Heterologous production of new lasso peptide koreensin based on genome mining

Title: Heterologous production of new lasso peptide koreensin based on genome mining

Running headline: Heterologous production of koreensin

Authors: Hiroki Fuwa,[1] Hikaru Hemmi,[2] Issara Kaweewan,[3] Ikko Kozaki,[4] Hiroyuki Honda,[4] Shinya Kodani[1,3,5*]

Affiliations: [1]Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka, Japan, [2]Food Research Institute, National Agriculture and Food Research Organization (NARO), Ibaraki, Japan, [3]Graduate School of Science and Technology, Shizuoka University, Shizuoka, Japan, [4]Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, [5]Academic Institute, Shizuoka University, Shizuoka, 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

Keywords: lasso peptide, heterologous production, *Sphingomonas koreensis*

Abstract

Lasso peptides are a class of ribosomally biosynthesized and posttranslationally modified peptides with a knot structure as a common motif. Based on a genome search, a new biosynthetic gene cluster of lasso peptides was found in the genome of the proteobacterium *Sphingomonas koreensis*. Interestingly, the amino acid sequence of the precursor peptide gene includes two cell adhesion motif sequences (KGD and DGR). Heterologous production of the new lasso peptide was performed using the cryptic biosynthetic gene cluster of *S. koreensis*. As a result, a new lasso peptide named koreensin was produced by the gene expression system in the host strain *S. subtarrenea.* The structure of koreensin was determined by NMR and ESI-MS analysis. The three-dimensional structure of koreensin was obtained based on an NOE experiment and the coupling constants. A variant peptide (koreensin-RGD) which had RGD instead of KGD was produced by heterologous production with site-directed mutagenesis experiment. Koreensin and koreensin-RGD did not show cell adhesion inhibitory activity, although the molecules possessed cell adhesion motifs. The possible presence of a salt bridge between the motifs in koreensin was indicated, and it may prevent the cell adhesion motif from functioning.

Introduction

Lasso peptides are a class of ribosomally biosynthesized and posttranslationally modified peptides[1-4] (RiPPs) with a knot structure as a common motif.[5-8] The amino group of the *N*-terminal amino acid forms a peptide bond with the β- or γ-carboxyl group of Asp or Glu in the 7th-9th position from the *N*-terminus, resulting in the formation of a macrolactam. The macrolactam looks like a loop of a "lasso" with the C-terminal linear peptide, which normally passes through the ring, as the tail. The lasso peptide microcin J25 was isolated from *Escherichia coli*, and it is regarded as the archetype of lasso peptides.[9] Its biosynthetic gene cluster consists of four genes coding a precursor peptide (gene A: *mcjA*), two maturation enzymes (gene B: *mcjB* and gene C: *mcjC*) and an ATP-binding cassette transporter (gene D: *mcjD*) in one gene cluster (approximately 4.8 kbp).[10] Normally, lasso peptide biosynthetic genes in proteobacteria have the same set of genes, although the transporter gene is optional.

Through genome mining, Maksimov et al. investigated the genome data of more than 3000 prokaryotic species in search of BGCs of lasso peptides and identified 76 bacteria that had BGCs of lasso peptides.[11] A prediction system of RiPPs named RODEO (Rapid ORF Description and Evaluation Online) was developed, and six lasso peptides were found based on the RODEO system.[12] A genome-mining approach and heterologous

53 expression using *Escherichia coli* as the host cell was performed to produce new lasso
54 peptides.[8, 13-22] Attempts have also been made to utilize lasso peptides as fine chemical
55 materials. Two cysteines were introduced into the amino acid sequence of microcin J25,
56 followed by enzymatic cleavage.[23] Catenane-like molecules were generated by self-
57 association.[23] The lasso peptide benenodin-I was reported to form two thermally
58 interchangeable conformers.[24] The benenodin-I variant was prepared and digested with
59 trypsin to afford a rotaxane-like molecule.[24] It is difficult to synthesize lasso peptides in
60 vitro due to the instability of the modification enzymes. Recently, the in vitro synthesis
61 of lasso peptides were reported by two different groups (fuscanodin/fusilassin).[25, 26]
62 However, heterologous production is still the main method for exploring new lasso
63 peptides due to the instability of the modification enzymes. For the heterologous
64 production of lasso peptides, we established a gene expression system using the host
65 strain *Sphingomonas subteranea,* a natural producer of the lasso peptide subterisin.[27] As
66 a result, a new lasso peptide named brevunsin was produced by this system with a high
67 yield.[28] Previously, the genetic engineering of lasso peptides afforded a microcin J25
68 (RGDF) variant that had the cell adhesion motif RGD in the loop structure.[29] This
69 engineering turned lasso peptide microcin J25 into a highly potent and selective αvβ3
70 integrin inhibitor.[29] A lasso peptide precursor gene that possesses the cell adhesion

71 motifs KGD[30, 31] and DGR[32, 33] in the amino acid sequence was found in the genome of
72 the proteobacterium *Sphingomonas koreensis* through genome mining. The KGD motif
73 was reported to exist in the snake venom protein barbourin, which functions as a platelet
74 aggregation inhibitor by binding specifically to integrin IIb-IIIa receptors on the platelet
75 surface.[30, 31] To clarify characteristic of KGD motif, the mimicking cyclic peptide
76 containing RGD (cyclo-(S, S)-GCGRGDWPCA-NH_2) or KGD (cyclo-(S, S)-
77 GCGKGDWPCA-NH_2) was subjected to inhibition assays of fibrinogen binding to
78 integrin IIb-IIIa, and vitronectin binding to the receptor $\alpha_v\beta_3$.[34] As a result, both
79 peptides indicated more than 90% inhibition at the concentration of 10 μM in fibrinogen
80 binding to integrin IIb-IIIa. [34] Interestingly, inhibition of vitronectin binding to the
81 receptor was observed in the assay of the cyclic RGD peptide, but not in the assay of the
82 cyclic KGD peptide. This result indicated that the cyclic KGD peptide possessed higher
83 inhibition selectivity to integrin IIb-IIIa.[34] The DGR motif is known to be the cell
84 adhesion motif of the bone-cell secreted-signal peptide osteopontin.[32, 33] The DGR
85 containing peptide was reported to inhibit the attachment of cells to laminin and type I
86 collagen.[35] However, DGR containing peptides were indicated to be less than 1000
87 times as active as the RGD containing peptides in inhibiting the attachment of cells to
88 fibronectin.[36] These reports prompted us to accomplish the heterologous production of

89 a new lasso peptide using the cryptic gene cluster of *S. koreensis*. As a result, we
90 succeed in the heterologous production of a new lasso peptide named koreensin (**1** in
91 Fig. 1A). In addition, koreensin variant peptide containing RGD (koreensin-RGD, **2** in
92 Fig. 1A) was obtained by the site-directed mutagenesis experiment to investigate the
93 effect on cell adhesion.
94

Results and Discussion

During genome mining, lasso peptide precursor coding genes were found in the genome data of bacteria belonging to the *Sphingomonadaceae* family by a BLASTP search using the amino acid sequence of the lasso peptide benenodin-1[24] precursor peptide coding gene (*benA1*). The alignment revealed the consensus motif -Leu-Ile/Val-Asp-Leu-Gly- in the leader peptide sequence (Fig. 2). However, the core peptide amino acid sequences of the core peptides had no conserved motifs except for 1st Gly and 8th Asp/Glu. The precursor-encoding genes had a common motif, -Thr-X- as the last two amino acids in the leader peptide. The core sequences started with Gly and had Asp/Glu in the 8th position counting from the 1st Gly of the core peptide (Fig. 2). Normally, the Lasso peptide forms a macrolactam between the 1st Gly and the 7th-9th Asp or Glu. Therefore, the macrolactam was expected to be formed between the 1st Gly and 8th Asp/Glu in the peptides (Fig. 2). Among the precursors, a precursor peptide (korA, coding gene accession number: WP_107525062.1) had the natural cell adhesion motifs KGD and DGR in ring and tail sequences (gray background letters in Fig. 2). Since the amino acid sequence of korA contains two motifs (KGD and DGR), we expect that the mature lasso peptide of *S. koreensis* may have cell adhesion inhibitory activity. Therefore, we performed heterologous production using a gene cluster containing *korA*.

As shown in Fig. 3, the lasso peptide biosynthetic gene cluster consisted of 8 genes coding GntR family transcriptional regulator (WP_066573229.1), precursor peptide (korA: WP_107525062.1), modification enzymes (korB: WP_157926355.1 and korC: WP_066573225.1), isopeptidase (WP_157926354.1), TonB dependent receptor (WP_075152903.1), FecR-like protein (WP_066573215.1), and FecI-like protein (WP_082737884.1). The organization in the gene cluster is the same as that of the lasso peptide astexin.[21] The minimum gene set required to produce a mature lasso peptide is three genes (*korA*, *korB*, and *korC*). The gene *korB* includes a peptidase domain that may cleave the leader peptide, and *korC* codes cyclase to form an isopeptide bond. As shown in Fig. 3, the new lasso peptide koreensin was expected to be produced from precursor KorA by modification by enzymes KorB and KorC.

To perform the heterologous production of the expected peptide, the minimum set of biosynthetic genes (*korA*, *korB*, and *korC*) was cloned. Previously, the expression vector pHSG396Sp was established to produce a new lasso peptide in the proteobacterium *Sphingomonas subtarranea*.[28] The DNA fragment including the genes *korA*, *korB*, and *korC* was amplified by PCR and integrated into the expression shuttle vector pHSG396Sp to afford pHSG396Sp-16723L (Fig. S1). The plasmid was cloned using *E. coli* DH5α and transformed into the host bacterial strain *S. subtarrenea* by

electroporation following a previous report.[28] The transformant (*S. subterranea* possessing the plasmid pHSG396Sp-16723L) was cultured in liquid modified basal medium at 30 °C for 9 days. The culture was evaporated and extracted with MeOH. The MeOH extract was analyzed by HPLC (Fig. S2) and ESI-MS. As a result, the transformant was indicated to produce the expected peptide koreensin (**1** in Fig. 1A). To obtain enough **1** for structure determination and biological testing, the transformant was cultured using 1 L of modified basal medium. After open column chromatography, the 60% MeOH fraction was repeatedly subjected to HPLC purification to obtain 2.8 mg of **1** from 1 L of culture.

The molecular formula of **1** was established as $C_{73}H_{113}N_{21}O_{23}$ by accurate ESI-MS analysis (Fig. S3), since the ion corresponding to $[M+2H]^{2+}$ (the calculated m/z value, 826.9237) was observed at m/z 826.9239. To determine the planar chemical structure, NMR spectroscopy, including ^{1}H, ^{13}C, DEPT135, DQF-COSY, TOCSY, NOESY, HMBC, and HSQC experiments (Fig. S4-S11), was performed on **1** (2.0 mg) in 0.5 mL of DMSO-d_6. All 17 amino acids in the peptide were assigned using spin system identification from the 2D NMR spectral data (Table 1). Three linear peptide sequences (Gly1-Pro2-Lys3-Gly4, Asp5-Phe6-Pro7-Asp8-Val9-Gly10-Asp11-Gly12-Arg13-Ile14, and Leu15-Ala16-Gly17) were established by NOESY and HMBC data (Fig. 1B). The

149 *N*-terminus of Gly1 was indicated to be attached to the side chain of Asp8 to form an
150 isopeptide bond based on HMBC cross-peaks from the amide proton of Gly1 to the β-
151 carbonyl carbon of Asp8 and the NOESY correlation between the amide proton of Gly1
152 and the β-proton of Asp8. Above all, the assignment of all the constituent amino acids
153 of **1** was accomplished. Interestingly, a C-terminal peptide with four amino acids
154 (LTDE) was truncated during heterologous production, possibly by endogenous
155 proteases of *S. subterranea*.
156 The absolute stereochemistries of the constituent amino acids in **1** were analyzed by
157 modified Marfey's method (Fig. S12-S25).[37] All the constituent amino acids were
158 determined to be in the L-form, as shown in Fig. 1A. Three-dimensional analysis was
159 performed on **1** based on NOE correlations and the coupling constants obtained in the
160 NMR experiments. As shown in Fig. 4, three-dimensional (3D) structure of **1** was
161 obtained. The amino acid sequence of Gly1 to Asp8 formed a macrolactam (yellow in
162 Fig. 4B). The amino acids from Val9 to Asp11 formed a loop structure, and the amino
163 acids of the C-terminus sequence from Gly12 to Gly17 formed the tail. The 3D structure
164 of **1** was indicated to be a typical "lasso" structure (the atomic coordinate data were
165 deposited in the Protein Data Bank, PDB ID: 7BW5). The bulky amino acid Arg13
166 seemed to be a steric lock preventing the tail from becoming unthreaded from the

167 macrolactam. The cell adhesion motif KGD was in the macrolactam exposed to the
168 outside, and the other cell adhesion motif (DGR) was located inside the macrolactam.
169 The thermostability of **1** was tested at 50, 70, and 95 °C for 3 h. Interestingly,
170 compound **1** was stable during incubation even at 95 °C for 3 h (Fig. S26). The lasso
171 peptide astexin-2 gave unthreaded form due to thermal stress, while the related lasso
172 peptide astexin-3 was thermostable.[38] As same as astexin-3, compound **1** was resistant
173 to the thermal stress.
174 To obtain koreensin variant peptide (koreensin-RGD, **2** in Fig. 1A) which have RGD
175 sequence instead of KGD sequence, the site-directed mutagenesis experiment was
176 performed on the koreensin expression vector pHSG396Sp-16723L. As a result, the
177 mutated vector pHSG396Sp-16723L-RGD was obtained. In the same manner with
178 koreensin, isolation of koreensin variant (koreensin-RGD, **2** in Fig. 1A) which had Arg3
179 instead of Lys3 (Fig. 1A) was performed. As a result, **2** was obtained by HPLC
180 purification with the yield of 5.2 mg from 1 L of culture. The structure of **2** was
181 determined by ESI-MS and NMR (Fig. S27 and S28, 2D NMR data of **2** is not shown)
182 in the same manner with **1**.
183 To determine whether **1** and **2** had inhibitory activity on cell adhesion, a cell adhesion
184 assay using human umbilical vein endothelial cells (HUVECs) was performed in the

185 presence of **1** or **2**. As a result, compounds **1** and **2** did not show inhibitory activity
186 against cell adhesion at doses from 5 to 5000 nM (Fig. S29). This result was unexpected
187 since compounds **1** and **2** had cell adhesion motif of KGD or RGD (Fig. S29). Further
188 investigation of the 3D structure of **1** indicated the possibility of the formation of a salt
189 bridge between the amino residue of Lys3 and the carboxyl residue of Asp11. The
190 distance between the nitrogen in the amino residue of the side chain of Lys3 and the
191 oxygen in the carboxyl residue of Asp11 was calculated to be approximately 4.4
192 angstroms in the lowest energy structure (Fig. 4C). This seems a little bit long for a salt
193 bridge since a salt bridge is normally within 4 angstroms. The peptide could be
194 conformational flexible in solution, so we could not exclude the possibility of salt
195 bridge formation. On the other hand, the distance between the nitrogen in the guanidine
196 group of Arg13 and the oxygen in the carboxyl residue of Asp5 was calculated to be
197 more than 7 angstroms, which seemed to be too far to form a salt bridge. The formation
198 of a salt bridge between Lys3 and Asp11 may be one possible reason why compound **1**
199 did not show cell adhesion inhibition activity. Although the 3D structure of peptide
200 variant koreensin-RGD (**2**) was not obtained, koreensin-RGD may possess similar 3D
201 structure with **1**. The reason of the inactivity of **2** may be explained by the
202 intramolecular salt bridge between Arg3 and Asp11. The existence of a salt bridge

explains the structural stability of **1** against high thermal stress. Further mutation experiments on the expression plasmid are required to obtain recombinant koreensin, in which Asp11 is replaced with a neutral amino acid such as Ala to preclude salt bridge formation and achieve inhibitory activity on cell adhesion.

Materials and methods

Bacterial strains

The microorganisms (Bacterial strains including *Sphingomonas subterrenea* NBRC16086^{T} and *Sphingomonas koreensis* NBRC16723^{T}) were obtained from the NBRC culture collection (NITE Biological Resource Center, Japan).

Construction of expression vector for koreensin

For the template for PCR amplification, genome DNA was extracted from the cells of *S. koreensis* following previous report.[28] The DNA fragment including a lasso peptide koreensin biosynthetic gene cluster (about 2.7 kbp, korA, korB, and korC) was amplified by PCR with template and the primer pair of 16723L-F (TGTCTCTAGATGCATCATGCGCAGCGCTCA) and 16723L-R (ATTACGGTACCGCCAAGCACGCGCAATATACG), using EmeraldAmp PCR Master

220 Mix (Takara Bio Inc., Shiga, Japan) by following the manufacturer's instructions. The
221 insert DNA fragment including koreensin biosynthetic gene cluster and the shuttle
222 vector pHSG396Sp[28] were digested with XbaI (New England Biolabs Inc., Ipswich,
223 United States) and KpnI (New England Biolabs Inc.) according to the manufacturer's
224 instructions. The DNA products were ligated using T4 DNA ligation mix (Takara Bio
225 Inc.) to afford the shuttle vector pHSG396Sp-16723L (Fig. S2). *E. coli* DH5α cells
226 were transformed with 2 μL of the ligation mixture by chemical competence
227 transformation and cells were placed onto LB agar plates containing chloramphenicol
228 (final concentration of 20 μg/mL).

229 **Transformation of expression vector to *Sphingomonas subterranea***

230 The expression shuttle vector pHSG396Sp-16723L was transformed into *S.*
231 *subterranea* using electroporation. The bacterium *S. subterranea* was cultured in 10 mL
232 of NBRC medium number 802 liquid medium (peptone 10 g, yeast extract 2 g,
233 $MgSO_4 \cdot 7H_2O$ 1 g in distilled water 1 L, pH 7.0) at 30 °C for 24h with shaking at 50
234 rpm. The cells were cooled on ice for 30 min, followed by centrifugation (4000 rpm,
235 4 °C, 10 min). The harvested cells were suspended in 10 ml of cold 10% glycerol
236 solution. After centrifugation (4000 rpm, 4 °C, 10 min), The cells were harvested and
237 re-suspended in 10 ml of cold 10% glycerol solution. After centrifugation (4000 rpm,

4 °C, 10 min), the harvested cells were re-suspended in 0.1 mL of cold 10% glycerol solution for electroporation. The electroporation experiment to 0.1 mL of suspension of the cells was performed with the program "Ec3" (Voltage: 3.0 kV, number of pulse: 1) using MicroPulser (Bio-Rad Laboratories, California, USA). After electroporation, 0.5 mL of SOC medium was immediately added to the suspension of cells. For recovery of damage, and the cells were incubated at 30 °C for 2 h. Then, the cells of *S. subterranea* were spread onto NBRC medium number 802 agar medium (peptone 10 g, yeast extract 2 g, $MgSO_4 \cdot 7H_2O$ 1 g, agar 15 g in distilled water 1 L, pH 7.0) containing chloramphenicol (final concentration of 10 μg/mL). After incubation at 30 °C for 6 days, colonies were picked and checked by colony PCR method to obtain *S. subterranea* harboring pHSG396Sp-16723L.

Isolation of koreensin

Sphingomonas subterranea harboring pHSG396Sp-16723L was cultured using 1 L of modified basal medium[39] containing chloramphenicol (20 μg/mL, final concentration) with shaking of 120 rpm at 30 °C for 9 days. The modified basal medium was prepared by adding the inorganic compounds (K_2SO_4, 2 g; K_2HPO_4, 3 g; NaCl, 1 g; NH_4Cl, 5 g; $MgSO_4 \cdot 7H_2O$, 80 mg; $CuCl_2$, 0.5 mg; $MnSO_4 \cdot H_2O$, 0.35 mg; $FeCl_3$, 0.5 mg; $CaCl_2 \cdot 2H_2O$, 0.5 mg) in 1L of distilled water with adjusting pH 7.0. After autoclaving, the medium was supplemented with separately sterilized glucose and yeast extract at final concentrations of 0.25%, 0.005%, respectively. The culture including bacterial

cells was evaporated using rotary evaporator to aqueous residue (about 20 mL). The aqueous residue was extracted with MeOH (300 mL). After filtration with paper filter (Wattman No. 1 filter, GE Healthcare Life Sciences, Illinois, USA), the MeOH extract was concentrated to aqueous residue using rotary evaporator. The aqueous residue was subjected to open column chromatography using hydrophobic resin CHP-20P (Mitsubishi Chemical Co., Tokyo, Japan), eluted with 10% MeOH, 60% MeOH, and 100% MeOH. The 60% MeOH fraction was subjected to HPLC analysis using ODS column (4.6 × 250 mm, 5 μm, Wakopak Handy-ODS, Wako Pure Chemical Industries Ltd., Osaka, Japan) with gradient elution from 20% to 50% MeCN containing 0.05% trifluoroacetic acid for 20 min. The UV detector in HPLC system was set at 220 nm to detect **1** (retention time: 11.2min, Fig. S2), along with subterisin (retention time: 14.3min, Fig. S2). For isolation of koreensin, 60% MeOH fraction was repeatedly subjected to HPLC purification using ODS column (4.6 × 250 mm, 5 μm, Wakopak Handy-ODS, Wako Pure Chemical Industries, Ltd., Osaka, Japan) with isocratic elution of 22% MeCN containing 0.05% trifluoroacetic acid with UV detector set at 220 nm to isolate 2.8 mg of **1** (retention time, 15.8min).

ESI-MS experiment

Koreensin (**1**) was dissolved in 30% MeCN (0.1 mg/mL) and accurate ESI-MS of **1** was measured with JMS-T100LP (JEOL Co. ltd, Tokyo, Japan) using polyethylene glycol 1000 as internal standard.

NMR experiments and Structure calculations

279 A NMR sample was prepared by dissolving **1** (2.0 mg) or **2** (2.0 mg) in 500 μl of
280 DMSO-d_6. All NMR spectra were obtained on Bruker Avance 600 and Avance III HD
281 800 spectrometers (Bruker, Massachusetts, United States) with quadrature detection in
282 the phase-sensitive mode by States-TPPI (time proportional phase incrementation) and
283 in the echo-antiecho mode. NMR spectra were acquired according to previous report.[28]
284 Structure calculations of the three-dimensional structure of koreensin in solution state
285 were also performed by using distant restraints and backbone Φ dihedral angle
286 restraints obtained from the NMR data according to the previous report.[23]

287 **Modified Marfey's analysis**

288 The modified Marfey's analyses of **1** were carried out in sealed vacuum hydrolysis tube.
289 Compound **1** (0.5 mg) was hydrolyzed in 500 μL of 6 N HCl at 110 °C for 16 h. After
290 cooling to room temperature, the hydrolysate was completely evaporated using rotary
291 evaporator and freeze-dried under vacuum. The hydrolysate was resuspended in 200 μL
292 of water, followed by adding 10 μL of the solution of *N*α-(5-fluoro-2,4-dinitrophenyl)-
293 L-leucinamide (L-FDLA; Tokyo Chemical Industry Co., LTD, Tokyo, Japan) in acetone
294 (10 μg/μL). The 100 μL of 1M $NaHCO_3$ solution was added to the hydrolysate,
295 followed by incubation at 80 °C for 3 min. The reaction mixture was cooled down to
296 room temperature and neutralized with 50 μL of 2N HCl. The 50% MeCN/water (1 mL)

297 was added to the mixture and then subjected to HPLC analysis. Each standard amino
298 acid (1 mg) was derivatized with L-FDLA and D-FDLA in the same manner.
299 Approximately 20 μL of each FDLA derivative was analyzed by HPLC (C18 column,
300 4.6×250 mm, Wakopak Handy ODS, Wako Pure Chemical Industries, Tokyo, Japan).
301 DAD detector (MD-2018, JASCO, Tokyo, Japan) was used to detect amino acid
302 derivatives, accumulating data of absorbance from 220 nm to 420 nm. HPLC analysis
303 for standard amino acids, including Lys, Arg, Asp, Pro, Ala, Val, allo-Ile, Ile, Leu and
304 Phe, was performed using solvent A (distilled water containing 0.05% TFA) and solvent
305 B (MeCN containing 0.05% TFA) in linear gradient mode from 0 min to 70 min,
306 increasing percentage of solvent B from 25% to 60% at a flow rate of 1 mL/min (HPLC
307 condition 1). The retention times (min) for L- and D-FDLA derivatized amino acids
308 (HPLC condition 1) were as follows; L-Lys-L-FDLA (21.03 min), L-Lys-D-FDLA
309 (16.60 min), L-Arg-L-FDLA (21.36 min), L-Arg-D-FDLA (18.31 min), L-Asp-L-FDLA
310 (26.97 min), L-Asp-D-FDLA (28.74 min), L-Pro-L-FDLA (33.28 min), L-Pro-D-FDLA
311 (38.92 min), L-Ala-L-FDLA (33.77 min), L-Ala-D-FDLA (40.09 min), L-Val-L-FDLA
312 (40.72 min), L-Val-D-FDLA (54.24 min), L-allo-Ile-L-FDLA (46.02 min), L-allo-Ile-D-
313 FDLA (61.71 min), L-Ile-L-FDLA (46.32 min), L-Ile-D-FDLA (61.84 min), L-Leu-L-
314 FDLA (47.28 min), L-Leu-D-FDLA (62.65 min), L-Phe-L-FDLA (48.66 min) and L-Phe-

315 D-FDLA (59.40 min). HPLC analysis for Val was performed using solvent A (distilled
316 water containing 0.05% TFA) and solvent B (MeCN containing 0.05% TFA) in isocratic
317 mode 40% of solvent B for 55 min at a flow rate of 1 mL/min (HPLC condition 2). The
318 retention times (min) for L- and D-FDLA derivatized amino acids (HPLC condition 2)
319 were as follows; L-Val-L-FDLA (20.77 min) and L-Val-D-FDLA (39.14 min).

320 **Thermostability test of koreensin**

321 Concentration of **1** was adjusted to 0.5 mg/mL in DMSO. Aliquot sample (100 μL
322 each) was heated at 50 °C, 70 °C, and 95 °C for 3 h, followed by immediate cooling to
323 4 °C. Each sample (50 μL) was subjected to HPLC analysis using ODS column (4.6 ×
324 250 mm, 5 μm, Wakopak Handy-ODS, Wako Pure Chemical Industries Ltd., Osaka,
325 Japan) with gradient elution from 20 to 50% MeCN containing 0.05% trifluoroacetic
326 acid for 20 min with UV detector set at 220 nm.

327 **Site-directed mutagenesis experiment on expression vector pHSG396Sp-16723L to**
328 **obtain koreensin variant peptide (koreensin-RGD)**

329 The DNA fragment (about 2.7 kbp) was obtained by PCR method using primer set
330 (16723L-KtoR-F: ATTAAGGGCCCCAGGGGGGACTTCCCCGAT, 16723L-KtoR-R:
331 TTGAGTGAGCTGATACCGCTCGCCGCAGCC) and template DNA (pHSG396Sp-
332 16723L). The insert DNA fragment and the vector pHSG396Sp-16723L[28] were

digested with ApaI (New England Biolabs Inc.) and KpnI (New England Biolabs Inc.) according to the manufacturer's instructions. The DNA fragments were ligated using T4 DNA ligation mix (Takara Bio Inc.) to afford the vector pHSG396Sp-16723L-RGD. The mutation in the vector was confirmed by DNA sequencing. Transformation and expression experiments using pHSG396Sp-16723L-RGD were performed in the same manner with the vector pHSG396Sp-16723L.

Cell culture

Human Umbilical Vein Endothelial Cells (HUVEC) cells were provided as frozen cells after primary culture by the supplier (Kurabo, Osaka, Japan). HUVEC were maintained in 75 cm^2 flasks (658170; Greiner Bio-One, Frickenhausen, Germany) cultured in HuMedia-EG2 (KE-2150S; Kurabo) supplemented with attached additive.

Cell adhesion assay

A cell adhesion assay was conducted on SPOT arrays according to a previously described method with slight modifications.[40] The spots which synthesized RGDS peptide were punched out as a disk and embedded in a 96-well plate. Cells were stained with calcein-AM (Life Technologies Corporation, Carlsbad, CA, USA) for 30 min, and 1.0×10^4 cells/well were directly seeded on the disks with serum-free Dulbecco's modified Eagle medium (DMEM, 08458-16; Nacalai Tesque) containing RGDS or

Lasso peptide. Cells, soluble peptides and peptide disks were incubated for 5 h for cell adhesion on the peptide disks. After two washes using phosphate-buffered saline (PBS) to remove unattached cells, fluorescence intensity was measured on a Fluoroskan Ascent (type 374; Labsystems, Helsinki, Finland) at 485/505 nm (excitation/emission) wavelengths. For reproducibility, data from triplicate spots were averaged.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science by Grants-in-aids (grant number 20K05848), Koyanagi-Foundation, and Takahashi Industrial and Economic Research Foundation. The NMR spectra were recorded on Bruker Avance 600 and Avance III HD 800 spectrometers at Advanced Analysis Center, NARO.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Budisa, N. Expanded genetic code for the engineering of ribosomally synthetized and post-translationally modified peptide natural products (RiPPs). *Curr Opin Biotechnol.* **24**, 591-598 (2013)
2. Letzel, A. C., Pidot, S. J. & Hertweck, C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics.* **15**, 983 (2014)
3. Link, A. J. Biosynthesis: Leading the way to RiPPs. *Nat Chem Biol.* **11**, 551-552 (2015)
4. Sardar, D. & Schmidt, E. W. Combinatorial biosynthesis of RiPPs: docking with marine life. *Curr Opin Chem Biol.* **31**, 15-21 (2016)
5. Maksimov, M. O., Pan, S. J. & James Link, A. Lasso peptides: structure, function, biosynthesis, and engineering. *Nat Prod Rep.* **29**, 996-1006 (2012)
6. Cheung-Lee, W. L. & Link, A. J. Genome mining for lasso peptides: past, present, and future. *J Ind Microbiol Biotechnol.* **46**, 1371-1379 (2019)
7. Maksimov, M. O. & Link, A. J. Prospecting genomes for lasso peptides. *J Ind Microbiol Biotechnol.* **41**, 333-344 (2014)
8. Hegemann, J. D., Zimmermann, M., Xie, X. & Marahiel, M. A. Lasso peptides:

an intriguing class of bacterial natural products. *Acc Chem Res.* **48**, 1909-1919 (2015)

9. Salomon, R. A. & Farias, R. N. Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli*. *J Bacteriol.* **174**, 7428-7435 (1992)

10. Solbiati, J. O., Ciaccio, M., Farias, R. N. & Salomon, R. A. Genetic analysis of plasmid determinants for microcin J25 production and immunity. *J Bacteriol.* **178**, 3661-3663 (1996)

11. Maksimov, M. O., Pelczer, I. & Link, A. J. Precursor-centric genome-mining approach for lasso peptide discovery. *Proc Natl Acad Sci U S A.* **109**, 15223-15228 (2012)

12. Tietz, J. I., et al. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol.* (2017)

13. Jeanne Dit Fouque, K., Bisram, V., Hegemann, J. D., Zirah, S., Rebuffat, S. & Fernandez-Lima, F. Structural signatures of the class III lasso peptide BI-32169 and the branched-cyclic topoisomers using trapped ion mobility spectrometry-mass spectrometry and tandem mass spectrometry. *Anal Bioanal Chem.* **411**, 6287-6296 (2019)

14. Hegemann, J. D., et al. Xanthomonins I-III: a new class of lasso peptides with a

seven-residue macrolactam ring. *Angew Chem Int Ed Engl.* **53**, 2230-2234 (2014)

15. Hegemann, J. D., Zimmermann, M., Zhu, S., Klug, D. & Marahiel, M. A. Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. *Biopolymers.* **100**, 527-542 (2013)

16. Hegemann, J. D., Zimmermann, M., Xie, X. & Marahiel, M. A. Caulosegnins I-III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. *J Am Chem Soc.* **135**, 210-222 (2013)

17. Hegemann, J. D., Schwalen, C. J., Mitchell, D. A. & van der Donk, W. A. Elucidation of the roles of conserved residues in the biosynthesis of the lasso peptide paeninodin. *Chem Commun (Camb).* **54**, 9007-9010 (2018)

18. Cheung-Lee, W. L., et al. Discovery of ubonodin, an antimicrobial lasso peptide active against members of the *Burkholderia cepacia* complex. *Chembiochem.* **21**, 1335-1340 (2019)

19. Cheung-Lee, W. L., Parry, M. E., Jaramillo Cartagena, A., Darst, S. A. & Link, A. J. Discovery and structure of the antimicrobial lasso peptide citrocin. *J Biol Chem.* **294**, 6822-6830 (2019)

20. Cheung-Lee, W. L., Cao, L. & Link, A. J. Pandonodin: a proteobacterial lasso peptide with an exceptionally long C-terminal tail. *ACS Chem Biol.* **14**, 2783-2792

(2019)

21. Zimmermann, M., Hegemann, J. D., Xie, X. & Marahiel, M. A. The astexin-1 lasso peptides: biosynthesis, stability, and structural studies. *Chem Biol.* **20**, 558-569 (2013)

22. Martin-Gomez, H., Linne, U., Albericio, F., Tulla-Puche, J. & Hegemann, J. D. Investigation of the biosynthesis of the lasso peptide chaxapeptin using an *E. coli*-based production system. *J Nat Prod.* **81**, 2050-2056 (2018)

23. Allen, C. D. & Link, A. J. Self-assembly of catenanes from lasso peptides. *J Am Chem Soc.* **138**, 14214-14217 (2016)

24. Zong, C., Wu, M. J., Qin, J. Z. & Link, A. J. Lasso peptide benenodin-1 is a thermally actuated [1] rotaxane switch. *J Am Chem Soc.* **139**, 10403-10409 (2017)

25. Koos, J. D. & Link, A. J. Heterologous and *in vitro* reconstitution of fuscanodin, a lasso peptide from *Thermobifida fusca*. *J Am Chem Soc.* **141**, 928-935 (2019)

26. DiCaprio, A. J., Firouzbakht, A., Hudson, G. A. & Mitchell, D. A. Enzymatic reconstitution and biosynthetic investigation of the lasso peptide fusilassin. *J Am Chem Soc.* **141**, 290-297 (2019)

27. Kuroha, M., Hemmi, H., Ohnishi-Kameyama, M. & Kodani, S. Isolation and structure determination of a new lasso peptide subterisin from Sphingomonas

438 subterranea. *Tetrahedron Letters.* **58**, 3429-3432 (2017)

439 28. Kodani, S., Hemmi, H., Miyake, Y., Kaweewan, I. & Nakagawa, H. Heterologous
440 production of a new lasso peptide brevunsin in *Sphingomonas subterranea*. *J Ind*
441 *Microbiol Biotechnol.* **45**, 983-992 (2018)

442 29. Hegemann, J. D., et al. Rational improvement of the affinity and selectivity of
443 integrin binding of grafted lasso peptides. *J Med Chem.* **57**, 5829-5834 (2014)

444 30. Minoux, H., Chipot, C., Brown, D. & Maigret, B. Structural analysis of the KGD
445 sequence loop of barbourin, an αIIbβ3-specific disintegrin. *J Comput Aided Mol*
446 *Des.* **14**, 317-327 (2000)

447 31. Scarborough, R. M., et al. Barbourin. A GPIIb-IIIa-specific integrin antagonist
448 from the venom of *Sistrurus m. barbouri*. *J Biol Chem.* **266**, 9359-9362 (1991)

449 32. Horii, A., Wang, X., Gelain, F. & Zhang, S. Biological designer self-assembling
450 peptide nanofiber scaffolds significantly enhance osteoblast proliferation,
451 differentiation and 3-D migration. *PLoS One.* **2**, e190 (2007)

452 33. Tsutsumi, H., Kawamura, M. & Mihara, H. Osteoblastic differentiation on
453 hydrogels fabricated from Ca^{2+}-responsive self-assembling peptides
454 functionalized with bioactive peptides. *Bioorg Med Chem.* **26**, 3126-3132 (2018)

455 34. Scarborough, R. M., et al. Design of potent and specific integrin antagonists.

Peptide antagonists with high specificity for glycoprotein IIb-IIIa. *J Biol Chem.* **268**, 1066-1073 (1993)

35. Yamada, K. M. & Kennedy, D. W. Peptide inhibitors of fibronectin, laminin, and other adhesion molecules: unique and shared features. *J Cell Physiol.* **130**, 21-28 (1987)

36. Hautanen, A., Gailit, J., Mann, D. M. & Ruoslahti, E. Effects of modifications of the RGD sequence and its context on recognition by the fibronectin receptor. *J Biol Chem.* **264**, 1437-1442 (1989)

37. Harada, K. I., Fujii, K., Hayashi, K., Suzuki, M., Ikai, Y. & Oka, H. Application of D,L-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method. *Tetrahedron Lett.* **37**, 3001-3004 (1996)

38. Allen, C. D., Chen, M. Y., Trick, A. Y., Le, D. T., Ferguson, A. L. & Link, A. J. Thermal unthreading of the lasso peptides astexin-2 and astexin-3. *ACS Chem Biol.* **11**, 3043-3051 (2016)

39. Imbert, M. & Blondeau, R. Effect of light on germinating spores of *Streptomyces viridosporus*. *FEMS Microbiol Lett.* **181**, 159-163 (1999)

40. Kanie, K., et al. Collagen type IV-specific tripeptides for selective adhesion of

474 endothelial and smooth muscle cells. *Biotechnol Bioeng*. **109**, 1808-1816 (2012)

475

476

Figure legends

Figure 1. A) Chemical structure of **1** and **2**, B) Key correlations of 2D NMR of **1**

Figure 2. Alignment of amino acid sequences of lasso peptide precursor coding genes similar to *benA1*. Bold letters: conserved amino acids. Underlined letters: core peptide. Gray back ground letters: cell adhesion motif.

Figure 3. Biosynthetic gene cluster for production of **1**

Figure 4. NMR-derived structures of **1**: (A) superposition of the 15 lowest-energy structures and (B) the lowest-energy structure of **1**. The isopeptide bond between Gly1 and Asp8 is shown in red. The ring-forming residues are shown in yellow, and the loop and the tail in blue. The side chain of "steric lock" Arg13 is shown. (C) The distance between nitrogen (orange) in amino residue of Lys3 and oxygen (cyan) in carboxyl residue of Asp11

Figure 1

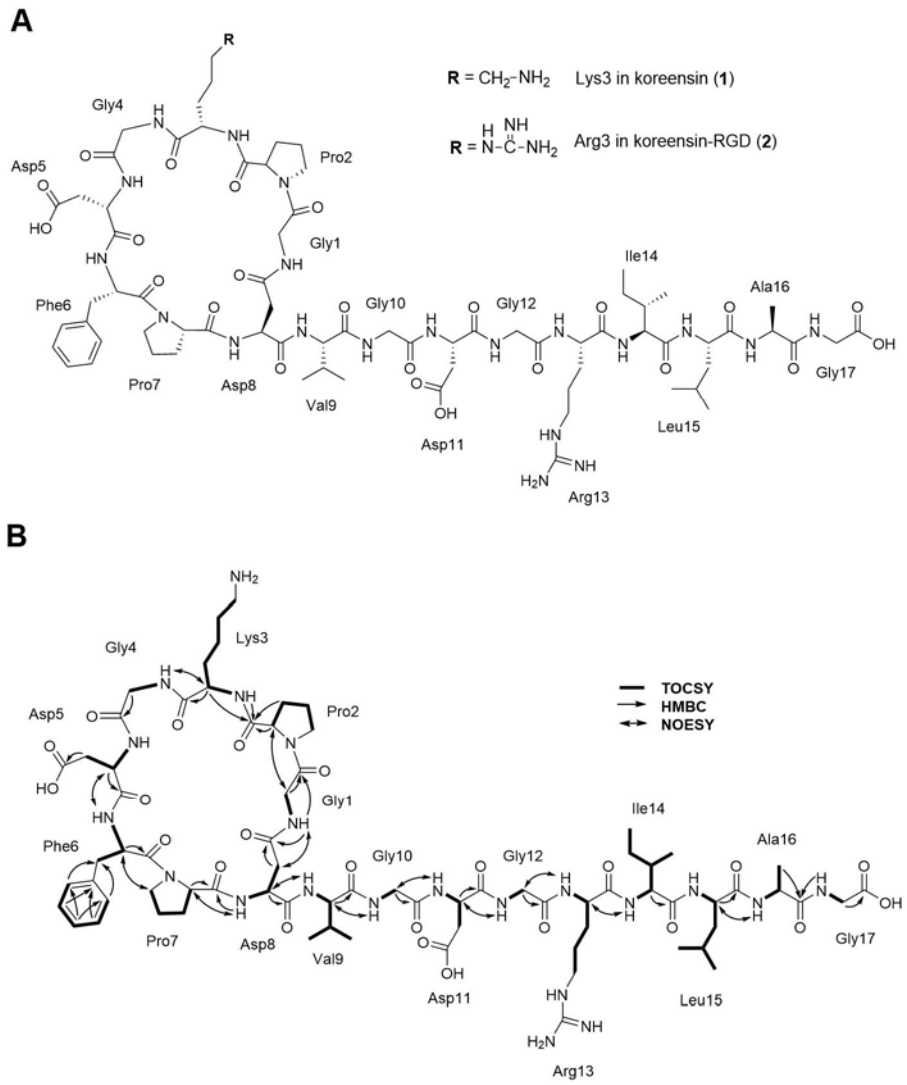

Figure 1. Fuwa et al.

496 Figure 2

497

498

499

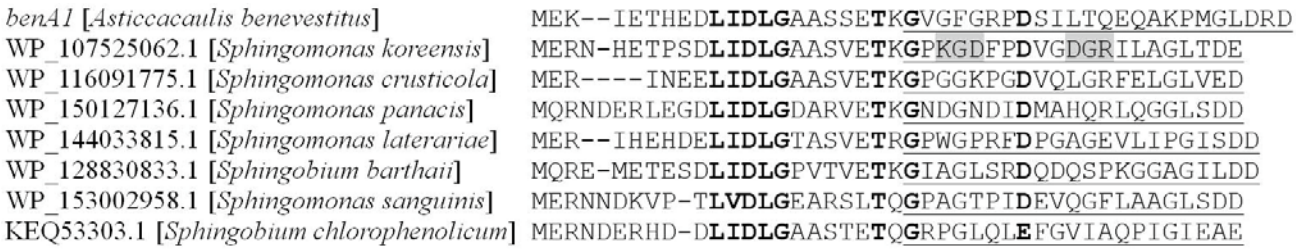

Figure 2. Fuwa et al.

500 Figure 3

501

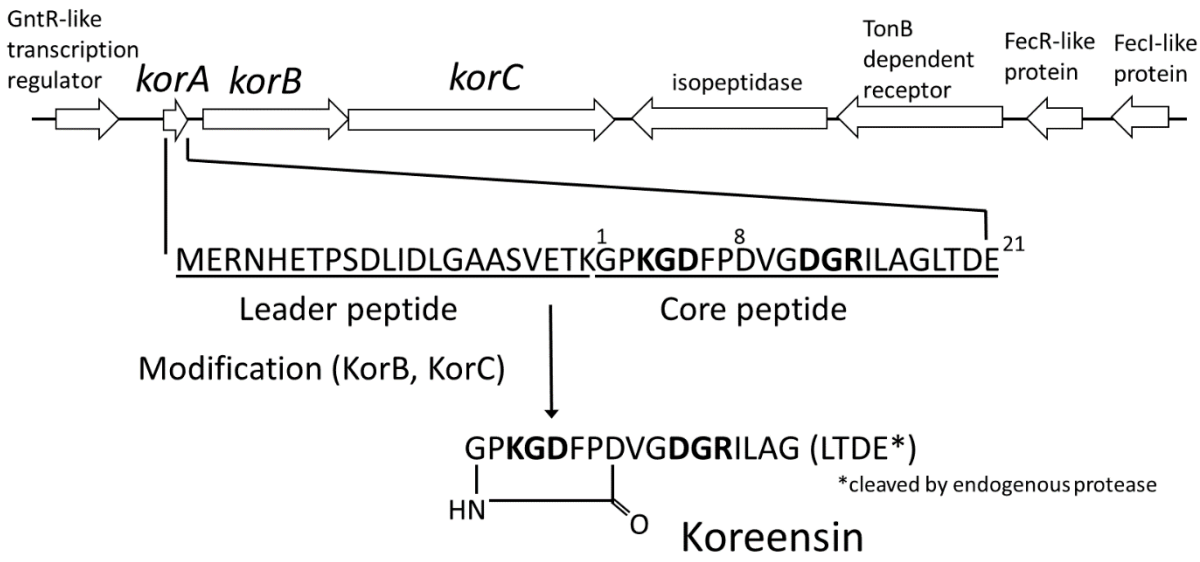

Figure 3. Fuwa et al.

502 Figure 4

503

504

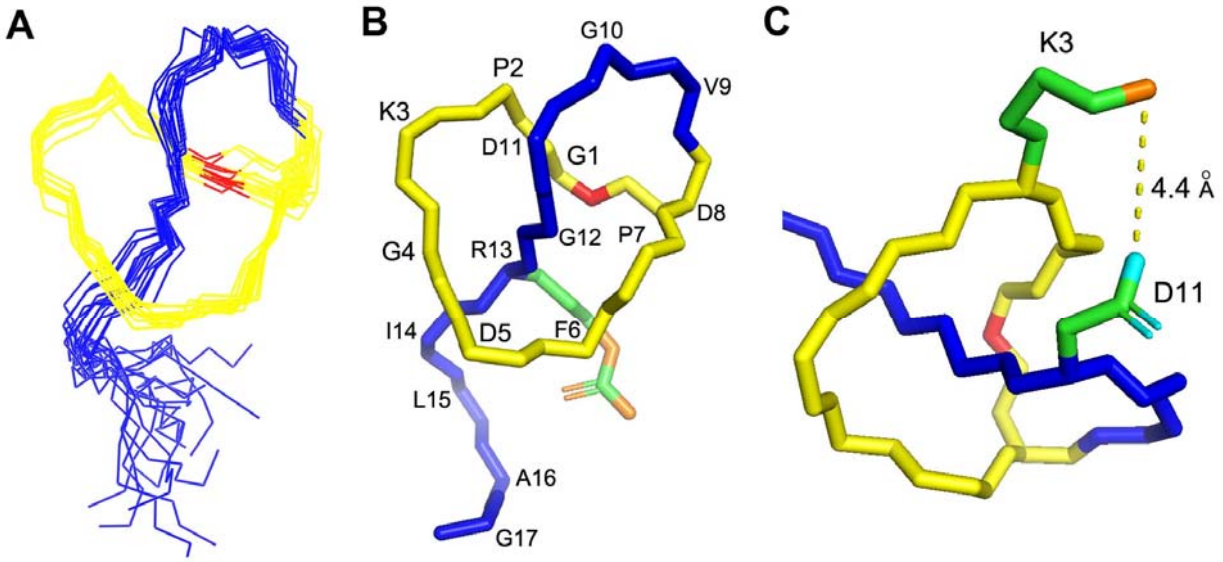

Figure 4. Fuwa et al.