# Heterologous expression of a cryptic gene cluster from a marine proteobacterium Thalassomonas actiniarum affords new lanthipeptides thalassomonasins A and B

Title: Heterologous expression of a cryptic gene cluster from a marine proteobacterium *Thalassomonas actiniarum* affords new lanthipeptides thalassomonasins A and B

Running title: Heterologous production of new lanthipeptides

Authors: Chanaphat Thetsana[1], Shinta Ijichi[2], Issara Kaweewan[2], Hiroyuki Nakagawa[3], and Shinya Kodani[1,2,4,5,*]

Affiliations: [1]Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka 422-8529, Japan; [2]Faculty of Agriculture, Shizuoka University, Shizuoka 422-8529, Japan; [3]Research center for Advanced Analysis, National Agriculture and Food Research Organization (NARO), Ibaraki 305-8642, Japan; [4]Shizuoka Institute for the Study of Marine Biology and Chemistry, Shizuoka University, Shizuoka 422-8529, Japan; [5]College of Agriculture, Academic Institute, Shizuoka University, Shizuoka 422-8529, Japan

*correspondence: 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

**KEYWORDS:** heterologous expression, lanthipeptide, biosynthesis, proteobacterium, *Thalassomonas actiniarum*

## Abstract

**Aims:** The aim of this study was to utilize a cryptic biosynthetic gene cluster of a marine proteobacterium *Thalassomonas actiniarum* for production of new lanthipeptides by heterologous expression system.

**Methods and Results:** Based on genome-mining, a new biosynthetic gene cluster of class I lanthipeptide was found in the genome sequence of a marine proteobacterium *Thalassomonas actiniarum.* Molecular cloning was performed to construct expression vector derived from commercial available plasmid pET-41a(+). Heterologous production of new lanthipeptides named thalassomonasins A and B was performed using the host *Escherichia coli* BL21(DE3) harboring the expression vector. The structure of thalassomonasin A was determined by interpretation of NMR and MS data. As a result, thalassomonasin A was determined to be a lanthipeptide with three units of lanthionine. The bridging pattern of the lanthionine rings in thalassomonasin A was determined by interpretation of NOESY data. The structure of thalassomonasin B was proposed by MS/MS experiment.

**Conclusions:** We succeeded in heterologous production of new class I lanthipeptides using a biosynthetic gene cluster of a marine proteobacterium *Thalassomonas actiniarum*.

**Significance and Impact of the Study:** To the best of our knowledge, this is the first report of heterologous production of lanthipeptides derived from proteobacterial origin. There are many cryptic biosynthetic gene clusters of this class of lanthipeptides in proteobacterial genomes. This study may lead to production of new lanthipeptides by utilizing the biosynthetic gene clusters.

## Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a large group of natural occurring peptides, which are classified into more than twenty subclasses (Budisa 2013; Montalban-Lopez et al. 2021). Lanthipeptides are one class of RiPPs, which possess unusual amino acids such as the thioether crosslinked amino acids: lanthionine (Lan) and methyllanthionine (MeLan) and the dehydrated amino acids: dehydroalanine (Dha) and dehydrobutyrine (Dhb). Initially, dehydrated amino acids Dha and Dhb are generated via dehydration of Ser and Thr, respectively (van der Donk and Nair 2014; Repka et al. 2017). Subsequent addition of Cys thiols to these dehydroamino acids afford Lan and MeLan, with formation of thioether bridges. Currently five groups of lanthipeptides (class I-V) have been reported according to biosynthetic system to give these unusual amino acids (van der Donk and Nair 2014; Repka et al. 2017; Ortiz-Lopez et al. 2020; Xu et al. 2020; Roman-Hurtado et al. 2021). Modification of class I lanthipeptide is accomplished by enzymes LanB and LanC (Repka et al. 2017). The enzyme LanB is a dehydratase to give dehydrated amino acids (Dha and Dhb) from Ser and Thr. Dehydration by LanB is taken place via glutamylation of hydroxy residues of Ser and Thr, followed by elimination (Ortega et al. 2015). The enzyme LanC is cyclase to give Lan and MeLan by generation of thioether bridges

between Cys and the dehydrated amino acid (Dha or Dhb). In biosynthesis of class II-IV lanthipeptides, one modification enzyme (LanM in class II, LanKC in class III, LanL in class IV) has two functions of dehydration of Ser and Thr and formation of thioether bond to give Lan and MeLan. Unlike LanB, the dehydration occurs via phosphorylation of hydroxy residues of Ser and Thr. Biosynthesis of class V lanthipeptide has three enzymes (LxmK: kinase, LxmX: dehydratase, LxmY: cyclase) to give Lan bridge (Ortiz-Lopez et al. 2020; Xu et al. 2020; Roman-Hurtado et al. 2021). In addition to dehydrated amino acids, Lan and MeLan, class V lanthipeptide contains D-Ala in the molecule. D-Ala is synthesized by stereochemically specific reduction of Dha caused by $F_{420}H_2$-dependent reductase: LxmJ.

So far, lanthipeptides have been reported mostly from Gram-positive bacteria including Actinobacteria and lactic acid bacteria (Knerr and van der Donk 2012). A large-scale bioinformatic analysis of actinobacterial genomes revealed a wide distribution of lanthipeptide like biosynthetic gene clusters (BGCs) (Zhang et al. 2015). The bacteria belonging to *Paenibacillus* genus were indicated to be promising bioresources for new lanthipeptide based on genome mining (Baindara et al. 2020). Based on a genome mining system RODEO (Tietz et al. 2017), nearly 8,500 lanthipeptide precursor peptide coding genes were identified from genomes of over

100,000 bacteria and archaea (Walker et al. 2020). These precursor peptide coding genes were indicated to be distributed in a broad range of bacterial phyla and the *Euryarchaeota* phylum of archaea. Heterologous expression of biosynthetic gene cluster (BGC) has been an efficient method to produce new lanthipeptide by utilizing cryptic BGCs found by genome mining. The class I lanthipeptides geobacillins were produced using a cryptic BGC of *Geobacillus thermodenitrificans* in *Escherichia coli* (Garg et al. 2012). A new duramycin/cinnamycin analogous peptide named kyamicin was produced using a cryptic BGC of the actinobacterium *Saccharopolyspora* sp. in the host actinobacterium *Streptomyces coelicolor* (Vikeli et al. 2020). A new two-component lantibiotic roseocin was produced by heterologous gene expression using a cryptic BGC of the actinobacterium *Streptomyces roseosporus* (Singh et al. 2020). Based on genome data of marine cyanobacterium *Prochlorococcus*, lanthipeptides named prochlorosins were produced by *in vitro* synthesis using lanthipeptide synthetase ProcM and precursor peptides (Mukherjee and van der Donk 2014; Yu et al. 2015; Jeanne Dit Fouque et al. 2021). The BGCs of this group of lanthipeptides are indicated to be distributed over the genomes of marine cyanobacteria (Cubillos-Ruiz et al. 2017; Bobeica and van der Donk 2018). As mentioned above, most lanthipeptides are mainly found from origins of Gram-positive bacteria and cyanobacteria.

From Gram-negative bacterium *Chitinophaga pinensis* (Bacteroidetes phylum), class I lanthipeptides pinensins A and B were isolated as antifungal reagents (Mohr et al. 2015). And the BGCs of pinensin-like lanthipeptides were indicated to exist in the genomes of Bacteroidetes (Caetano et al. 2020). The class I lanthipeptide PedA15.1 was produced by heterologous expression using BGC of *Pedobacter lusitanus* which belongs to Bacteroides (Bothwell et al. 2021). To the best of our knowledge, there are just two reports about lanthipeptides derived from Gram-negative bacterial origins (Caetano et al. 2020; Bothwell et al. 2021), although there are many BGCs of lanthipeptides found in the genome data of Gram-negative bacteria (Walker et al. 2020). On the basis of these backgrounds, we performed heterologous expression of a cryptic gene cluster of a marine proteobacterium *Thalassomonas actiniarum* (Hosoya et al. 2009) using *Escherichia coli* as the host. As a result, we succeeded in production of new lanthipeptides named thalassomonasins A and B. Here we describe heterologous production and structural determination of new lanthipeptides thalassomonasins A and B.

## Materials and methods

### Bacterial strains

Bacterial strains were obtained from Biological Resource Center, NITE (NBRC), Japan. The strains include a marine proteobacterium; *Thalassomonas actiniarum* (NBRC 104231$^{T}$), Gram-positive bacteria; *Bacillus subtilis* (NBRC 13719$^{T}$), *Micrococcus luteus* NBRC 3333 $^{T}$); *Staphylococcus aureus* (NBRC 100910 $^{T}$), and Gram-negative bacteria; *Escherichia coli* (NBRC 102203 $^{T}$); *Pseudomonas aeruginosa* (NBRC 12689 $^{T}$).

### Construction of the expression vector pET41a-104231CBA1

PCR amplification (KAPA HiFi HotStart ReadyMix, Kapa Biosystems, Inc., USA) was performed using primers (104231-F-XbaI and 104231-R-KpnI, Table S1) and genomic DNA of *T. actiniarum* as a template to obtain a gene fragment including *tlnC*, *tlnB* and *tlnA1*. The genomic DNA of *T. actiniarum* was obtained using DNA extraction kit (DNeasy Blood & Tissue, Qiagen, Netherlands). The PCR amplified gene fragment (*tlnC*, *tlnB* and *tlnA1*) and a pET-41a (+) vector were subjected to double digestion using the enzymes *XbaI* and *KpnI* (New England Biolabs, USA). To obtain the expression vector pET41a-104231CBA1 (Fig. S1), ligation of two DNA fragments was accomplished using Ligation-Convenience Kit (Nippon Gene Co., Ltd., Japan). The constructed plasmid pET41a-104231CBA1 was transformed into *E. coli* DH5α competent cells for cloning. LB agar medium supplemented with 30 µg/mL kanamycin was used for cultivation of the transformants. The expression vector pET41a-104231CBA1 was extracted by FastGene Plasmid Mini Kit (Nippon Genetics Co., Ltd., Japan) and introduced into *E. coli* BL21 (DE3) by electroporation.

### Construction of the expression vector pET41a-104231CBA2

A gene fragment including *tlnC* and *tlnB* was obtained by PCR amplification

(EmeraldAmp PCR Master Mix, Takara Bio Inc., Japan) using primers (104231-F-XbaI and 104231-R2-KpnI, Table S1) and a template (vector pET41a-104231CBA1). To obtain the vector pET41a-104231CB (Fig. S2), the enzymes *XbaI* and *KpnI* (New England Biolabs, USA) were used to digest a PCR amplified gene fragment and pET-41a (+) vector, followed by ligation using Ligation-Convenience Kit (Nippon Gene Co., Ltd., Japan). The constructed plasmid pET41a-104231CB was transformed into *E. coli* DH5α competent cells and cultured on LB agar medium supplemented with 30 μg/mL of kanamycin. The genomic DNA of *T. actiniarum* NBRC104231$^{T}$ was used as a template for PCR amplification along with primers (104231-A2-F-KpnI and 104231-A2-R-EcoRI, Table S1) to obtain a gene fragment of *tlnA2*. The gene fragment including *tlnA2* amplified by PCR and the vector pET41a-104231CB vector were subjected to double digestion using the enzymes *KpnI* and *EcoRI* (New England Biolabs, USA). The expression vector pET41a-104231CBA2 (Fig. S3) was constructed by ligation using Ligation-Convenience Kit (Nippon Gene Co., Ltd., Japan). The constructed plasmid pET41a-104231CBA2 was transformed into *E. coli* DH5α competent cells for cloning. LB agar medium supplemented with 30 μg/mL of kanamycin was used for cultivation of the transformants. The expression vector pET41a-104231CBA2 was extracted by FastGene Plasmid Mini Kit (Nippon Genetics Co., Ltd., Japan) and transformed into the cells *E. coli* BL21 (DE3) by electroporation.

**Heterologous production of thalassomonasins A and B**

The *E. coli* BL21 (DE3) harboring the vector (pET41a-104231CBA1 or pET41a-104231CBA2) was cultured on modified basal agar medium supplemented with 30

158 μg/mL kanamycin and 0.1 mM IPTG at 30°C for 24 h for the production of
159 thalassomonasins A and B, respectively. For extraction, the cells were collected from
160 modified basal agar medium by a laboratory spatula, followed by addition of MeOH.
161 After centrifugation, the MeOH extract was concentrated to aqueous residue by rotary
162 evaporation, and dissolved in distilled water. For production of thalassomonasin A,
163 HPLC separation using an ODS column (4.6 × 250 mm, Wakopak Handy-ODS,
164 Fujifilm Wako Pure Chemical Co., Japan) was applied. Elution was performed by an
165 isocratic mode (0 - 10 min, 21% MeCN containing 0.05% TFA, flow rate: 1 mL/min)
166 and a gradient mode (10- 20 min, from 21% MeCN containing TFA to 40% MeCN
167 containing TFA, flow rate: 1 mL/min), and UV detector was set at 220 nm. The
168 retention time of thalassomonasin A was 12.5 min (Fig. S4). Though the procedure, 3.8
169 mg of thalassomonasin A was obtained from 2 L culture (wet weight of cells: 9.0 g).
170 For production of thalassomonasin B, HPLC separation using C30 column (4.6 × 250
171 mm, Wakopak Navi C30-5, Fujifilm Wako Pure Chemical Co., Japan) was applied.
172 Elution was performed by a gradient mode (0 - 20 min, from 20% MeCN containing
173 TFA to 25% MeCN containing TFA, flow rate: 1 mL/min), and UV detector was set at
174 220 nm. The retention time of thalassomonasin B was 21.5 min (Fig. S5).
175 **Mass spectrometry experiments**

The accurate mass measurement was accomplished using an ESI Orbitrap mass spectrometer (Orbitrap Velos ETD, Thermo Fisher Scientific, Waltham, MA, USA). The peptide was diluted using 50% MeOH and supplied to ESI Orbitrap mass spectrometer with electrospray ionization (ESI) in the positive polarity by direct infusion. MALDI-TOF MS and MS/MS analysis was performed with a MALDI-TOF/TOF mass spectrometer (4800 Plus TOF/TOF analyzer, SCIEX, Redwood City, CA, USA). Each peptide was mixed with equal volume of α-Cyano-4-hydroxycinnamic acid (CHCA) (Shimadzu GLC Ltd., Tokyo, Japan) matrix solution (10 mg/mL in 50% MeCN containing 0.1% TFA), and the mixed solution (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. Calibration of the mass spectrometer was performed using standard reagents of polytyrosine (Thermo Fisher Scientific Pierce Biotechnology, Rockford, IL, USA) and the peptide mixture (Peptide Calibration Standard II, Bruker Daltonik GmbH, Bremen Germany), respectively, prior to the measurements.

**NMR experiments**

NMR sample was prepared by dissolving 3.3 mg of thalassomonasin A in 500 μL of DMSO-$d_6$. 1D and 2D NMR spectra ($^{1}H$, $^{13}C$, DEPT-135, DQF-COSY, TOCSY, NOESY, HMBC, and HSQC) were obtained using Bruker Avance800 spectrometer

194 with quadrature detection as described in the previous report (Kodani et al. 2017).

195 **Antibacterial activity assay of thalassomonasin A**

196 The antibacterial activity assay of thalassomonasin A was performed using minimum
197 inhibitory concentrations (MICs) test in 96-well microplates, following the previous
198 report (Suzuki et al. 2021). The bacteria strains (Gram-negative bacteria including
199 *Escherichia coli* and *Pseudomonas aeruginosa,* and Gram-positive bacteria including
200 *Bacillus subtilis*, *Staphylococcus aureus*, and *Micrococcus luteus*) were used for the
201 antibacterial activity assay. Tetracycline hydrochloride was used as positive control.
202 The MICs of tetracycline hydrochloride were 0.25 μg/mL (*E. coli*), 4.0 μg/mL (*P.*
203 *aeruginosa*), 0.06 μg/mL (*M. luteus*), 0.03 μg/mL (*S. aureus*), and 8.0 μg/mL (*B.*
204 *subtilis*).

205 **Results**

206 **Analysis of the biosynthetic gene cluster of thalassomonasins A and B.**

207 The previous genome mining report indicated putative precursor peptide encoding
208 genes in proteobacterial genomes (Walker et al. 2020). In the genome of a marine
209 proteobacterium *Thalassomonas actiniarum*, a typical BGC of class I lanthipeptide was
210 found (Fig. 1A). In the BGC, two precursor peptide coding genes (*tlnA1,* accession
211 number: WP_160298205.1 and *tlnA2,* accession number: WP_160298204.1) and

212 modification enzyme coding genes (*tlnB,* accession number: WP_044830464.1 and
213 *tlnC*: WP_044830465.1) are tandemly located with same direction (Fig. 1A). The
214 protein TlnB corresponds to LanB of class I lanthipeptide biosynthesis, which functions
215 as dehydratase. And the protein TlnC corresponds to LanC which affords Lan/MeLan
216 bridge in the molecule. Regarding the precursor peptides *tlnA1* and *tlnA2*, we performed
217 an alignment with other similar precursor peptide coding genes in proteobacterial
218 genomes (Fig. 1B). Normally precursor peptides have leader peptide region at *N*-
219 terminus and core peptide region at *C*-terminus. The alignment indicated leader peptide
220 and core peptide regions are flanked by double Gly motif (under lined in Fig. 1B).
221 Double Gly motif is possibly specific cleavage site of peptidase domain of dual-
222 functional ATP-binding cassette transporter (Bobeica et al. 2019). However, there is no
223 ATP-binding cassette transporter coding gene close to the BGC. In class I lanthipeptide,
224 FNLD box sequence is conserved in leader peptide sequence for recognition of
225 modification enzymes (LanB and LanC). In the leader peptide sequences (Fig. 1B),
226 FNLD motif was not found in amino acid sequences of these precursor peptides and no
227 specific conserved motif was found. Interestingly, at $2^{nd}$ -$12^{th}$ amino acids of *N*-
228 terminus, Lys rich region was found (bold letters in Fig. 1B). In the core peptide

sequences, there is no conserved motif. The positions of amino acids (Cys, Ser/Thr) which can be involved in Lan/MeLan were not conserved among the peptides (Fig.1B).

**Production of thalassomonasins A and B.**

For heterologous production of new lanthipeptides, we performed molecular cloning of a DNA region which included *tlnB*, *tlnC*, and either *tlnA1* or *tlnA2*. The expression vector pET41a-104231CBA1 (gene cluster: *tlnC-tlnB-tlnA1*, Fig. S1) was constructed by molecular cloning. For production of thalassomonasin B, the expression vector pET41a-104231CBA2 (gene cluster: *tlnC-tlnB-tlnA2*, Fig. S2 and S3) which contained a precursor peptide coding gene *tlnA2* instead of *tlnA1* was constructed by manipulating the vector pET41a-104231CBA1. After cultivation, the cells of *E. coli* BL21(DE3) possessing the vector pET41a-104231CBA1 were extracted with double volume of MeOH according to cell volume and the MeOH extract was analyzed by HPLC and ESI-MS. As a result, a new lanthipeptide named thalassomonasin A was detected in the MeOH extract of the transformant (Fig. S4). In the same manner with thalassomonasin A, thalassomonasin B was produced by *E. coli* BL21(DE3) possessing the vector pET41a-104231CBA2 (Fig. S5).

**Structure determination of thalassomonasins A and B.**

246 The molecular formula of thalassomonasin A was determined to be $C_{78}H_{118}N_{20}O_{22}S_3$
247 by accurate MS measurement (Fig. S47), since the ion peak was observed at $[M+2H]^{2+}$
248 892.4023 (calculated value: 892.4020). By MS/MS experiment, *b*-series ions (*b9*, *b10*,
249 *b11*, *b16*) and *y*-series ions (*y7*, *y8*, *y16*) were observed, which indicated partial amino
250 acid sequence of Lys-Pro in the middle, Val at *N*-terminus, and Phe at *C*-terminus (Fig.
251 2A, Fig. S48 - S52, and Table S2). We deduced that partial cyclic structures hampered
252 to give fragmentation on the MS/MS experiment. Thalassomonasin A was deduced to
253 be a 17 amino acid length peptide (VSVVHSCVCKPSEVDCF) synthesized from the
254 precursor peptide TlnA1 (Fig. 1B). Since the molecular formula of the peptide
255 (VSVVHSCVCKPSEVDCF) was $C_{78}H_{124}N_{20}O_{25}S_3$, thalassomonasin A was deduced to
256 be the peptide with removal of three units of $H_2O$, considering the discrepancy with the
257 molecular formula ($C_{78}H_{118}N_{20}O_{22}S_3$) obtained by the accurate MS measurement. To
258 determine the structure, the analyses of NMR spectra ($^1H$, $^{13}C$, DEPT-135, DQF-COSY,
259 TOCSY, NOESY, HMBC, and HSQC) were performed on thalassomonasin A
260 dissolved in 0.5 mL of DMSO-$d_6$ (Fig. S6 - S46). The spin system identification using
261 the 2D NMR spectral data indicated the presence of all amino acids including three
262 units of Lan (Ala-S-Ala), except for His (Tables S4-1 to S4-6 and Fig. 2B). Regarding
263 the His residue, protons and carbons of the imidazole ring in His in $^1H$, $^{13}C$, DEPT-135

and 2D NMR spectra were not observed with this NMR condition. However, α-, β-, and amide protons of the 5$^{th}$ residue from *N*-terminus were observed (Fig. 2B and Table S4-2), and we concluded that this residue was His based on the amino acid sequence of the precursor peptide TlnA1. The HMBC correlation was observed from amide proton in the amino acids (S-Ala2, Val3, Val4, S-Ala6, S-Ala9, Lys10, S-Ala12, Glu13, Val14, Asp15, S-Ala16, and Phe17) to carbonyl carbon of adjacent amino acids (Val1, S-Ala2, Val3, His5, Val8, S-Ala9, Pro11, S-Ala12, Glu13, Val14, Asp15, and S-Ala16), respectively (Fig. 2B). The HMBC correlations indicated four partial sequences (Val1-S-Ala2-Val3-Val4, His5-S-Ala6, Val8-S-Ala9-Lys10, and Pro11-S-Ala12-Glu13-Val14-Asp15-S-Ala16-Phe17). In addition to the analysis of HMBC correlations for amino acid sequence elucidation, the sequential walk using the NOESY correlation (Fig. 2B) between α-proton and amide proton of adjacent amino acids revealed the two partial sequences (Val1 to Lys10 and Pro11 to Phe17). The NOESY correlation between β-protons of Lys10 and δ-proton of Pro11 indicated the connection between Lys10 and Pro11. As mentioned above, the amino acid sequence of thalassomonasin A was determined from Val1 to Phe17 (Fig. 2B). Regarding determination of bridging pattern of Lan (Ala-S-Ala), NOESY correlations were used. The NOESY correlations between β-protons were observed (red double ended arrows in Fig. 2B and Fig. S24), which

282 indicated the connections of S-Ala residues (S-Ala2/S-Ala7, S-Ala6/S-Ala9, S-Ala12/S-
283 Ala16). Each Lan residue was indicated to have thioether bond which had the bridging
284 pattern depicted in Fig. 2B.
285 Isolation of thalassomonasin B was performed in the similar manner with
286 thalassomonasin A using HPLC (Fig. S5). The extract contained many related peptides
287 derived from the core peptide of the precursor A2, judging by ESI-MS experiment
288 (Data not shown). During production in cells of *E. coli*, endogenous protease of *E. coli*
289 may digest and give various length peptides from the precursor A2 (Fig. S5). We
290 succeeded in identification of one compound named thalassomonasin B (Fig. S5). The
291 molecular formula of thalassomonasin B was determined to be $C_{129}H_{188}N_{34}O_{38}S_3$ by
292 accurate MS measurement (Fig. S53), since the ion peak was observed at $[M + 3H]^{+3}$
293 973.4402 (calculated value: 973.4401, in Fig. S53). It was difficult to obtain enough
294 amount of thalassomonasin B for NMR analysis, since the production yield was low
295 possibly due to digestion of endogenous protease of *E. coli* (Fig. S5). To analyze the
296 chemical structure, we performed MS/MS analysis on thalassomonasin B to obtain
297 amino acid sequence (Table S3, Fig.S54-S64). As a result, *b*-series ions *b23-b26*
298 indicated the *C*-terminus amino acid sequence of Gln-Val-Glu-Met (Fig. 2A). And *b*-
299 series ions *b8-b16* indicated the amino acid sequence of Glu-His-Ile-Dhb-Dhb-His-Dhb-

300 Tyr. Considering the partial amino acid sequence obtained by MS/MS experiment and
301 the molecular formula, we proposed that thalassomonasin B was the peptide
302 biosynthesized from the peptide (GAVTLKDCEHITTHTYTVAWGDCQVEM) with
303 removal of five units of water. Since the cyclic part tends not to give fragmentation on
304 MS/MS experiment, we proposed that the structure of thalassomanasin B contained two
305 units of MeLan rings and three units Dhb, with MeLan bridging pattern as shown in Fig.
306 2A.

307 **Antibacterial activity assay on thalassomonasin A.**

308 Thalassomonasin A was subjected to the antibacterial activity assays using the testing
309 bacteria (Gram-negative bacteria including *Escherichia coli* and *Pseudomonas*
310 *aeruginosa,* and Gram-positive bacteria including *Bacillus subtilis*, *Staphylococcus*
311 *aureus*, and *Micrococcus luteus*) following the previous report (Suzuki et al. 2021).
312 Since the yield of thalassomonasin B was too low, the antibacterial activity of
313 thalassomonasin B was not tested. As a result, thalassomonasin A did not show
314 antibacterial activity against all the tested bacteria at a final concentration of 64 μg/mL.
315

## Discussion

Thalassomonasins A and B were obtained by *in vivo* production in the cell of *Eschelichia coli*. During the biosynthesis, proteolysis in *E. coli* occurs to give thalassomonasins A and B. The actual cleavage sites of the precursor peptides TlnA1 and TlnA2 in the biosynthesis of the original strain *Thalassomonas actiniarum* may be different with those of thalassomonasins A and B. The precursor peptides TlnA1 and TlnA2 have double Gly motif in the sequence (specific cleavage site for protease domain of dual functional transporter, underlined in Fig. 1B) while dual functional transporter coding gene is lacking in the BGC. As the matter of fact, we tried to detect lanthipeptides in the original strain *Thalassomonas actiniarum* by chemical analysis. Briefly, *T. actiniarum* was cultured using marine broth 2216 medium. Expected peptides were not detected in the culture by analysis using HPLC and ESI-MS (data not shown). So far, the cleavage sites of the precursor peptides TlnA1 and TlnA2 in the original strain *T. actiniarum* are not clear. In this study, thalassomonasin A did not show antibacterial activity. However, different cleavage site may possibly confer antibacterial activity to thalassomonasin A. Further *in vitro* production experiment is needed to clarify this point.

In this study, heterologous production of thalassomonasins A and B was accomplished by co-expression of three genes (*tlnA1*/*tlnA2*, *tlnB*, *tlnC*) in *E. coli*. The precursor peptides TlnA1 and TlnA2 contained a leader peptide and a core peptide with double Gly motif in the middle. Normally, leader peptide region contains conserved FNLD sequence in class I lanthipeptide biosynthesis. This conserved motif is needed for recognition of modification enzymes. The interaction between dehydratase NisB and the leader peptide of NisA is caused largely by hydrophobic packing interactions (Ortega et al. 2015). In the FNLD sequence of NisA, the hydrophobic side chain of Phe is inserted in a hydrophobic pocket of NisB. However, FNLD motif was not found in leader sequences of TlnA1 and TlnA2 and other similar peptides (Fig. 1B). And no clear conserved motif was found among leader peptides of similar peptides except for Lys rich region ($2^{nd}$-$11^{th}$ amino acids). In the previous report (Bothwell et al. 2021), Lys rich region in the leader peptide was reported in the biosynthesis of a class I lanthipeptide PedA15.1. We proposed that this Lys rich region may be the modification enzyme recognition motif of this group of class I lanthipeptides. In the biosynthesis of microviridin which belongs to RiPPs, α-helix structure in leader peptide is indicated to be the modification enzyme recognition motif (Li et al. 2016). The leader peptide region of the precursor peptide of lipolanthine, a novel type of class-III lanthipeptide, was

351 indicated to have α-helix structure for the modification enzyme to recognize (Wiebach
352 et al. 2020). We subjected the amino acid sequence of TlnA1 and A2 to secondary
353 structure prediction software CRNPRED (Kinjo and Nishikawa 2006). As a result, the
354 sequence LTLKKKAVKQL in TlnA1 and LTLKKKNLA in TlnA2 were indicated to
355 have possibility of α-helix formation (underlined letters in Fig. 3). In the biosynthesis of
356 thalassomonasins, α-helix structure may be the recognition motif for modification
357 enzymes like microviridin (Li et al. 2016) and lipolanthine (Wiebach et al. 2020).
358 In the biosynthesis of thalassomonasin A, Ser residues of core peptide in the
359 precursor TlnA1 is deduced to be dehydrated by dehydratase TlnB to give Dha (Fig. 3).
360 By the function of cyclase TlnC, Cys residue form thioether bond with Dha which is 3-5
361 amino acids away to the side of *N*-terminus (Fig. 3). This bridging formation pattern
362 seems to be similar with other class I lanthipeptides such as nisin. The Lan/MeLan ring
363 formation normally occurs with *C* to *N* terminus direction from Cys to Ser/Thr on core
364 peptide in biosynthesis of class I lanthipeptides (Fig. 4). The topologies of Lan/MeLan
365 rings of the class I lanthipeptides are summarized in Fig. 4. The lanthipeptides nisin and
366 subtilin have five rings (rings A-E) in common. The topology of rings A-C seems to be
367 conserved in class I lanthipeptides of Gram-positive bacteria, except for cebulantin
368 (Moon et al. 2019). Regarding lanthipeptides derived from Gram-negative bacteria, the

lanthipeptide pinensin A (Fig. 4) has two units of unusual large size MeLan rings (Mohr et al. 2015). Recently unusual ring formation pattern was indicated in class I lanthipeptide PedA15.1 (Fig. 4) from *Pedobacter lusitanus* which belongs to Bacteroides. The lanthipeptide PedA15.1 has three Lan rings, and one Lan ring is uniquely formed with *N* to *C* terminus direction of Cys to Ser which is opposite of the other Lan rings (Bothwell et al. 2021). In the present report, thalassomonasins A and B have different Lan/MeLan topologies, compared with other lanthipeptides. Interestingly, the amino acid sequences and bridging patterns of thalassomonasins A and B are completely different (Fig. 4), which indicates the possibility that the modification enzymes TlnB and TlnC may recognize a wide variety of amino acid sequences in core peptide.

The alignment of amino acid sequences of TlnB with other LanBs in class I lanthipeptide biosynthesis indicated the presence of typical glutamylation and elimination domains in TlnB (Fig. S65). The domains for enolate protonation and zinc-binding were found in TlnC by alignment of amino acid sequences of TlnC with other LanCs (Fig. S65). In the process of dehydration of Ser or Thr by dehydratase (LanB), glutamylation of hydroxy residue occurs. Then dehydrated amino acid is generated with removal of glutamyl residue. Glutamyl residue is provided from glutamyl tRNA (tRNA-

Glu). In the case of heterologous production of class I lanthipeptide, tRNA-Glu and glutamyl tRNA synthase (GluRS) are required from the same bacterium which have dehydratase. Because lanthionine dehydratase is indicated to recognize specific tRNA-Glu of same bacterial origin (Hudson et al. 2015; Ozaki et al. 2016; Bothwell et al. 2021). In this report, we observed production of class I lanthipeptides thalassomonasins A and B without introducing tRNA-Glu and GluRS of *Thalassomonas actiniarum* into the co-expression system. The dehydratase TlnB seems to utilize tRNA-Glu of *E. coli* origin (Fig. 3), since the sequences of tRNA-Glu are very similar between the expression host, *E. coli*, and the origin of the BGC, *T. actiniarum* (Fig. S66). In addition, the amino acid sequence of GluRS of *T. actiniarum* (accession number: WP_044831289.1) has a high similarity (identity, 66.17%) with that of *E. coli* (accession number: P04805.1). These results indicate that *E. coli* gene expression system may be suitable for heterologous production of class I lanthipeptide of proteobacteria which have similar tRNA-Glu and GluRS.

**Author contributions**

CT performed molecular cloning and isolation of compounds. SI performed molecular cloning. IK and HN performed analysis using NMR and MS. SK designed the

405 experiments and wrote the manuscript. All the authors gave intellectual input and
406 critically revised the manuscript.
407
408 **Acknowledgements**
409 We thank Ms. Tomoko Sato for measurement of MS data. This study was supported by
410 Grants-in-Aid for Scientific Research: JSPS KAKENHI grant number 20K05848 and
411 21F21095, Institute for Fermentation, OSAKA (IFO): grant number G-2021-3-010,
412 and Kobayashi Foundation: grant number 196. I. K. was supported by JSPS
413 Postdoctoral Fellowships for Research in Japan (Standard Program).
414 **Conflict of Interest**
415 The authors declare that they have no conflict of interest.
416 **Ethical statement**
417 This article does not contain any studies with human participants or animals performed
418 by any of the authors.
419
420

## References


Baindara, P., Nayudu, N. and Korpole, S. (2020) Whole genome mining reveals a diverse repertoire of lanthionine synthetases and lanthipeptides among the genus *Paenibacillus*. *J Appl Microbiol* **128**, 473-490.

Bobeica, S.C., Dong, S.H., Huo, L., Mazo, N., McLaughlin, M.I., Jimenez-Oses, G., Nair, S.K. and van der Donk, W.A. (2019) Insights into AMS/PCAT transporters from biochemical and structural characterization of a double glycine motif protease. *Elife* **8**, e42305.

Bobeica, S.C. and van der Donk, W.A. (2018) The enzymology of prochlorosin biosynthesis. *Methods Enzymol* **604**, 165-203.

Bothwell, I.R., Caetano, T., Sarksian, R., Mendo, S. and van der Donk, W.A. (2021) Structural analysis of class I lanthipeptides from *Pedobacter lusitanus* NL19 reveals an unusual ring pattern. *ACS Chem Biol* **16**, 1019-1029.

Budisa, N. (2013) Expanded genetic code for the engineering of ribosomally synthetized and post-translationally modified peptide natural products (RiPPs). *Curr Opin Biotechnol* **24**, 591-598.

Caetano, T., van der Donk, W. and Mendo, S. (2020) Bacteroidetes can be a rich source of novel lanthipeptides: The case study of *Pedobacter lusitanus*. *Microbiol Res* **235**,

126441.

Cubillos-Ruiz, A., Berta-Thompson, J.W., Becker, J.W., van der Donk, W.A. and Chisholm, S.W. (2017) Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci U S A* **114**, E5424-E5433.

Garg, N., Tang, W., Goto, Y., Nair, S.K. and van der Donk, W.A. (2012) Lantibiotics from *Geobacillus thermodenitrificans*. *Proc Natl Acad Sci U S A* **109**, 5241-5246.

Hosoya, S., Adachi, K. and Kasai, H. (2009) *Thalassomonas actiniarum* sp. nov. and *Thalassomonas haliotis* sp. nov., isolated from marine animals. *Int J Syst Evol Microbiol* **59**, 686-690.

Hudson, G.A., Zhang, Z., Tietz, J.I., Mitchell, D.A. and van der Donk, W.A. (2015) *In vitro* biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J Am Chem Soc* **137**, 16012-16015.

Jeanne Dit Fouque, K., Hegemann, J.D., Santos-Fernandez, M., Le, T.T., Gomez-Hernandez, M., van der Donk, W.A. and Fernandez-Lima, F. (2021) Exploring structural signatures of the lanthipeptide prochlorosin 2.8 using tandem mass spectrometry and trapped ion mobility-mass spectrometry. *Anal Bioanal Chem*, 4815–4824.

Kinjo, A.R. and Nishikawa, K. (2006) CRNPRED: highly accurate prediction of one-

457 dimensional protein structures by large-scale critical random networks. *BMC*
458 *Bioinformatics* **7**, 401.
459 Knerr, P.J. and van der Donk, W.A. (2012) Discovery, biosynthesis, and engineering of
460 lantipeptides. *Annu Rev Biochem* **81**, 479-505.
461 Kodani, S., Inoue, Y., Suzuki, M., Dohra, H., Suzuki, T., Hemmi, H. and Ohnishi-
462 Kameyama, M. (2017) Sphaericin, a lasso peptide from the rare actinomycete
463 *Planomonospora sphaerica*. *Eur J Org Chem* **8**, 1177-1183.
464 Li, K., Condurso, H.L., Li, G., Ding, Y. and Bruner, S.D. (2016) Structural basis for
465 precursor protein-directed ribosomal peptide macrocyclization. *Nat Chem Biol* **12**, 973-
466 979.
467 Mohr, K.I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J.,
468 Gerth, K., Stadler, M. and Muller, R. (2015) Pinensins: the first antifungal lantibiotics.
469 *Angew Chem Int Ed Engl* **54**, 11254-11258.
470 Montalban-Lopez, M., Scott, T.A., Ramesh, S., Rahman, I.R., van Heel, A.J., Viel, J.H.,
471 Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., Grande Burgos, M.J., Hill, C.,
472 Kim, S., Koehnke, J., Latham, J.A., Link, A.J., Martinez, B., Nair, S.K., Nicolet, Y.,
473 Rebuffat, S., Sahl, H.G., Sareen, D., Schmidt, E.W., Schmitt, L., Severinov, K.,
474 Sussmuth, R.D., Truman, A.W., Wang, H., Weng, J.K., van Wezel, G.P., Zhang, Q.,

475 Zhong, J., Piel, J., Mitchell, D.A., Kuipers, O.P. and van der Donk, W.A. (2021) New
476 developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* **38**, 130-
477 239.
478 Moon, K., Xu, F. and Seyedsayamdost, M.R. (2019) Cebulantin, a cryptic lanthipeptide
479 antibiotic uncovered using bioactivity-coupled HiTES. *Angew Chem Int Ed Engl* **58**,
480 5973-5977.
481 Mukherjee, S. and van der Donk, W.A. (2014) Mechanistic studies on the substrate-
482 tolerant lanthipeptide synthetase ProcM. *J Am Chem Soc* **136**, 10450-10459.
483 Ortega, M.A., Hao, Y., Zhang, Q., Walker, M.C., van der Donk, W.A. and Nair, S.K.
484 (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB.
485 *Nature* **517**, 509-512.
486 Ortiz-Lopez, F.J., Carretero-Molina, D., Sanchez-Hidalgo, M., Martin, J., Gonzalez, I.,
487 Roman-Hurtado, F., de la Cruz, M., Garcia-Fernandez, S., Reyes, F., Deisinger, J.P.,
488 Muller, A., Schneider, T. and Genilloud, O. (2020) Cacaoidin, first member of the new
489 lanthidin RiPP family. *Angew Chem Int Ed Engl* **59**, 12654-12658.
490 Ozaki, T., Kurokawa, Y., Hayashi, S., Oku, N., Asamizu, S., Igarashi, Y. and Onaka, H.
491 (2016) Insights into the biosynthesis of dehydroalanines in goadsporin. *Chembiochem*
492 **17**, 218-223.

493 Repka, L.M., Chekan, J.R., Nair, S.K. and van der Donk, W.A. (2017) Mechanistic
494 understanding of lanthipeptide biosynthetic enzymes. *Chem Rev* **117**, 5457-5520.
495 Roman-Hurtado, F., Sanchez-Hidalgo, M., Martin, J., Ortiz-Lopez, F.J. and Genilloud,
496 O. (2021) Biosynthesis and heterologous expression of cacaoidin, the first member of
497 the lanthidin family of RiPPs. *Antibiotics (Basel)* **10**.
498 Singh, M., Chaudhary, S. and Sareen, D. (2020) Roseocin, a novel two-component
499 lantibiotic from an actinomycete. *Mol Microbiol* **113**, 326-337.
500 Suzuki, M., Komaki, H., Kaweewan, I., Dohra, H., Hemmi, H., Nakagawa, H.,
501 Yamamura, H., Hayakawa, M. and Kodani, S. (2021) Isolation and structure
502 determination of new linear azole-containing peptides spongiicolazolicins A and B from
503 *Streptomyces* sp. CWH03. *Appl Microbiol Biotechnol* **105**, 93-104.
504 Tietz, J.I., Schwalen, C.J., Patel, P.S., Maxson, T., Blair, P.M., Tai, H.C., Zakai, U.I.
505 and Mitchell, D.A. (2017) A new genome-mining tool redefines the lasso peptide
506 biosynthetic landscape. *Nat Chem Biol* **13**, 470-478.
507 van der Donk, W.A. and Nair, S.K. (2014) Structure and mechanism of lanthipeptide
508 biosynthetic enzymes. *Curr Opin Struct Biol* **29**, 58-66.
509 Vikeli, E., Widdick, D.A., Batey, S.F.D., Heine, D., Holmes, N.A., Bibb, M.J., Martins,
510 D.J., Pierce, N.E., Hutchings, M.I. and Wilkinson, B. (2020) *In situ* activation and

heterologous production of a cryptic lantibiotic from an african plant ant-derived *Saccharopolyspora* species. *Appl Environ Microbiol* **86**.

Walker, M.C., Eslami, S.M., Hetrick, K.J., Ackenhusen, S.E., Mitchell, D.A. and van der Donk, W.A. (2020) Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics* **21**, 387.

Wiebach, V., Mainz, A., Schnegotzki, R., Siegert, M.J., Hugelland, M., Pliszka, N. and Sussmuth, R.D. (2020) An Amphipathic Alpha-Helix Guides Maturation of the Ribosomally-Synthesized Lipolanthines. *Angew Chem Int Ed Engl* **59**, 16777-16785.

Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S.X., Lin, S., Deng, Z. and Tao, M. (2020) Functional genome mining reveals a class V lanthipeptide containing a D-amino acid Introduced by an F420 $H_2$ -dependent reductase. *Angew Chem Int Ed Engl* **59**, 18029-18035.

Yu, Y., Mukherjee, S. and van der Donk, W.A. (2015) Product formation by the promiscuous lanthipeptide synthetase ProcM is under kinetic control. *J Am Chem Soc* **137**, 5140-5148.

Zhang, Q., Doroghazi, J.R., Zhao, X., Walker, M.C. and van der Donk, W.A. (2015) Expanded natural product diversity revealed by analysis of lanthipeptide-like gene

529 clusters in actinobacteria. *Appl Environ Microbiol* **81**, 4339-4350.

530

531

532 Figure legends

533 Figure 1 A) biosynthetic gene cluster for thalassomonasins A and B, red: precursor

534 peptide coding gene, blue: dehydratase, yellow: cyclase, B) alignment of amino acid

535 sequences of precursor peptide coding genes related to thalassomonasins A and B,

536 amino acid residues which possibly modified are highlighted with colors (light blue:

537 Cys, red: Ser/Thr). Underlined double Gly motif is a specific cleavage site of protease

538 domain of dual-functional ATP-binding cassette transporter.

539 Figure 2 A) fragmentation ions obtained by MS/MS experiment on thalassomonasins A

540 and B, B) Key correlations of NMR experiments for structure determination of

541 thalassomonasin A

542 Figure 3 Proposed biosynthetic pathway for thalassomonasin A, underlined sequence is

543 Lys rich region possibly recognized by modification enzymes.

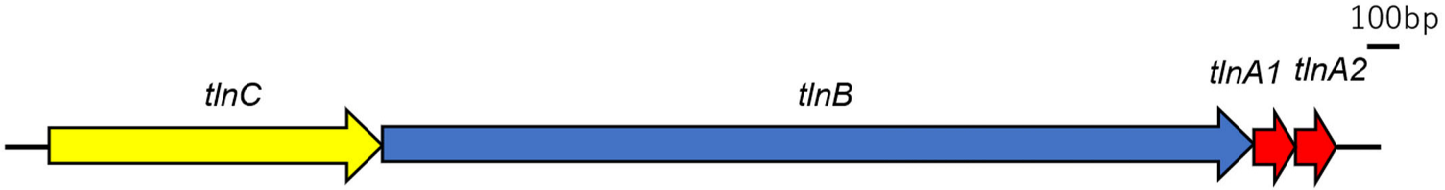
544 Figure 4 Lan/MeLan bridging patterns of class I lanthipeptides, dehydrated amino acids:

545 yellow box, Ser/Thr forming Lan/MeLan ring: red box, Cys forming Lan/MeLan ring:

546 blue box; thioether bridging patterns are indicated by lines.

547

A




B

*Thalassomonas actiniarum* (*tln A1*) MKLTLKKKAVKQLSNDVNLLPKKETAAIAGGLASANKTIVSVVHSCVCKPSEVDCF
*Thalassomonas actiniarum* (*tln A2*) MKLTLKKKNLANLSKDLNELPNEQTKDINGGAVTLKDCEHITTHTYTVAWGDCQVEM
*Thalassomonas viridans* MKLTLKKKNLANLSKNLNEVPNEQTGKINGGKLATSERDCDHITTHTYTVAWGDCGVEM
*Thalassomonas viridans* MKLSLKKKNLTNLSKDLNELPNEQTKDINGGAGTKDCQYPTTHTYTVAWGDCQAEM
*Thalassomonas viridans* MKLTLKKNAVKQLSNDAKALPKQQTAAIAGGQLQTNVKCVVSVKHSCVCLPSEVDCS
*Pseudoalteromonas luteoviolacea* MKLTVTKKNIKNLSQDNKRLPSEMTKFINGGTMPQLRADSSDTSVTGPSSPSNQD
*Pseudoalteromonas luteoviolacea* MKLAVTKKNIKNLSQDNKRLPSEMTKFINGGTMAQLRDSSDTSVTVPSGPTHQL
*Pseudoalteromonas luteoviolacea* MKLAITKKNIKNLSQDKKRLPANMTKFINGGSMAQPQDSSDTSVTVPSGPTHQL
*Pseudoalteromonas piscicida* MKLALNKKCIKALSNDSKSIPANMTPQIAGGRMQADYLLTDRFGDCWTGRNHSCREHTCQIP
*Pseudoalteromonas flavipulchra* MKLSLNKKKIKALSADKQTVPVVMTPQVAGGALRFSQDRLTHCVANTCGRHTCPDIY
*Pseudoalteromonas peptidolytica* MLKLKLIKKNLKQLNDKAHIDGSMTKKIAGGNTDGIKTTSHISQPEPDPYHSNSLTKPQL
*Pseudoalteromonas peptidolytica* MKLQLNKKNIKALSNDGKSVPAAMTPQVVGGNMQPLSPVGSNYCSDDYNCWTGRPRSCREHTCQME
*Pseudoalteromonas amylolytica* MKLKLNKKKIKNLSNDAHKLQQDMTPQVAGGIYYSRAGQFGCGSWMFCELPTEDKNNKM
*Pseudoalteromonas caenipelagi* MKLKIKKNIVKQLVGGTALNQKQTAEIA-GGMKETNGCPITSKNHCTLAFK

Figure 1. Thetsana et al.

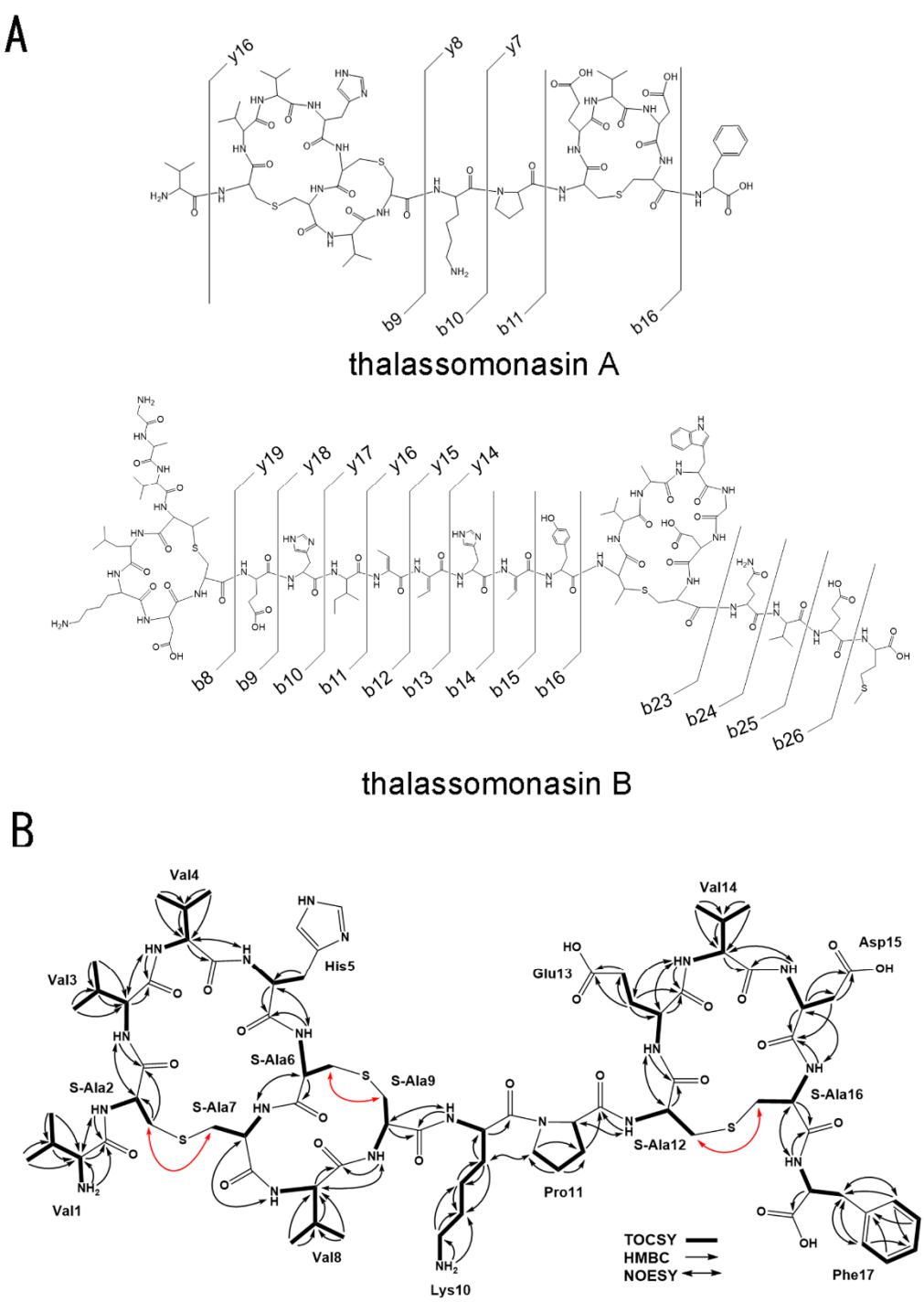



Fig. 2 Thetsana et al.

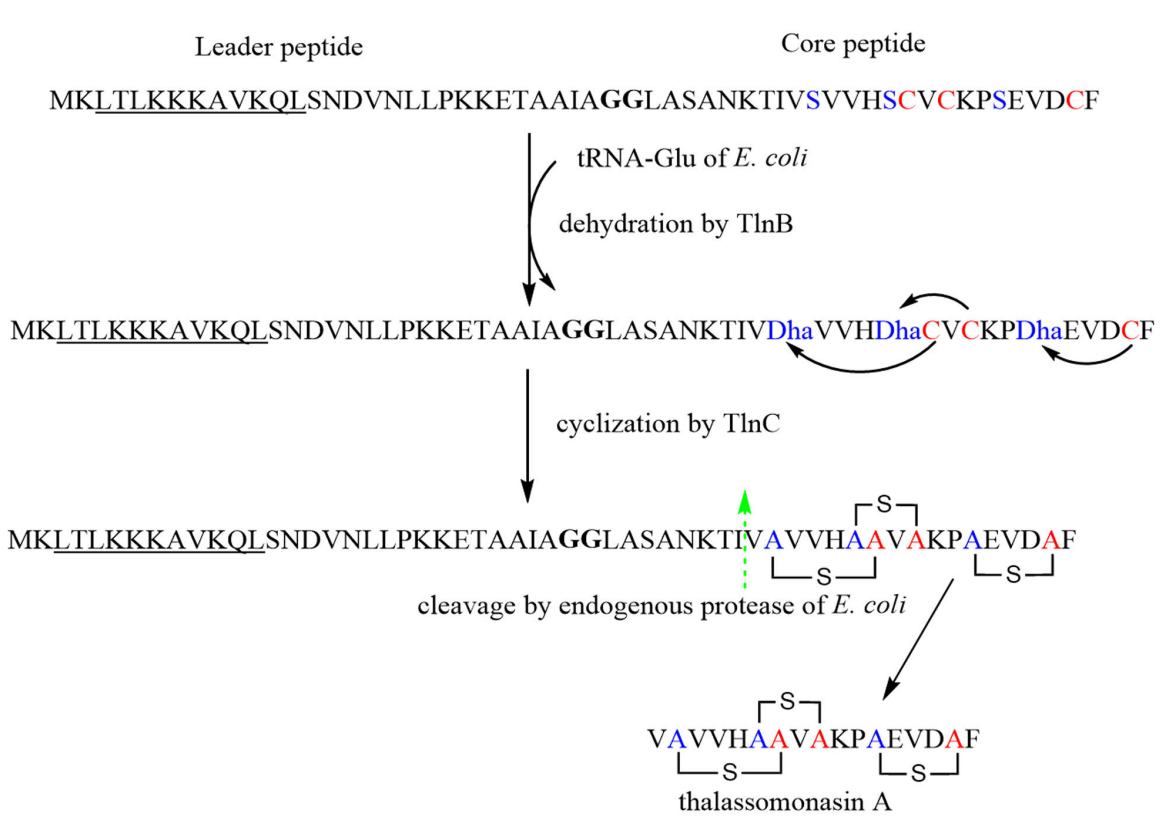



**Fig. 3. Thetsana et al.**

568

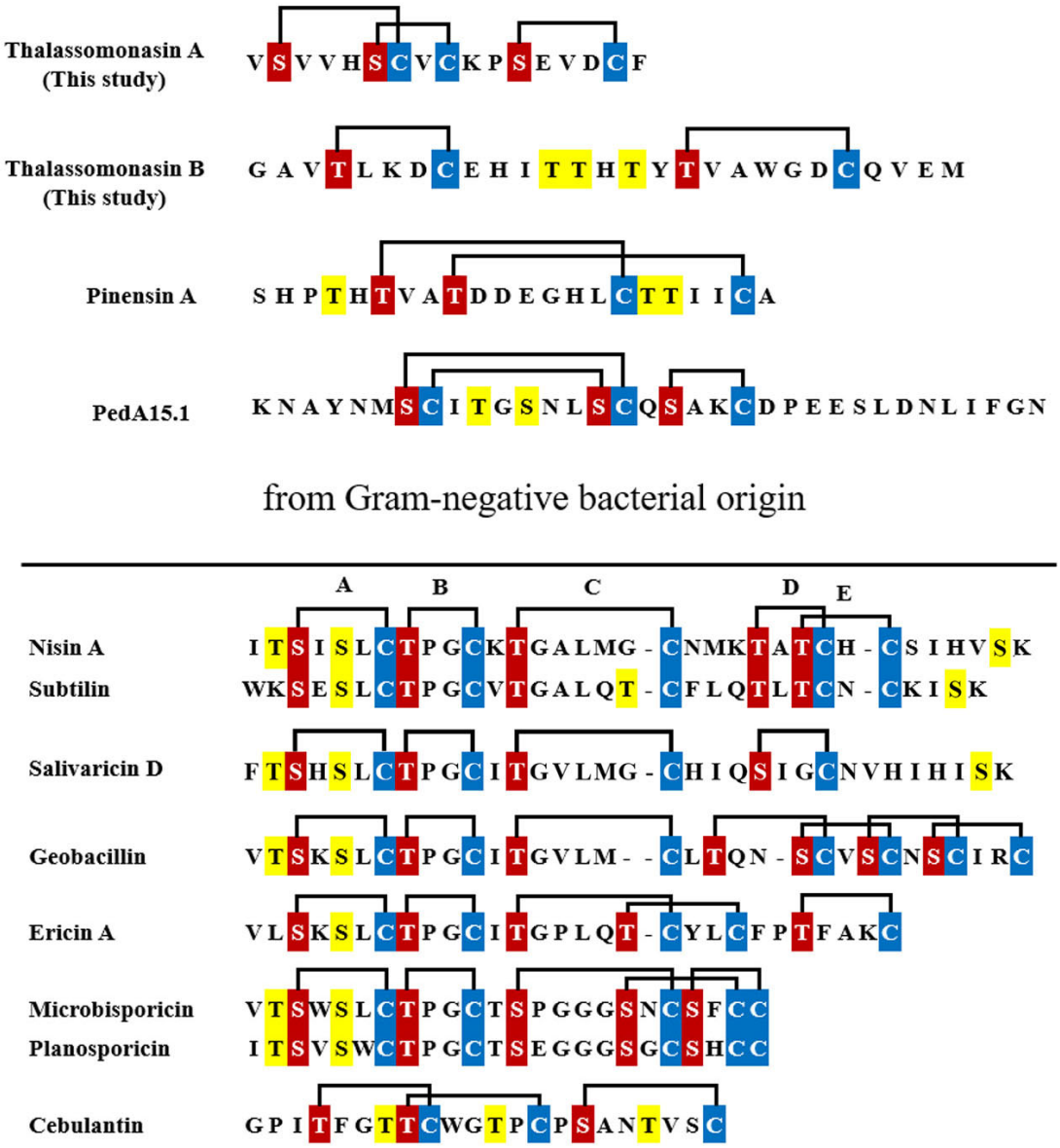



Fig. 4. Thetsana et al.