Isolation and structure determination of new linear azole-containing peptides spongiicolazolicins A and B from Streptomyces sp. CWH03

Isolation and structure determination of new linear azole-containing peptides spongiicolazolicins A and B from *Streptomyces* sp. CWH03

Running headline: Structure determination of spongiicolazolicins

Authors: Mana Suzuki[1], Hisayuki Komaki[2], Issara Kaweewan[3], Hideo Dohra[1,4,5], Hikaru Hemmi[6], Hiroyuki Nakagawa[6,7], Hideki Yamamura[8], Masayuki Hayakawa[8] and Shinya Kodani[1,3,5,*]

Affiliations: [1]Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka, Japan; [2]Biological Resource Center, National Institute of Technology and Evaluation (NBRC), Chiba, Japan; [3]Graduate School of Science and Technology, Shizuoka University, Shizuoka, Japan; [4]Research Institute of Green Science and Technology; [5]Academic Institute, Shizuoka University, Shizuoka, Japan; [6]Food Research Institute, National Agriculture and Food Research Organization (NARO), Ibaraki, Japan; [7]Advanced Analysis Center, National Agriculture and Food Research Organization (NARO), Ibaraki, Japan; [8]Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Yamanashi, Japan

*To whom correspondence should be addressed: Shinya Kodani, College of Agriculture, Academic Institute, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka 422-8529 Japan, Tel/Fax; +81(54)238-5008, E-mail; kodani.shinya@shizuoka.ac.jp, ORCID:0000-0002-6792-1184

Abstract

Linear azole-containing peptides are a class of ribosomally synthesized and post-translationally modified peptides. We performed a chemical investigation on marine actinomycetes, and new linear azole-containing peptides named spongiicolazolicins A and B was found in the MeOH extracts of a newly isolated strain *Streptomyces* sp. CWH03 (NBRC 114659) and two strains of *S. spongiicola* (strain HNM0071^{T}: DSM 103383^{T} and strain 531S: NBRC 113560). The strain *Streptomyces* sp. CWH03 was indicated to be a new species closely related to *S. spongiicola* by phylogenetic analysis using the genome sequence. The new peptides named spongiicolazolicins A and B was isolated from the cell of *Streptomyces* sp. CWH03. The partial structure of spongiicolazolicin A was determined by 2D NMR experiments. Based on data of MS/MS experiments, the chemical structures of spongiicolazolicins A and B were proposed using the amino acid sequence deduced from the precursor-encoding gene, which was found from whole-genome sequence data of *Streptomyces* sp. CWH03. The biosynthetic gene cluster of spongiicolazolicins was propoaed based on comparative analysis with that of a known linear azole peptide goadsporin.

Keywords: linear azole-containing peptide, NMR, MS/MS, biosynthesis, *Streptomyces spongiicola*

Key Points

- *Streptomyces* sp. CWH03 was a new species isolated from marine sediment.
- New linear azole-containing peptides named spongiicolazolicins A and B were isolated.
- Biosynthetic pathway of spongiicolazolicins was proposed.

Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a class of naturally occurring peptides that are classified into several groups, such as lanthipeptides (McAuliffe et al. 2001; Willey and van der Donk 2007), lasso peptides (Hegemann et al. 2015; Maksimov et al. 2012), and thiazole/oxazole-modified microcins (TOMMs). TOMMs can be subdivided into several structural classes such as linear azol(in)e-containing peptides (LAPs) and thiopeptides (Cox et al. 2015). LAPs are defined as linear peptides containing several azole and/or azoline rings biosynthesized from Cys, Ser, and Thr residues in their precursor peptides (Melby et al. 2011). LAPs, such as microcin B17 (Liu 1994; Yorgey et al. 1994), streptolysin S (Molloy et al. 2011; Todd 1938; Wessels 2005), plantazolicin (Kalyon et al. 2011; Molohon et al. 2011; Scholz et al. 2011), coryneazolicin (Deane et al. 2016; Takuma et al. 2019), phazolicin (Travin et al. 2019), klebsazolicin (Metelev et al. 2017; Travin et al. 2018) and goadsporin (Igarashi et al. 2001; Onaka et al. 2005; Onaka et al. 2001; Ozaki et al. 2016; Ozaki et al. 2017), have been reported to have a wide variety of bioactivities such as antibacterial.

Goadsporin was isolated from *Streptomyces* sp. TP-A0584 as an inducing compound of morphological differentiation of streptomycetes (Igarashi et al. 2001; Onaka et al.

2001). The biosynthetic gene cluster (BGC) for goadsporin contains a structural gene, *godA*, and six modification enzyme-encoding genes (*godB*, *godD-H*). The precursor peptide GodA consists of an *N*-terminal 30-amino-acid leader peptide for recognition of the modification enzymes and a C-terminal 19-amino-acid core peptide that is modified to produce mature goadsporin (Onaka et al. 2005). Several Ser, Thr, and Cys residues in the core peptide region are first modified to azolines by the cyclodehydratase (GodD), followed by conversion to azoles by the flavin mononucleotide-dependent dehydrogenase GodE. Some Ser residues in the core peptide are converted to dehydroalanine (Dha) residues by cooperative modification of GodF and GodG, which are homologous to the *N*-terminal and *C*-terminal domains of the lanthipeptide dehydratase LanB, respectively. The *N*-terminal leader peptide is removed by the function of an *N*-terminal peptidase domain in a putative ABC transporter GodB. Finally, the *N*-terminal amino group of the core peptide is acetylated by the GNAT domain-containing acetyltransferase GodH to produce mature goadsporin.

Marine actinomycetes have been indicated as promising bioresources for secondary metabolites (Dhakal et al. 2017). *Streptomyces spongiicola* (strain HNM0071^{T}) was isolated as a novel actinomycete from an unidentified marine sponge collected on the coast of Sanya city, China (Huang et al. 2016). We previously isolated a new strain,

531S (deposited as NBRC 113560 in NITE Biological Resourse Center, Japan), of *S. spongiicola* from marine sediment in Okinawa and reported its whole-genome sequencing (Dohra et al. 2019). The closely related marine actinomycete *Streptomyces tirandamycinicus* was reported to produce antibiotics tirandamycins A and B (Huang et al. 2019). However, there is no report of secondary metabolites on *S. spongiicola*, so far. To find new secondary metabolites, we performed chemical screening on MeOH extracts of the marine actinomycetes including the two strains of *S. spongiicola* (DSM 103383^T = HNM0071^T and NBRC 113560 = 531S) and a newly isolated strain *Streptomyces* sp. CWH03 using HPLC and ESI-MS. As a result, we found new LAPs named spongiicolazolicins A and B (Fig. 1) in the extracts of *Streptomyces* sp. CWH03 and the two strains of *S. spongiicola*. Here, we describe the isolation and structure determination of spongiicolazolicins A and B from *Streptomyces* sp. CWH03. In addition, the biosynthetic pathway of spongiicolazolicins A and B was proposed based on the genome data of *Streptomyces* sp. CWH03 and the two strains of *S. spongiicola* (HNM0071^T and 531S).

Materials and methods

Bacterial strains

Bacterial strains including *Micrococcus luteus* NBRC 3333^{T}, *Pseudomonas aeruginosa* NBRC 12689^{T}, *Bacillus subtilis* subsp. *subtilis* NBRC 13719^{T}, *Staphylococcus aureus* subsp. *aureus* NBRC 100910^{T} and *Escherichia coli* NBRC 102203^{T} were obtained from the NBRC culture collection (NITE Biological Resource Center, Japan). *Streptomyces spongiicola* DSM 103383^{T} (HNM0071^{T}) was obtained from Leibniz Institute (DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Germany). *S. spongiicola* NBRC 113560 (531S) was previously reported (Dohra et al. 2019). *Streptomyces* sp. Strain CWH03 was deposited in the NBRC culture collection (NITE Biological Resource Center, Japan) and designated as NBRC 114659.

Isolation of *Streptomyces* sp. CWH03

A marine sediment sample was collected from Ishigaki Island, Okinawa, Japan, and was then suspended in sterilized distilled water. The sample suspension was spread on humic acid-vitamin (HV) agar containing nalidixic acid (20 mg/L) and cycloheximide (50 mg/L) and then incubated at 30 ºC for 2 weeks. Following the incubation period, the strain CWH03 was isolated and transferred to ISP No. 2 (ISP2: International

Streptomyces project medium No. 2) agar plate (Shirling and Gottlieb 1966). The strain was maintained in 20% (v/v) glycerol at -80 ºC until use.

Genome sequencing of *Streptomyces* sp. CWH03

Genomic DNA of *Streptomyces* sp. CWH03 was extracted using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) and fragmented using the Covaris M220 ultrasonicator (Covaris, Brighton, UK) according to the manufacturer's protocol for a 550-bp fragment. A genomic library was constructed using a TruSeq DNA PCR-free library preparation kit (Illumina) according to the manufacturer's instructions and sequenced on the Illumina MiSeq platform to generate 301-bp paired-end reads. The raw read sequences were cleaned up using Trimmomatic v0.36 (Bolger et al. 2014) by trimming adapter sequences and low-quality reads with parameters CROP:300, SLIDINGWINDOW:4:15, and MINLEN:150. The resultant 1,476,501 high-quality read pairs totaling 808.4 Mb, corresponding to 120-fold coverage of the genome, were assembled using the SPAdes ver. 3.13.0 (Bankevich et al. 2012) with a default *k*-mer setting and the following options: careful, only-assembler, and cov-cutoff auto. The draft genome sequence of *Streptomyces* sp. CWH03 was annotated using the DFAST-core ver. 1.2.5 (Tanizawa et al. 2018) with an in-house database created from 1,627

genome sequences of the genus *Streptomyces* deposited in the NCBI RefSeq database (January 31, 2020).

Average nucleotide identity (ANI) analysis (Konstantinidis and Tiedje 2005) was performed using a ruby script (ani.rb) from the enveomics collection (Rodriguez-R and Konstantinidis 2016) with default parameters. Briefly, nucleotide sequence fragments were extracted from the query genome with a window size of 1,000 bp and an overlapping step size of 200 bp. The fragments were then aligned to a reference genome using blastn, and an average of identity values of the best matches for the fragments was calculated. Multilocus sequence analysis (MLSA) was performed using the single-copy housekeeping genes, *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* of 13 strains except *Streptomyces* sp. I6, because *rpoB* and *trpB* of this strain were predicted to be pseudogenes. Nucleotide sequences of these five genes were individually aligned using MAFFT v7.427 (Katoh and Standley 2013) with an -auto option. To improve a subsequent phylogenetic analysis, poorly aligned regions were trimmed using trimAl v1.2 (Capella-Gutierrez et al. 2009) with an -automated1 option. The trimmed alignments were used to generate a maximum-likelihood phylogenetic tree using IQ-TREE v1.6.12 (Nguyen et al. 2015) with 1,000 replicates of the ultrafast bootstrap using UFBoot2 (Hoang et al. 2018) implemented in IQ-TREE. In both ANI analysis and

MLSA, two strains (S187^T and ATCC19609) of *Streptomyces xinghaiensis* (Zhao et al. 2009) isolated from marine sediment were used as outgroups.

Chemical investigation of marine actinomycetes

Marine actinomycetes were cultured using 100 mL of ISP2 agar medium for 7 days at 30 °C. The aerial hyphae and spore cells on the agar surface were harvested with a steel spatula. Double volume of MeOH was added to the harvested cells, followed by centrifugation at 3000 rpm for 10 min. The supernatant was subjected to HPLC purification using ODS column (4.6 × 250 mm, Wakopak Handy-ODS, Wako Pure Chemical Industries, Ltd., Osaka, Japan) with gradient elution from 20% to 80% MeCN (v/v) containing 0.05% trifluoroacetic acid (v/v) with flow rate of 1 ml/min for 20 min, setting UV detector at 220 nm to detect compounds. Each compound was isolated and subjected to ESI-MS analysis to determine molecular weight.

Isolation of spongiicolazolicins A and B

Streptomyces sp. CWH03 was cultured using 3L of ISP2 agar medium for 7 days at 30 °C. The aerial hyphae and spore cells on the agar surface were harvested with a steel spatula. Double volume of MeOH was added to the harvested cells, followed by centrifugation at 3000 rpm for 10 min. The supernatant was repeatedly subjected to HPLC purification using ODS column (4.6 × 250 mm, Wakopak Handy-ODS, Wako

Pure Chemical Industries, Ltd., Osaka, Japan) with gradient elution from 20 to 80% MeCN (v/v) containing 0.05% trifluoroacetic acid (v/v) for 20 min (flow rate, 1 ml/min) with UV detector set at 220 nm to yield spongiicolazolicins A and B, along with compound A (Fig. S2).

MS experiments

Peptides were dissolved in 50% MeCN (v/v). The accurate mass measurement was conducted using an ESI Orbitrap mass spectrometer (Orbitrap Velos ETD, Thermo Fisher Scientific, Waltham, MA, USA). The peptide was appropriately diluted with 50% MeOH (v/v) and was supplied to ESI Orbitrap mass spectrometer by direct infusion with electrospray ionization (ESI) in the positive polarity. Both MALDI-TOF MS and MS/MS analyses were conducted using a MALDI-TOF/TOF mass spectrometer (4800 Plus TOF/TOF analyzer, SCIEX, Redwood City, CA, USA). The peptide was mixed with equal volume of α-Cyano-4-hydroxycinnamic acid (4-CHCA) (Shimadzu GLC Ltd., Tokyo, Japan), matrix solution: 10 mg/mL in 50% MeCN (v/v) containing 0.1% TFA (v/v), and an aliquot of the mixture (0.5 μL) was spotted onto a standard stainless plate. After dried up, MS and MS/MS spectra were measured in the positive-ion mode. The mass spectrometer was tuned and calibrated using calibration standards of polytyrosine (Thermo Fisher Scientific Pierce Biotechnology, Rockford, IL, USA) and

the peptide mixture (Peptide Calibration Standard II, Bruker Daltonics), respectively, prior to the measurements.

NMR analysis

NMR sample was prepared by dissolving spongiicolazolicin A (2.0 mg) in 500 µl of DMSO-d_6. ^{1}H, ^{13}C, DEPT-135, and all 2D NMR spectra were obtained on Bruker Avance800 spectrometer with quadrature detection following the previous report (Kodani et al. 2017).

Antibacterial activity assay

Antibacterial activity assay using five testing bacterial strains (*E. coli*, *P. aeruginosa*, *B. subtilis*, *S. aureus*, and *M. luteus*) was performed by minimum inhibitory concentrations (MICs) test according to the previous report (Kaweewan et al. 2020). Serial twofold dilution of peptides spongiicolazolicins A and B (50 µL/well) were prepared in 96 well microplates. Testing bacteria were diluted with Muller Hinton broth to give a final concentration of 10^5 colony-forming unit (CFU)/mL. The bacterial solution (50 µL) was transferred to each well which contain a various concentration of compounds. After incubate the microplates at 30˚c for 24h, the MICs was determined as the lowest concentration that gave no visible growth.

Data availability

The raw read sequences of *Streptomyces* sp. CWH03 was deposited in DDBJ Sequence Read Archive (DRA) under the accession no. DRR207591. This whole genome shotgun sequencing project has been deposited in DDBJ/ENA/GenBank under the accession no. BLLG00000000.

Results

Genome sequencing of *Streptomyces* sp. CWH03

The strain *Streptomyces* sp. CWH03 was isolated from marine sediment on Ishigaki Island, Okinawa, Japan. To clarify the phylogenetic characteristics of the strain CWH03, the draft genome sequencing was accomplished. The draft genome sequence of *Streptomyces* sp. CWH03 consisted of 120 contigs totaling 6,714,233 bp with an N_{50} value of 264,498 bp and a G+C content of 72.32%. The genome contained 5,800 protein-coding genes and 64 transfer RNA genes. The ribosomal RNA operon copy number was unclear in the contigs generated by the short-read assembly. Sequencing coverages of the regions of the 23S (BLLG01000043, nt. 425–3,538) and 16S rRNA genes (BLLG01000003, nt. 546,993–548,507) were 769.8× and 637.8×, respectively, which were 6.4 and 5.3 times higher than the average sequencing coverage, respectively.

Phylogenetic analysis based on the 16S rRNA genes did not show enough resolution to define the phylogenetic position of *Streptomyces* sp. CWH03 because the 16S rRNA gene of CWH03 was 100% identical to those of *S. spongiicola* HNM0071^T and 531S and *Streptomyces* sp. I6. Thus, we performed average nucleotide identity (ANI) analysis and multilocus sequence analysis (MLSA) for species definition of *Streptomyces* sp. CWH03 among closely related species of the genus *Streptomyces*. As a result, the genome sequence of *Streptomyces* sp. CWH03 showed ANI values (88–93%) similar to those of *S. spongiicola* HNM0071^T (Zhou et al. 2019) and 531S (Dohra et al. 2019), *S. tirandamycinicus* HNM0039^T (Huang et al. 2019), *S. wuyuanensis* CGMCC 4.7042^T (Zhang et al. 2013), and seven strains of *Streptomyces* spp. (Fig. 2). These ANI values were lower than the species threshold of 95–96% (Goris et al. 2007; Richter and Rossello-Mora 2009), suggesting that *Streptomyces* sp. CWH03 is a novel species belonging to the genus *Streptomyces*. ANI analysis also showed that five unspecified species, strains CNT302, CNS615, CNR698, ICN441, and PKU-MA01144, belong to *S. tirandamycinicus*. These ANI analyses demonstrated that the group that is phylogenetically distant to other species of the genus *Streptomyces* is composed of *S. wuyuanensis*, *S. spongiicola*, *S. tirandamycinicus*, and three unspecified species (strains CWH03, I6, and ICN19) in the NCBI genome database, most of which are associated

with marine environments. MLSA based on the single-copy housekeeping genes *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* was also performed to infer the phylogenetic position of strain CWH03 in genetically related species of the genus *Streptomyces* (Fig. S1). The strain CWH03 clustered in the same clade as *S. spongiicola* HNM0071^T and 531S, which was supported by high bootstrap values. Taken together with the results of the ANI analysis, these findings suggested that strain CWH03 is a novel species closely related to *S. spongiicola.*

Isolation and structure determination of spongiicolazolicins A and B

We performed a chemical investigation of marine actinomycetes using HPLC and ESI-MS and detected new peptides including spongiicolazolicins A and B (Fig. 1) in the MeOH extract of *Streptomyces* sp. CWH03 and two strains of *S. spongiicola* (DSM 103383^T = HNM0071^T and NBRC 113560 = 531S). Among the three strains, *Streptomyces* sp. CWH03 stably produced the largest amount of spongiicolazolicins A and B, along with the compound A (Fig. S2). To obtain enough amount of the peptides for structure determination, *Streptomyces* sp. CWH03 was cultured using 3 L of modified ISP2 agar medium. In preliminary screening, spongiicolazolicins were observed to accumulate in the bacterial cells during agar cultivation (data not shown). Therefore, the cells were harvested from the surface of the agar medium for extraction.

The cells were extracted with MeOH, and the extract was repeatedly subjected to HPLC separation to obtain three compounds including spongiicolazolicins A and B (3.6 and 3.2 mg), and the compound A (2.5 mg, Fig. S2).

The molecular formula of spongiicolazolicin A was confirmed to be $C_{99}H_{117}N_{29}O_{30}S_2$ by accurate MS (Fig. S3) since the monoisotopic ion corresponding to $[M+2H]^{2+}$ was observed at *m/z* 1128.9062 (calculated *m/z* value, 1128.9054). To determine the chemical structure, NMR analyses including ^{1}H, ^{13}C, DEPT-135, DQF-COSY, TOCSY, NOESY, HMBC, and HSQC were performed on spongiicolazolicin A using 0.5 mL of DMSO-d_6 as a solvent (Fig S4-S33). The presence of amino acids (one unit each of Val, Ile, Phe and Pro; five units each of Gly and Ala; two units of Ser) was indicated by spin system analysis based on the HSQC, HMBC, DQF-COSY, and TOCSY spectra (Fig. 3). The presence of four units of dehydroalanine (Dha) was indicated based on the HMBC correlation from the amide proton to the β-carbon (Fig. 3). In the same manner, the presence of azole groups (three units of oxazole, two units of thiazole, and four units of methyl oxazole) was deduced from the NMR experiments (Table S1 and Fig. S5). Sequential analysis of the amino acids was performed based on HMBC and NOESY correlations (Fig. 3). As a result, ten partial structures (Gly1-Ser1, Val-Gly2, Ala1-Ala2, Gly3-Dha1, Ile-Dha2, Pro-Ser2-Gly4, Phe-Dha3, Ala3-Gly5, Ala4-Dha4, and Ala5)

were constructed (Fig. 3). However, the full sequence was not constructed by NMR data due to lack of long-range HMBC correlations.

Since the structure of spongiicolazolicin A contained multiple units of Dha and an azole ring, similar to goadsporin, we speculated that spongiicolazolicin A was biosynthesized by the biosynthetic system of RiPPs. By blastp similarity search using biosynthetic genes including *godF* and *godG* (Fig. 4A) for dehydroalanine biosynthesis in the goadsporin pathway, similar genes (*szlF* and *szlG* in Fig. 4B) were found in the genome of *Streptomyces* sp. CWH03. Regarding azole formation, genes encoding cyclodehydratase, YcaO protein (*szlD)* and YcaO partner protein (*szlC*) and dehydrogenases (*szlB* and *szlE*) were also found in the same gene cluster (Fig. 4B). Interestingly, four precursor-encoding genes *szlA1-A4* were found in the same BGC (Fig. 4B). Alignment of the four genes identified the conserved motif sequence **E**XX**E**X**EVLP** at the *N*-terminus (Fig. 5A), which was thought to be the leader sequence in the precursor peptide of RiPP biosynthesis. One of the precursor genes (*szlA4*) included the partial amino acid sequences (Gly-Ser, Val-Gly, Ala-Ala, Pro-Ser-Gly, and Ala-Gly) that were determined by NMR experiments. Therefore, we speculated that spongiicolazolicin A was biosynthesized from the precursor-encoding gene (*szlA4*) and the amino acid residues Ser, Thr, and Cys in the C-terminus region

(core peptide, Fig. 5B) formed Dha or azole rings. Comparison of the core peptide sequence and NMR-determined sequence was performed, as shown in Fig. 5B. The sequences matched very well, and the Ser residues at the 11^{th}, 14^{th}, 21^{st}, and 27^{th} positions were indicated to be converted to Dha by dehydration (blue in Fig. 5B). The molecular formula of the core peptide (GSTVGCAATGSSISTPSGCFSSAGTASSA) was $C_{99}H_{161}N_{29}O_{43}S_2$. Given that the formation of four dehydrated amino acids (-H_2O X 4) and nine azole ring formations (-H_4O X 9) had taken place in the core peptide, the expected molecular formula matched the determined molecular formula ($C_{99}H_{117}N_{29}O_{30}S_2$). Since four units of Dha were already detected by NMR experiments (* in Fig. 5B), the amino acids (red in Fig. 5B) in the core peptide must form azole rings during biosynthesis. Thus, we proposed the structure of spongiicolazolicin A to be that shown in Fig. 1. To confirm the structure, an MS/MS experiment was performed on spongiicolazolicin (Fig. S34-S37). As a result, fragmentation peaks were observed (Fig. 5C), which confirmed the chemical structure of spongiicolazolicin A to be shown in Fig. 1.

Since solubility of spongiicolazolicin B was low in water and organic solvents including MeOH, MeCN, Acetone, and DMSO at the high concentration (>0.1 mg/mL), it was difficult to determine the chemical structure by NMR. The MS/MS experiment

was accomplished to deduce the structure. The fragmentation ions were observed as shown in Fig. 5D. The fragmentation peaks (2229.7 and 2130.7) indicated the presence of valine, and the fragmentation peak 2229.7 indicated the two amino acids (GN) at *N*-terminus. Among amino acid sequences of precursor peptides SzlA1-A4, SzlA1 contains the sequence of GNV at *N*-terminus of core peptide (underlined in Fig. 5A). Thus, we proposed that spongiicolazolicin B was biosynthesized from the precursor Sz1A1. The accurate MS of spongiicolazolicin B was measured and the monoisotopic ion corresponding to $[M+2H]^{2+}$ was observed at *m/z* 1200.9883 (Fig. S38). Based on the hypothesis, the molecular formula of spongiicolazolicin B was proposed to be $C_{107}H_{137}N_{31}O_{30}S_2$ (The calculated *m/z* value,1200.9867 for $[M+2H]^{2+}$ of the proposed compound). The molecular formula of the core peptide of SzlA1(GNVSSAATVGSGSCPAMTWGTASTLSSK) was $C_{107}H_{175}N_{31}O_{41}S_2$. Given that the formation of three dehydrated amino acids (-H_2O X 3) and eight azole ring formations (-H_4O X 8) had taken place in the core peptide, the expected molecular formula matched the proposed molecular formula ($C_{107}H_{137}N_{31}O_{30}S_2$). Based on this molecule, we proposed the structure of spongiicolazolicin B as shown in Fig. 1, by interpretating the fragmentations obtained by MS/MS experiment (Fig. 5D and Fig. S39-S42).

We performed accurate MS and MS/MS experiments on the compound A in the extract of *Streptomyces* sp. CHW03 (Fig. S2). The monoisotopic ion was detected at $[M+2H]^{2+}$ 1221.9620 by accurate MS (Fig. S43). Judging from the accurate MS data, compound A may be biosynthesized from the precursors SzlA2 or SzlA3. However, the structure was not able to be determined by MS/MS fragmentations (Fig. S44-S47). As same as spongiicolazolicin B, insoluble characteristic of compound A hindered the measurement of NMR. The structure of compound A was unidentified in this study.

The HPLC profiles of three strains were compared (Fig. S2). *Streptomyces* sp. CWH03 produced spongiicolazolicins A and B, along with compound A. However, two strains of *S. spongiicola* (DSM 103383 and 531S) produced spongiicolazolicins A and B, with compound B (Fig. S2). Accurate MS of compound B was measured (Fig. S48), and the monoisotopic ion was detected at $[M+2H]^{2+}$ 1115.3886. Judging from the accurate MS data, compound B may also be biosynthesized from the precursors SzlA2 or SzlA3. The structure of compound B was not able to be proposed by MS/MS experiment (Fig. S49-S51). Three strains (CWH03, DSM 103383, and 531S in Fig. S2) commonly produced spongiicolazolicins A and B, assumably by the almost identical biosynthetic gene clusters. The reason of difference of HPLC profiles (compounds A and B in Fig. S2) was not clear.

Antibacterial activity test

One of LAPs, microcin B17, was reported to be antibacterial compound with potent inhibitory activity on DNA gyrase(Heddle et al. 2001; Zamble et al. 2001) Another LAP, plantazolicin, showed specific activity against *Bacillus anthracis*.(Molohon et al. 2016; Molohon et al. 2011). To clarify antibacterial characteristic of spongiicolazolicin A, we performed antibacterial activity test of spongiicolazolicin A against Gram-negative bacteria, including *E. coli* and *P. aeruginosa,* and Gram-positive bacteria, including *B. subtilis*, *S. aureus*, and *M. luteus*. Spongiicolazolicin A did not show antibacterial activity against all tested bacteria at a final concentration of 64 μg/mL.

Proposed biosynthesis of spongiicolazolicin

As stated above, we found a RiPP biosynthetic gene cluster in the draft genome sequence of *Streptomyces* sp. CWH03, which encoded genes for four precursors and other proteins for biosynthesis of azoles and dehydrated amino acids (Fig. 4B). The genes coding modification and processing are assigned as *szlB* - *szlG,* as color coded in Fig. 5B. The BGC of spongiicolazolicins contained encoding genes of split proteins SzlD (blue, YcaO-like protein in Fig. 4B) and SzlC (light blue, YcaO-partner protein in Fig. 4B), corresponding to the encoding gene of two domain-fused protein GodD in the

BGC of goadsporin (Fig. 4A). Since SzlC possesses YcaO-partner protein domain (light blue in Fig. 4D), it is indicated to be the partner protein for SzlD to form protein complex. YcaO cyclodehydratases form catalytic center in complex with the partner proteins (E1 ubiquitin-activating enzyme or ocin-ThiF-like protein family) which normally contain the domain called RiPP precursor peptide recognition element (RRE) (Burkhart et al. 2015; Burkhart et al. 2017). The RRE domain recognizes leader peptide region of precursor and recruits core peptide region to the enzyme reaction (Kloosterman et al. 2020). Interestingly, SzlC lacks of domain of RRE (Fig. 4D), which means that other protein possessing RRE domain is necessary for recognition of precursor. The protein SzlB has three domains including RRE (yellow, Fig. 4D), YcaO partner protein domain (light blue, Fig. 4D) and FMN-dependent dehydrogenase (purple, Fig. 4D). Interestingly this domain architecture is same as azole biosynthetic protein SulE in biosynthesis of thiopeptide sulfomycin (Du et al. 2020). The protein SzlE is indicated to possess dehydrogenase domain which has similarity with cyanobactin biosynthetic dehydrogenase ThcOx (purple, Fig. 4D). SzlH contains RRE domain (yellow, Fig. 4D) similar to that of PqqD family peptide modification chaperone and domain of site-2 protease (S2P) class of zinc metalloprotease (orange, Fig. 4D).

The gene clusters were well conserved among the three producers examined in this study (Fig. 4B). According to the functions of biosynthetic genes, we proposed the biosynthetic pathway of spongiicolazolicin A (Fig. 6). The precursor peptide SzlA4, including the *N*-terminal leader sequence, is ribosomally translated. The enzymes SzlB-SzlE catalyze the post-translational modification of SzlA4, including cyclization and dehydration. Nine Ser, Thr and Cys residues in the core peptide region composed of 29 amino acid residues (Fig. 5A) of SzlA4 are first modified to azolines by a YcaO partner protein (SzlC) and a YcaO cyclodehydratase (SzlD) and subsequently converted to form oxazole or thiazole rings by the dehydrogenases (SzlB and SzlE). Four Ser residues adjacent to the rings are converted to Dha residues by two lanthipeptide dehydratases, SzlF and SzlG. Finally, the *N*-terminal leader peptide is removed by a protease (SzlH) to yield mature spongiicolazolicin A (Fig. 6). Spongiicolazolicin B is proposed to be biosynthesized by the exact same biosynthetic system. Unlike the BGC of goadsporin, transporter-encoding genes are not found in/near the BGC of spongiicolazolicins.

Discussion

The essential proteins in LAP biosynthesis (Arnison et al. 2013; Cox et al. 2015; Melby et al. 2011) are the precursor peptide (protein A), modification enzymes including dehydrogenase (protein B), and YcaO-partner protein/YcaO like

cyclodehydratase (protein C/D). In many LAP biosynthetic gene clusters, the C-D proteins are combined in a single polypeptide, like GodD in goadsporin biosynthesis (Onaka et al. 2005; Ozaki et al. 2016; Ozaki et al. 2017). In the biosynthesis of spongiicolazolicins, YcaO-like cyclodehydratase (SzlD) and YcaO-partner protein (SzlC) are independent, as same as biosynthesis in microcin B17 (Collin and Maxwell 2019) and streptolysin S (Molloy et al. 2011). Interestingly, two genes (*szlB* and *szlE*) coding protein which has dehydrogenase domain exist in the cluster. In the biosynthesis of goadsporin and spongiicolazolicins, additional lanthipeptide dehydratases (green and light green in Fig 4) are present to give dehydrated amino acids. The similar lanthipeptide dehydratases are present in the BGC of thiopeptides (Cox et al. 2015). In lantibiotic nisin biosynthesis (Field et al. 2015; Lubelski et al. 2008), NisB is a lanthipeptide dehydratase classified in LanB protein (lanthionine biosynthetic enzyme B: dehydratase). In the NisB-catalyzed reaction, tRNA Glu-dependent glutamylation of serines and threonines is catalyzed by the *N*-terminus domain of NisB, and the *C*-terminal domain removes glutamate to afford dehydroamino acids (Bothwell et al. 2019). The BGC of goadsporin possessed split LanB proteins including GodF and GodG corresponding to *N*- and *C*-terminus regions of LanB, respectively (Ozaki et al. 2016). This split LanB proteins are present in BGCs of

thiopeptides such as nosiheptide (Wang et al. 2013). In biosynthesis of reported LAPs, the split LanB proteins were found only in the BGCs of goadsporin and spongiicolazolicins. The combination of azole biosynthesis (dehydrogenase and YcaO like cyclodehydratase/ YcaO-partner protein) and lanthipeptide biosynthesis (dehydratases) is likely a hybrid system to afford the peptides containing the azoles and dehydrated amino acids.

The distribution of spongiicolazolicin-type BGCs was searched by BLAST similarity search using the precursor peptide-encoding genes (*szlA1-A4*). BGCs similar to that of spongiicolazolicins (Table S2) were found in the genomes of 11 strains out of the phylogenetically related 12 strains (highlighted in purple, Fig. 2). The BGC of this class peptide was not found in the genome of *S. wuyuanensis* CGMCC4.7042^{T}. Phylogenetic analysis showed that *S. wuyuanensis* CGMCC4.7042^{T} formed a single monophyletic clade with *Streptomyces* sp. ICN19 containing the BGC (Fig. S1), suggesting that *S. wuyuanensis* CGMCC4.7042^{T} lost the BGC after diverging from *Streptomyces* sp. ICN19. The gene organizations and orientations of the BGCs are the same with that of the BGC of spongiicolazolicins, and the amino acid sequences of the precursors are highly similar (Table S2). The amino acid sequences of spongiicolazolicins A and B showed no similarity with that of goadsporin. In addition, the amino acid sequences of

the four-precursor peptide-encoding genes *szlA1-4* had low similarities each other (Fig. 5A).

In the course of searching for spongiicolazolicin type BGC, the similar BGCs were found in genome sequences of 11 strains of *Streptomyces* (Table S2; *Streptomyces qinglanensis* CGMCC 4.6825^{T}, *S. qinglanensis* SCSIO M10379, *Streptomyces* sp. CNT318, *Streptomyces* sp. AA1529, *Streptomyces* sp. LHW50302, *Streptomyces gardneri* NBRC 12865, *Streptomyces* sp. WAC05292, *Streptomyces viridifaciens* ATCC 11989, *Streptomyces* sp. 769, *S. venezuelae* ATCC 15439, and *Streptomyces* sp. CNZ279). For instance, the BGC of *S. qinglanensis* CGMCC 4.6825^{T} is shown in Fig. 4C. In the case of BGC of *S. qinglanensis* CGMCC 4.6825^{T}, all necessary genes encoding for four precursor genes and modification enzymes (SzlB - SzlG) were in the cluster, although order of genes were different from BGC of spongiicolazolicin. In the cluster, two genes coding for transporter are present as same as the BGC of goadsporin. The amino acid sequences of precursors of this related BGCs are different from those of BGC of spongiicolazolicins.

Above all, the BGCs of spongiicolazolicins and related compounds are distributed over many streptomycetes (Table S2). In this study, the azole-containing peptides spongiicolazolicins A and B were isolated from a new strain *Streptomyces sp.* CWH03.

Further exploration of the streptomycetes which have the related BGC will lead to find new azole containing peptides.

Author contribution

SK designed the research and wrote the manuscript. MS and IK conducted experiments of structure determination. HY and MH performed isolation of marine actinomycetes. HD conducted genome sequencing and genetic analysis on streptomycetes. HK accomplished bioinformatics analysis on biosynthetic pathway. HN and HH conducted MS and NMR experiments. All authors read and approved the manuscript.

Acknowledgment

We thank Ms. Tomoko Sato for measurement of MS data.

Funding information

This study was supported by Grants-in-aids of the Japan Society for the Promotion of Science (grant number 20K05848) and Research Grant from Koyanagi Foundation (grant number: 19060023).

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Goransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Muller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Sussmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30(1):108-160. doi:10.1039/c2np20085f

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol

19(5):455-477. doi:10.1089/cmb.2012.0021

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120. doi:10.1093/bioinformatics/btu170

Bothwell IR, Cogan DP, Kim T, Reinhardt CJ, van der Donk WA, Nair SK (2019) Characterization of glutamyl-tRNA-dependent dehydratases using nonreactive substrate mimics. Proc Natl Acad Sci U S A 116(35):17245-17250. doi:10.1073/pnas.1905240116

Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nat Chem Biol 11(8):564-570. doi:10.1038/nchembio.1856

Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, Mitchell DA (2017) YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function. Chem Rev 117(8):5389-5456. doi:10.1021/acs.chemrev.6b00623

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-1973. doi:10.1093/bioinformatics/btp348

Collin F, Maxwell A (2019) The microbial toxin microcin B17: prospects for the

development of new antibacterial agents. J Mol Biol 431(18):3400-3426. doi:10.1016/j.jmb.2019.05.050

Cox CL, Doroghazi JR, Mitchell DA (2015) The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. BMC Genomics 16:778. doi:10.1186/s12864-015-2008-0

Deane CD, Burkhart BJ, Blair PM, Tietz JI, Lin A, Mitchell DA (2016) *In vitro* biosynthesis and substrate tolerance of the plantazolicin family of natural products. ACS Chem Biol 11(8):2232-2243. doi:10.1021/acschembio.6b00369

Dhakal D, Pokhrel AR, Shrestha B, Sohng JK (2017) Marine rare actinobacteria: Isolation, characterization, and strategies for harnessing bioactive compounds. Front Microbiol 8:1106. doi:10.3389/fmicb.2017.01106

Dohra H, Kaweewan I, Casareto BE, Suzuki Y, Kodani S (2019) Draft genome sequence of *Streptomyces spongiicola* strain 531S, an actinobacterium isolated from marine sediment. Microbiol Resour Announc 8(3):e01198-18. doi:10.1128/MRA.01198-18

Du Y, Qiu Y, Meng X, Feng J, Tao J, Liu W (2020) A heterotrimeric dehydrogenase complex functions with 2 distinct YcaO proteins to install 5 azole heterocycles into 35-membered sulfomycin thiopeptides. J Am Chem Soc 142(18):8454-

8463. doi:10.1021/jacs.0c02329

Field D, Cotter PD, Ross RP, Hill C (2015) Bioengineering of the model lantibiotic nisin. Bioengineered 6(4):187-192. doi:10.1080/21655979.2015.1049781

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57(Pt 1):81-91. doi:10.1099/ijs.0.64483-0

Heddle JG, Blance SJ, Zamble DB, Hollfelder F, Miller DA, Wentzell LM, Walsh CT, Maxwell A (2001) The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. J Mol Biol 307(5):1223-1234. doi:10.1006/jmbi.2001.4562

Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2015) Lasso peptides: an intriguing class of bacterial natural products. Acc Chem Res 48(7):1909-1919. doi:10.1021/acs.accounts.5b00156

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35(2):518-522. doi:10.1093/molbev/msx281

Huang X, Kong F, Zhou S, Huang D, Zheng J, Zhu W (2019) *Streptomyces*

tirandamycinicus sp. nov., a novel marine sponge-derived actinobacterium with antibacterial potential against *Streptococcus agalactiae*. Front Microbiol 10:482. doi:10.3389/fmicb.2019.00482

Huang X, Zhou S, Huang D, Chen J, Zhu W (2016) *Streptomyces spongiicola* sp. nov., an actinomycete derived from marine sponge. Int J Syst Evol Microbiol 66(2):738-743. doi:10.1099/ijsem.0.000782

Igarashi Y, Kan Y, Fujii K, Fujita T, Harada K, Naoki H, Tabata H, Onaka H, Furumai T (2001) Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. II. Structure determination. J Antibiot (Tokyo) 54(12):1045-1053. doi:doi.org/10.7164/antibiotics.54.1045

Kalyon B, Helaly SE, Scholz R, Nachtigall J, Vater J, Borriss R, Sussmuth RD (2011) Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from *Bacillus amyloliquefaciens* FZB42. Org Lett 13(12):2996-2999. doi:10.1021/ol200809m

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772-780. doi:10.1093/molbev/mst010

Kaweewan I, Hemmi H, Komaki H, Kodani S (2020) Isolation and structure

determination of a new antibacterial peptide pentaminomycin C from *Streptomyces cacaoi* subsp. *cacaoi*. J Antibiot (Tokyo) 73(4):224-229. doi:10.1038/s41429-019-0272-y

Kloosterman AM, Shelton KE, van Wezel GP, Medema MH, Mitchell DA (2020) RRE-finder: a genome-mining tool for class-independent RiPP discovery. mSystems 5(5):e00267-20. doi:10.1128/mSystems.00267-20

Kodani S, Inoue Y, Suzuki M, Dohra H, Suzuki T, Hemmi H, Ohnishi-Kameyama M (2017) Sphaericin, a lasso peptide from the rare actinomycete *Planomonospora sphaerica*. Eur J Org Chem(8):1177-1183. doi:10.1002/ejoc.201601334

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102(7):2567-2572. doi:10.1073/pnas.0409727102

Liu J (1994) Microcin B17: posttranslational modifications and their biological implications. Proc Natl Acad Sci U S A 91(11):4618-4620.

Lubelski J, Rink R, Khusainov R, Moll GN, Kuipers OP (2008) Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. Cell Mol Life Sci 65(3):455-476. doi:10.1007/s00018-007-7171-2

Maksimov MO, Pan SJ, James Link A (2012) Lasso peptides: structure, function,

biosynthesis, and engineering. Nat Prod Rep 29(9):996-1006. doi:10.1039/c2np20070h

McAuliffe O, Ross RP, Hill C (2001) Lantibiotics: structure, biosynthesis and mode of action. FEMS Microbiol Rev 25(3):285-308. doi:10.1111/j.1574-6976.2001.tb00579.x

Melby JO, Nard NJ, Mitchell DA (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Curr Opin Chem Biol 15(3):369-378. doi:10.1016/j.cbpa.2011.02.027

Metelev M, Osterman IA, Ghilarov D, Khabibullina NF, Yakimov A, Shabalin K, Utkina I, Travin DY, Komarova ES, Serebryakova M, Artamonova T, Khodorkovskii M, Konevega AL, Sergiev PV, Severinov K, Polikanov YS (2017) Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel. Nat Chem Biol 13(10):1129-1136. doi:10.1038/nchembio.2462

Molloy EM, Cotter PD, Hill C, Mitchell DA, Ross RP (2011) Streptolysin S-like virulence factors: the continuing sagA. Nat Rev Microbiol 9(9):670-681. doi:10.1038/nrmicro2624

Molohon KJ, Blair PM, Park S, Doroghazi JR, Maxson T, Hershfield JR, Flatt KM, Schroeder NE, Ha T, Mitchell DA (2016) Plantazolicin is an ultra-narrow

spectrum antibiotic that targets the *Bacillus anthracis* membrane. ACS Infect Dis 2(3):207-220. doi:10.1021/acsinfecdis.5b00115

Molohon KJ, Melby JO, Lee J, Evans BS, Dunbar KL, Bumpus SB, Kelleher NL, Mitchell DA (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. ACS Chem Biol 6(12):1307-1313. doi:10.1021/cb200339d

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32(1):268-274. doi:10.1093/molbev/msu300

Onaka H, Nakaho M, Hayashi K, Igarashi Y, Furumai T (2005) Cloning and characterization of the goadsporin biosynthetic gene cluster from *Streptomyces* sp. TP-A0584. Microbiology 151(Pt 12):3923-2933. doi:10.1099/mic.0.28420-0

Onaka H, Tabata H, Igarashi Y, Sato Y, Furumai T (2001) Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. I. Purification and characterization. J Antibiot (Tokyo) 54(12):1036-1044. doi:doi.org/10.7164/antibiotics.54.1036

Ozaki T, Kurokawa Y, Hayashi S, Oku N, Asamizu S, Igarashi Y, Onaka H (2016) Insights into the biosynthesis of dehydroalanines in goadsporin. Chembiochem

17(3):218-223. doi:10.1002/cbic.201500541

Ozaki T, Yamashita K, Goto Y, Shimomura M, Hayashi S, Asamizu S, Sugai Y, Ikeda H, Suga H, Onaka H (2017) Dissection of goadsporin biosynthesis by *in vitro* reconstitution leading to designer analogues expressed *in vivo*. Nat Commun 8:14207. doi:10.1038/ncomms14207

Richter M, Rossello-Mora R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106(45):19126-19131. doi:10.1073/pnas.0906412106

Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1. doi:doi:10.7287/peerj.preprints.1900v1

Scholz R, Molohon KJ, Nachtigall J, Vater J, Markley AL, Sussmuth RD, Mitchell DA, Borriss R (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. J Bacteriol 193(1):215-224. doi:10.1128/JB.00784-10

Shirling BE, Gottlieb D (1966) Methods for characterizing *Streptomyces* species. Int J Syst Bacteriol 16 doi:10.1099/00207713-16-3-313

Takuma M, Kuroha M, Nagano Y, Kaweewan I, Hemmi H, Oyoshi T, Kodani S (2019)

Heterologous production of coryneazolicin in *Escherichia coli*. J Antibiot (Tokyo) 72(11):800-806. doi:10.1038/s41429-019-0212-x

Tanizawa Y, Fujisawa T, Nakamura Y (2018) DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34(6):1037-1039. doi:10.1093/bioinformatics/btx713

Todd EW (1938) The differentiation of two distinct serological varieties of streptolysin, streptolysin O and streptolysin S. J Pathol Bacteriol 47(3):423-445. doi:doi:10.1002/path.1700470307

Travin DY, Metelev M, Serebryakova M, Komarova ES, Osterman IA, Ghilarov D, Severinov K (2018) Biosynthesis of translation inhibitor klebsazolicin proceeds through heterocyclization and *N*-terminal amidine formation catalyzed by a single YcaO enzyme. J Am Chem Soc 140(16):5625-5633. doi:10.1021/jacs.8b02277

Travin DY, Watson ZL, Metelev M, Ward FR, Osterman IA, Khven IM, Khabibullina NF, Serebryakova M, Mergaert P, Polikanov YS, Cate JHD, Severinov K (2019) Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. Nat Commun 10(1):4563. doi:10.1038/s41467-019-12589-5

Wang S, Zhou S, Liu W (2013) Opportunities and challenges from current investigations into the biosynthetic logic of nosiheptide-represented thiopeptide antibiotics. Curr Opin Chem Biol 17(4):626-634. doi:10.1016/j.cbpa.2013.06.021

Wessels MR (2005) Streptolysin S. J Infect Dis 192(1):13-15. doi:10.1086/430625

Willey JM, van der Donk WA (2007) Lantibiotics: peptides of diverse structure and function. Annu Rev Microbiol 61:477-501. doi:10.1146/annurev.micro.61.080706.093501

Yorgey P, Lee J, Kordel J, Vivas E, Warner P, Jebaratnam D, Kolter R (1994) Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. Proc Natl Acad Sci U S A 91(10):4519-4523. doi:doi.org/10.1073/pnas.91.10.4519

Zamble DB, Miller DA, Heddle JG, Maxwell A, Walsh CT, Hollfelder F (2001) In vitro characterization of DNA gyrase inhibition by microcin B17 analogs with altered bisheterocyclic sites. Proc Natl Acad Sci U S A 98(14):7712-7717. doi:10.1073/pnas.141225698

Zhang X, Zhang J, Zheng J, Xin D, Xin Y, Pang H (2013) *Streptomyces wuyuanensis* sp. nov., an actinomycete from soil. Int J Syst Evol Microbiol 63(Pt 8):2945-2950.

doi:10.1099/ijs.0.047050-0

Zhao XQ, Li WJ, Jiao WC, Li Y, Yuan WJ, Zhang YQ, Klenk HP, Suh JW, Bai FW (2009) *Streptomyces xinghaiensis* sp. nov., isolated from marine sediment. Int J Syst Evol Microbiol 59(Pt 11):2870-2874. doi:10.1099/ijs.0.009878-0

Zhou S, Xiao K, Huang D, Wu W, Xu Y, Xia W, Huang X (2019) Complete genome sequence of *Streptomyces spongiicola* HNM0071^{T}, a marine sponge-associated actinomycete producing staurosporine and echinomycin. Marine Genomics 43:61-64. doi:10.1016/j.margen.2018.08.002

Figure legends

Fig. 1. Chemical structure of spongiicolazolicins A and B. Modified residues are colored; dehydroalanine (Dha): blue, azole rings: red. *_C_-terminus modified amino acid residues.

Fig. 2. Average nucleotide identity (ANI) matrix of *Streptomyces* sp. CWH03 and related species. ANI values are visualized by heat map and the relationship of the strains for ANI values are showed by dendrogram. Two strains of *S. xinghaiensis* are used as an outgroup. Blue bands represent the species threshold of ANI values (95–96%). Strains in purple have the BGCs of spongiicolazolicin class peptide.

Fig. 3. Partial amino acid sequences determined by 2D NMR correlations, * indicates proposed azole rings by the biosynthesis.

Fig. 4. Biosynthetic gene clusters for (A) Goadsporin and (B) Spongiicolazolicins and (C) Related biosynthetic gene cluster in *Streptomyces qinglanensis* CGMCC 4.6825^{T}. (D) Domain structure of proteins. The accession numbers for the encoded proteins are shown in Table S2

Fig. 5. (A) Amino acid sequences of the precursor peptide genes. Residues which are modified in mature peptide are indicated by color. Blue: dehydroalanine (Dha) forming residue. Red: azole ring forming residue. (B) Comparison between core peptide (*szlA4*) and determined partial amino acid sequence by NMR, * indicates dehydroalanine (Dha). Fragmentations obtained by MS/MS experiment on (C) spongiicolazolicin A and (D) spongiicolazolicin B. Modified residues are indicated by color. Blue: dehydroalanine (Dha). Red: azole ring.

Fig. 6. Putative biosynthetic pathway of spongiicolazolicin A.

Fig. 1

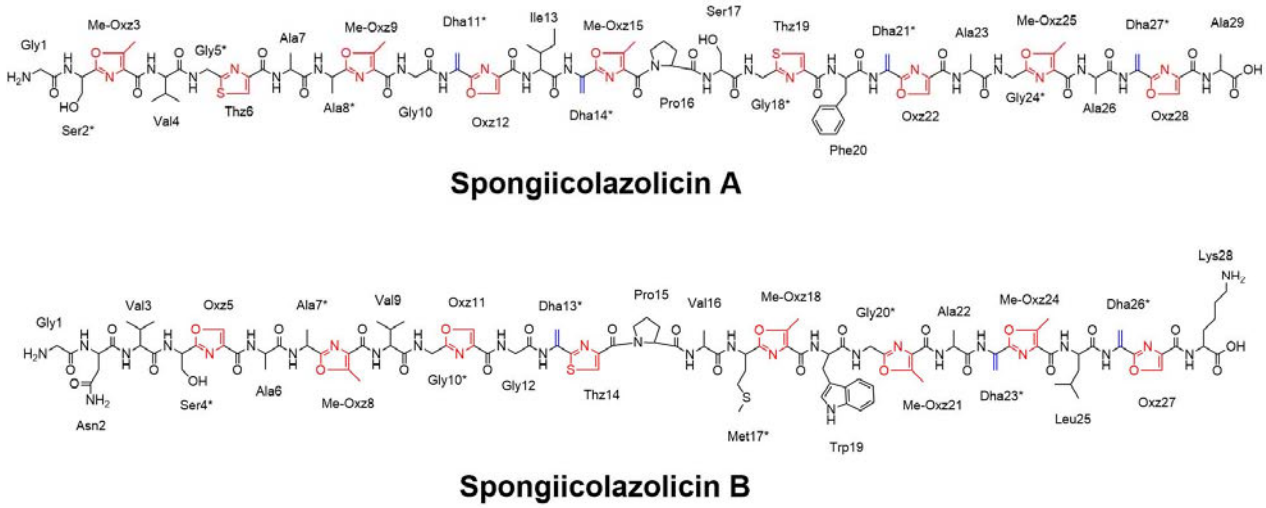


Fig. 1. Suzuki et al.

Fig. 2

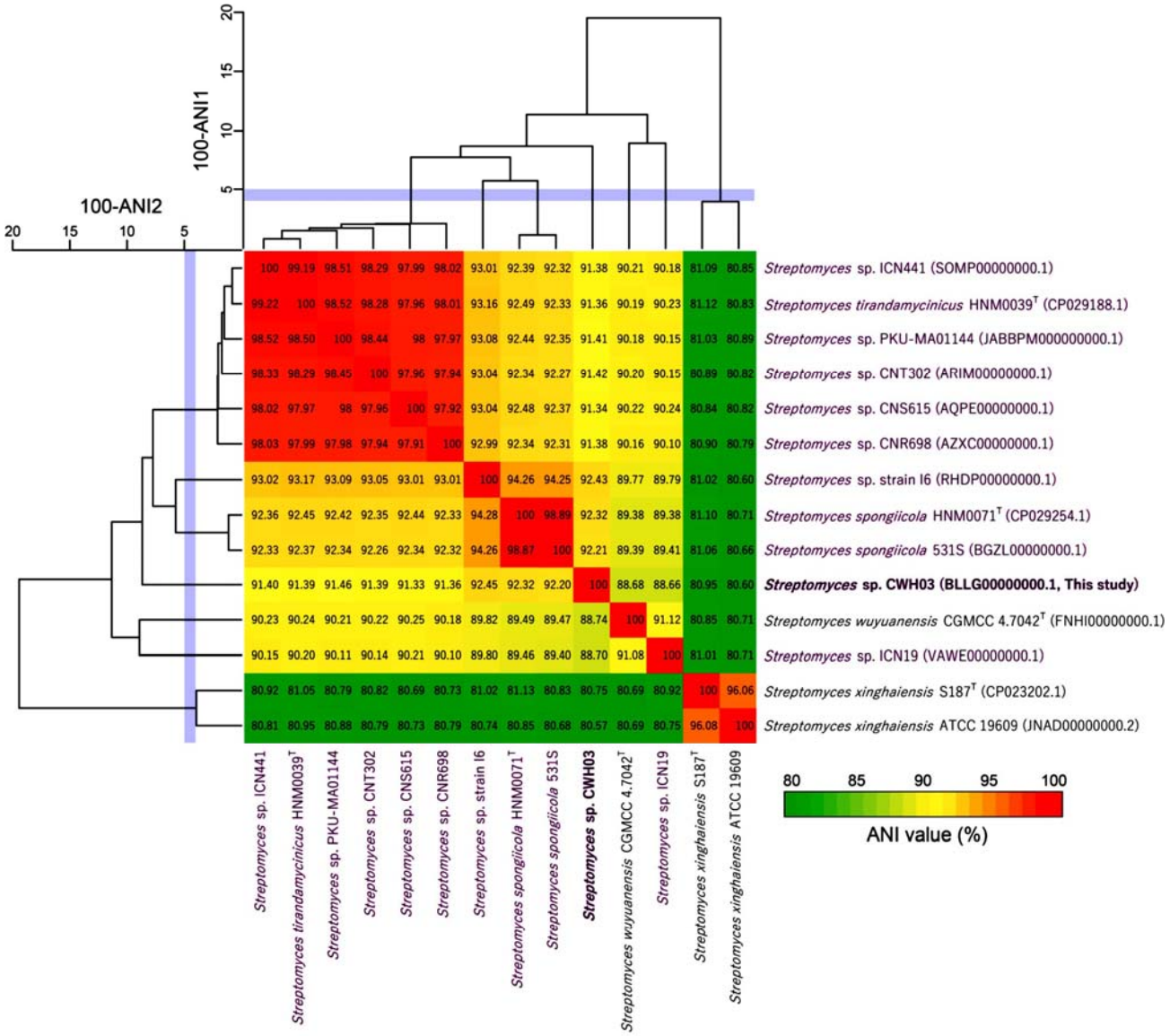


Fig. 2. Suzuki et al.

Fig. 3

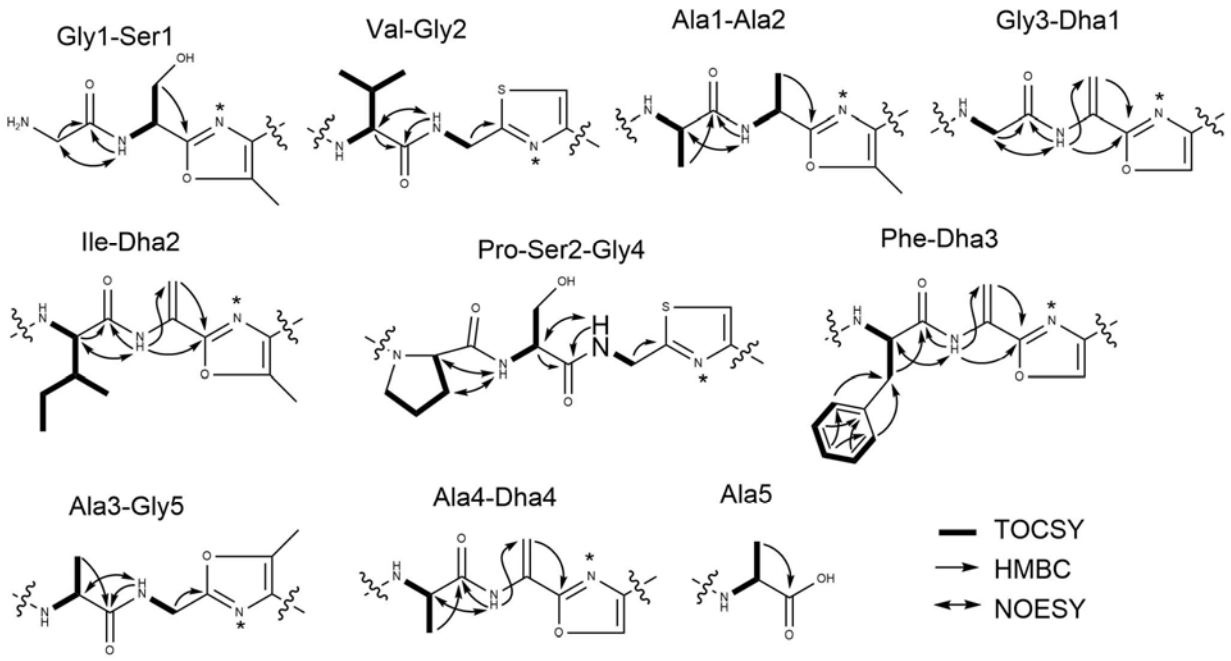


Fig. 3. Suzuki et al.

Fig. 4

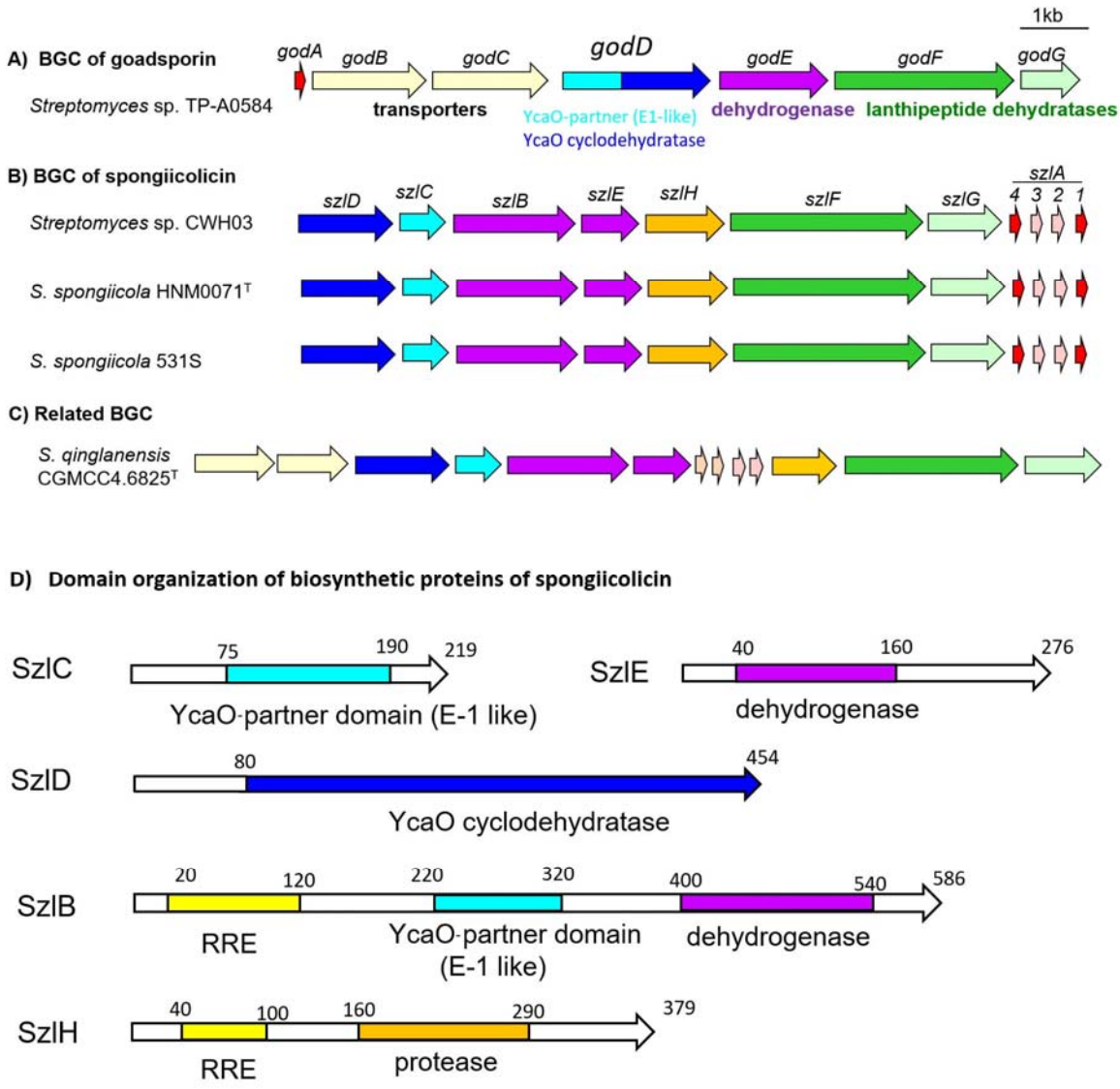


Fig. 4. Suzuki et al.

Fig. 5

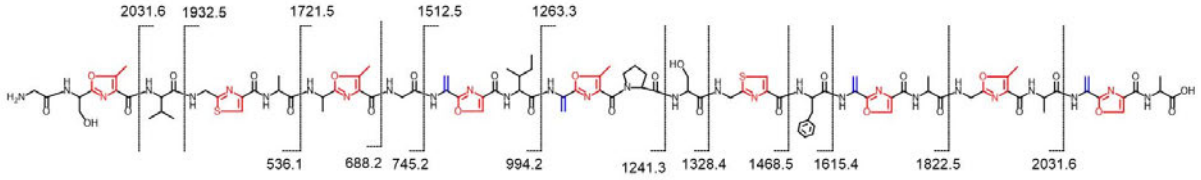
A

```
szlA1   MN--------ELDLRLDQEMPELEVLPEEYSAGN-VSSAATVGSGSCPAMTWGTASTLSSK
szlA2   MN--------ELDLRIEQESLEIEVLPDTQSAGNCAGTASTAGSFSCPAGSIGSAGSASST
szlA3   MDRLPETD--VLELVLEGERPELEVLPAEYAPGSSVGCAGSISCASCPAATISSGSTASSH
szlA4   MNDITPATGLALDLRLDQEAPELEVLPTSHSPGSTVGCAATGSSISTPSGCFSSAGTASSA
```

B

```
Core peptide of szlA4   GSTVGCAATGSSISTPSGCFSSAGTASSA
Determined by NMR       GS VG AA G* I* PSG F* AG A* A
```

C



D

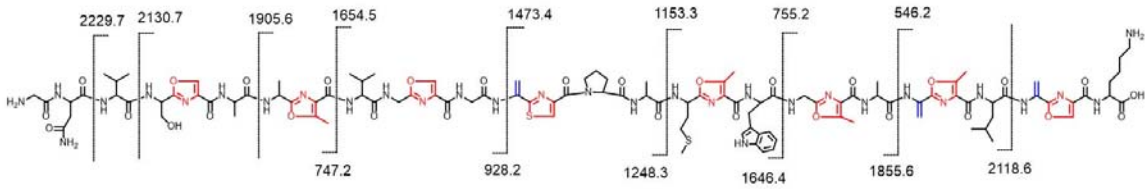


Fig. 5. Suzuki et al.

Fig. 6

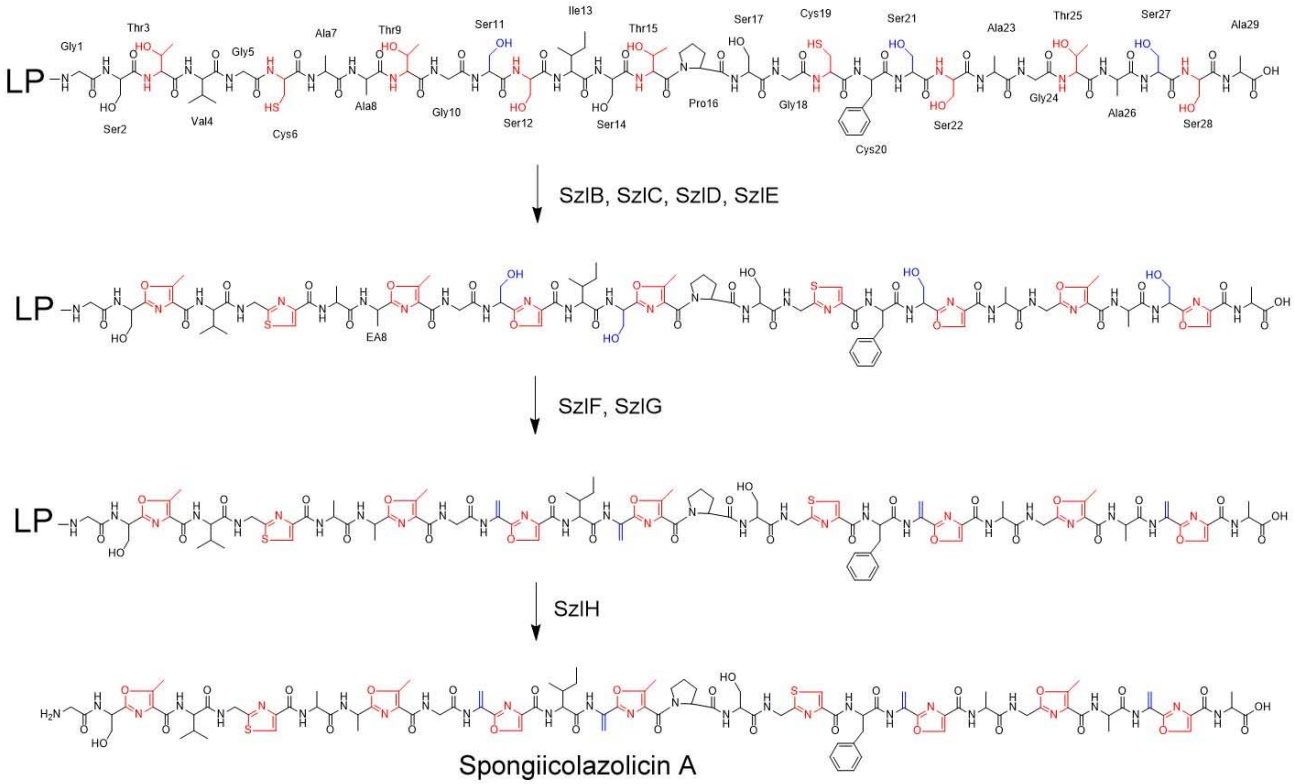


Fig. 6. Suzuki et al.