous study, we have found that alanine and cysteine were the optimum substrates for the two adenylation domains (Zhang et al., 2018). The current heterologous expression study further elucidates the biosynthetic mechanism of bacillamide C. The *in vitro* biosynthesis of the bacillamide C intermediate- AlaCys_{thiazole} provides an opportunity for bacillamide C and analog production in the future.

4.1. Synthesis of the bacillamide C intermediate AlaCys_{thiazole}

Duerfahrt et al. (2004) suggested that Cy domain in an NRPS catalyzes the heterocyclization of cysteine or serine/threonine to thiazoline and oxazoline rings through a bimodular model system. However, the productivity of the engineered synthetases is hindered by poor substrate selectivity (Hura et al., 2012). Bloudoff et al. (2017) provided insight into the catalytic mechanisms of the Cy domain and implicated Aspartic acid- Threonine dyad in the catalysis of the cyclodehydration

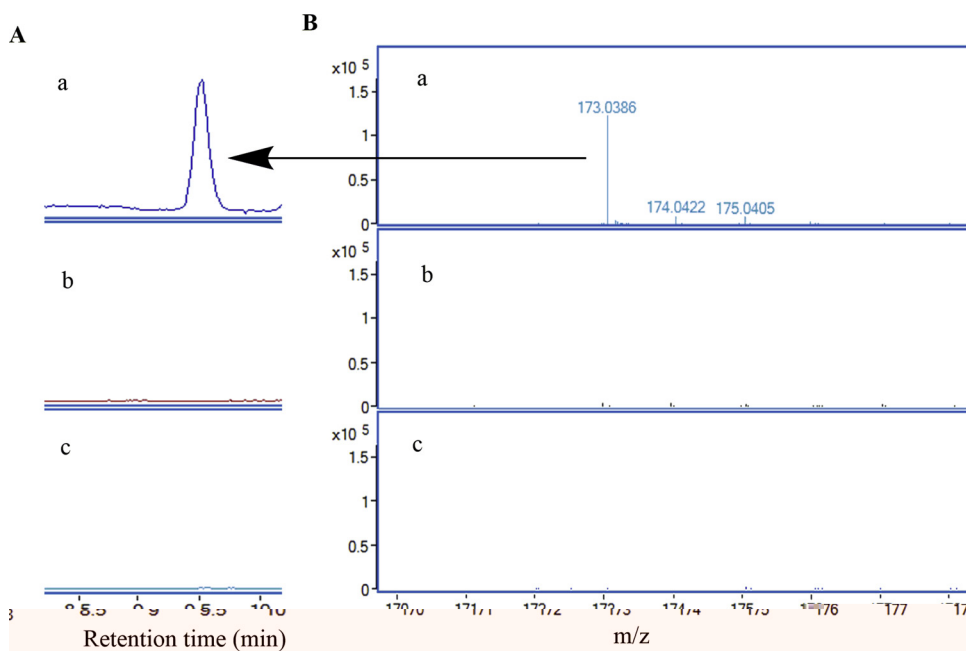


Fig. 3. High-performance liquid chromatography–mass spectrometry analysis of product formation catalyzed by *in vitro* combinatorial biosynthesis. Mass spectrum of the peak at 9.516 min, $[M+H]^+ = 173.0386\ m/z$ (a). Reaction in the absence of substrate amino acids was as a control reaction (b). Reaction mixtures with the boiled protein were as a control (c). Each experiment was performed in triplicate.

reaction. In this study, an *in vitro* assay for dipeptide formation was conducted. After mixing two recombinant protein modules, namely module A1-P1 and module Cy-A2-P2, the Sfp protein, and the corresponding substrate amino acids in an appropriate buffer, we conducted an *in vitro* assay for the synthesis of the bacillamide intermediate thiazole derivative (Figs. 3 and 4). We anticipated that AlaCys_{thiazoline} would be produced instead of AlaCys_{thiazole} because of the absence of the Ox domain. Notably, after the *in vitro* biosynthesis reaction, AlaCys_{thiazole} with a molecular weight of 172.03 was detected (Figs. 3, 4 and Fig. S2), and these results are inconsistent with the results of Bloudoff et al. (2017), who observe a thiazoline product. The Ox domain can catalyze the oxidation of the heterocyclic rings to produce corresponding thiazole and oxazole derivatives (Duerfahrt et al., 2004; Schneider and Walsh, 2004). The Ox domain from MtaD is involved in myxothiazol biosynthesis (Schneider et al., 2003). In the study of Li et al. (2013), the oxidation of 2-thiazolines and 2-oxazolines was reported to be crucial, and various methods have been reported for this transformation. The introduction of the second double bond into the AlaCys_{thiazole} moiety we observe in our experiments remains unexplained.

Heterocycles such as thiazolidines, thiazoles, and oxazoles exhibit a notable biological activity (Credico et al., 2011), and 4-carboxy oxazoles, thiazoles, and thiazolidines are condensed from serine or cysteine with aldehydes or acids. Fenner et al. (2016) chemically prepared arginine containing thiazole fragments as building blocks of plantazolicin A, which is a compound with promising selective activity against the causative agent of anthrax toxicity. IleCys_{thiazoline} is catalyzed by the heterogenetic Cy domain, which is the available Cy domain (Jaitzig et al., 2014). In our study, AlaCys_{thiazole} as a building block of bacillamide C was produced by *in vitro* reaction of heterologous expression NRPS. The rational and flexible design of NRPS templates is a critical technique for the synthesis of novel bacillamide analogs.

4.2. Heterologous expression of NRPS

Heterologous expression and mutating specificity-conferring residues are valuable in the investigation of the biosynthetic mechanism of bacillamide C. The heterologous expression of NRPS clusters is a

promising strategy for the mass production of bacillamide C. NRPSs are large multifunctional proteins with a modular organization. One module induces all catalytic activities that are necessary for the incorporation of one amino acid residue into the peptide product. In this study, we heterologously expressed A1-PCP1 and Cy-A2-PCP2 module proteins (Fig. 2). Heterologous expression generates novel synthetic bioactive compounds (Zhao and Kuipers, 2016). Bacillales may produce vital antimicrobial compounds that could be utilized for medical, food, or agricultural applications. A total of 1231 putative nonribosomal antimicrobial gene clusters were detected and subgrouped into 23 types of NRPSs, 5 types of PKs, and 3 types of NRPS/PKS hybrid synthesized compounds distributed over 49 species of Bacillales (Zhao and Kuipers, 2016). Only report from Saini et al. (2017) has heterologously expressed an enzyme from *B. atrophaeus* and characterized it in *E. coli*. The effective use of heterologous hosts for the characterization of biosyn-

Funding

The National Key Research and Development Program of China

- containing basic substance from a *Thermo actinomyces* species. *J. Antibiot.* 28, 609–610.
- Saini, M., Bindal, S., Gupta, R., 2017. Heterologous expression of gamma-glutamyl transpeptidase from *Bacillus atrophaeus* GS-16 and its application in the synthesis of gamma-d-glutamyl-L-tryptophan, a known immunomodulatory peptide. *Enzyme Microb. Technol.* 2017 (99), 67–76.
- Sasse, F., Steinmetz, H., Schupp, T., Petersen, F., Memmert, K., Hofmann, H., Heusser, C., Brinkmann, V., von Matt, P., Höfle, G., Reichenbach, H., 2002. Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. production, isolation, physico-chemical and biological properties. *J. Antibiot.* 55, 543–551.
- Schneider, T.L., Walsh, C.T., 2004. Portability of oxidase domains in nonribosomal peptide synthetase modules. *Biochemistry* 43, 15946–15955.
- Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., Stuibler, H.-P., 2003. The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proc. Natl. Acad. Sci.* 100, 8601–8606.
- Schwarzer, D., Marahiel, M.A., 2001. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* 88, 93–101.
- Selva, E., Gastaldo, L., Saddler, G., Toppo, G., Ferrari, P., Carniti, G., Goldstein, B., 1996. Antibiotics A21459 A and B, new inhibitors of bacterial protein synthesis. I. taxonomy, isolation and characterization. *J. Antibiot.* 49, 145–149.
- Shao, J., Li, R., Lepo, J.E., Gu, J.-D., 2013. Potential for control of harmful cyanobacterial blooms using biologically derived substances: problems and prospects. *J. Environ. Manage.* 125, 149–155.
- Socha, A.M., Long, R.A., Rowley, D.C., 2007. Bacillamides from a hypersaline microbial mat bacterium. *J. Nat. Prod.* 70, 1793–1795.
- Tabata, N., Tomoda, H., Zhang, H., Uchida, R., Omura, S., 1999. Zerkovamycin, a new cyclic peptide antibiotic from *Streptomyces* sp. K96-0670. II. Structure elucidation. *J. Antibiot.* 52, 34–39.
- Turgay, K., Krause, M., Marahiel, M.A., 1992. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* 6, 529–546.
- Wang, W., Joyner, S., Khoury, K.A.S., Dömling, A., 2010. (–)-Bacillamide C: the convergent approach. *Org. Biomol. Chem.* 8, 529–532.
- Watanabe, K., 2014. Effective use of heterologous hosts for characterization of biosynthetic enzymes allows production of natural products and promotes new natural product discovery. *Chem. Pharm. Bull.* 62, 1153–1165.
- Weckermann, R., Furbass, R., Marahiel, M.A., 1988. Complete nucleotide sequence of the *tycA* gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. *Nucleic Acids Res.* 16, 11841.
- Williams, B.B., Van Benschoten, A.H.V., Cimermancic, P., Donia, M.S., Zimmermann, M., Taketani, M., Ishihara, A., Kashyap, P.C., Fraser, J.S., Fischbach, M.A., 2014. Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe* 16, 495–503.
- Yu, L.L., Li, Z.Y., Peng, C.S., Li, Z.Y., Guo, Y.W., 2009. Neobacillamide A, a novel thiazole-containing alkaloid from the marine bacterium *Bacillus vallismortis* C89, associated with South China Sea sponge *Dysidea avara*. *Helv. Chim. Acta* 92, 607–612.
- Yu, L.Y., Hu, Z.F., Ma, Z.J., 2015. Production of bioactive tryptamine derivatives by co-culture of marine *Streptomyces* with *Bacillus mycoides*. *Nat. Prod. Res.* 29, 2087–2091.
- Yuwen, L., Zhang, F.-L., Chen, Q.-H., Lin, S.-J., Zhao, Y.-L., Li, Z.-Y., 2013. The role of aromatic L-amino acid decarboxylase in bacillamide C biosynthesis by *Bacillus atrophaeus* C89. *Sci. Rep.* 3, 1753.
- Zhang, F., Wang, Y., Jiang, Q., Chen, Q., Karthik, L., Zhao, Y.-L., Li, Z., 2018. Substrate selection of adenylation domains for nonribosomal peptide synthetase (NRPS) in bacillamide C biosynthesis by marine *Bacillus atrophaeus* C89. *J. Ind. Microbiol. Biotechnol.* 45, 335–344.
- Zhao, X., Kuipers, O.P., 2016. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species. *BMC Genomics* 17, 882.