Biosynthetic studies on a D-tryptophan-containing lasso peptide, MS-271

D-トリプトファン含有ラッソペプチド天然物 MS-271 の生合成研究

Feng Zhi

Graduate School of Chemical Sciences and Engineering
Hokkaido University

Abstract

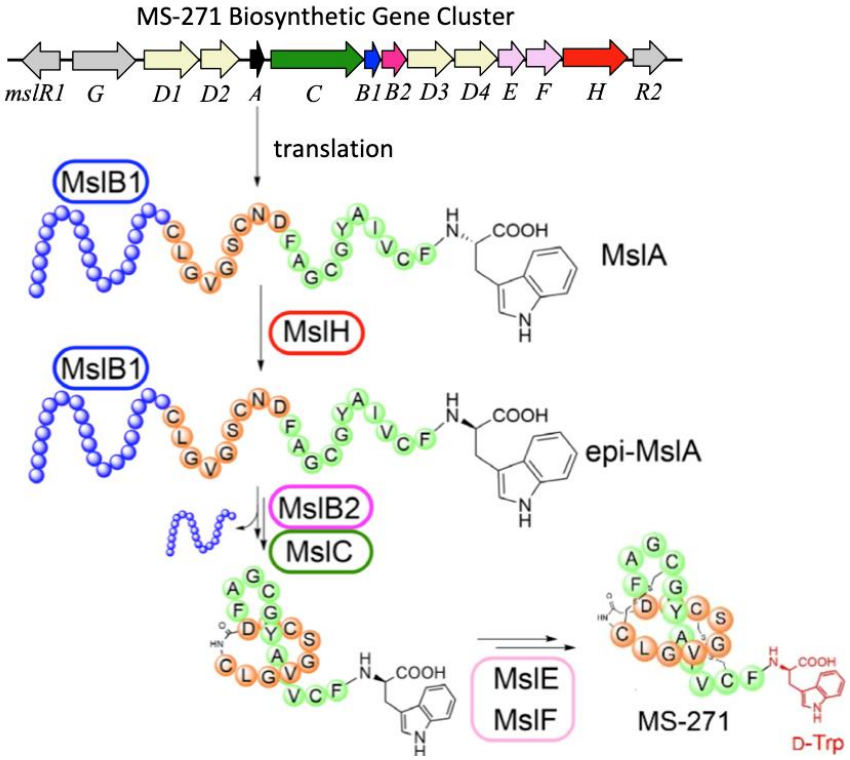
MS-271, originally isolated from *Streptomyces* sp. M-271 as a potent inhibitor of calmodulin-activated myosin light-chain kinase, is a lasso peptide natural product comprising 21 amino acids (aa) with a D-tryptophan (Trp) at its C-terminus. Lasso peptides are peptide natural products that have a characteristic isopeptide-bonded slipknot structure. In terms of their biosynthesis, lasso peptides belong to a group of ribosomally synthesized and post-translationally modified peptides (RiPPs). Thus, the biosynthesis of MS-271, especially the mechanism of D-Trp introduction, is of great interest. In this study, I investigated the biosynthesis of MS-271.

In chapter 2, I first identified MS-271 biosynthetic gene cluster (*msl*) spanning a ca. 11-kbp region from *mslR1* to *mslH* by draft genome sequencing followed by searching for DNA region encoding the amino acid sequence of MS-271. Sequence analysis revealed that precursor peptide gene (*mslA*) encoded 42-residue peptide with a leader peptide at its N-terminus, and most importantly, the C-terminal core region contained all 21 amino acid residues of MS-271 including the C-terminal Trp. This suggested that the D-Trp residue is introduced via epimerization into a ribosomal peptide as a post-translational modification. The cluster also contained genes for modification enzymes such as a macrolactam synthetase (*mslC*), precursor peptide recognition element (*mslB1*), cysteine protease (*mslB2*), disulfide oxidoreductases (*mslE*, *mslF*). Although obvious epimerase genes were absent in the cluster, the cluster encoded a protein of unknown function (*mslH*). Hence, I next carried out heterologous expression of the *msl* cluster in *Streptomyces lividans*. As the results, the production of MS-271 was confirmed by LC-MS and chiral amino acid analysis, indicating that the cluster contains all the necessary genes for MS-271 production including a novel peptide epimerase gene. I also showed that MslB1, B2, C and H were indispensable for MS-271 production by gene knockout experiments. Overall, these results suggested that MslH is responsible for the epimerization of the C-terminal Trp.

In chapter 3, I characterized the function of MslH *in vivo* and *in vitro*. Considering that many modification enzymes involved in RiPP biosynthesis require leader peptides for their substrate recognition, I speculated that the epimerization occurs on the nascent full-length MslA in MS-271 biosynthesis. As expected, *in vivo* experiments revealed the formation of epi-MslA when *mslA* was expressed with *mslH* in *Escherichia coli*. Additional coexpression of precursor peptide recognition element (*mslB1*) enhanced

the formation of MslA. Furthermore, *in vitro* experiments revealed that MslH catalyzed epimerization of C-terminal Trp of MslA in metal- and cofactor-independent manner and that the leader peptide in MslA is indispensable for the substrate recognition by MslH. I also examine substrate specificity of MslH by heterologous expression of the *msl* cluster to produce MS-271 derivatives by altering the core peptide sequences of the *mslA* gene, and demonstrated that MslH exhibited broad substrate specificities toward the N-terminal region of core peptides while the C-terminal "CFW" sequence is important for substrate recognition. Overall, I fully characterized MslH as a novel peptide epimerase. This is the first example epimerase that catalyzes epimerization at the Cα center adjacent to a carboxylic acid in a cofactor-independent manner.

Taken together, I identified MslH, previously annotated as a hypothetical protein, as a novel epimerase involved in the post-translational epimerization of the C-terminal Trp residue of the precursor peptide MslA. I also demonstrated that MslH exhibited broad substrate specificity toward the N-terminal region of the core peptide, showing that MslH-type epimerases offer opportunities in peptide bioengineering.



Biosynthesis of MS-271.

TABLE OF CONTENTS

Chapter 1

General introduction

1.1. Introduction of ribosomally synthesized and post-translationally modified peptide natural products (RiPPs)

Over the past century, natural products have played a key role in promoting our understanding of biology and the development of medicine. Along with the four major groups, terpenoids, alkaloids, polyketides and non-ribosomal peptides, identified in the last century, recent development of genome sequencing technology has greatly accelerated the discovery of ribosomally synthesized and post-translationally modified peptides (RiPPs) as a novel major superfamily of natural products[1].

In terms of the biosynthesis, RiPPs, found in all three kingdoms of life, are first synthesized as a linear precursor peptide (20 – 110 residues) encoded by a structural gene in the biosynthetic gene cluster (BGC) (Figure 1-1-1). Precursor peptides generally contain a leader peptide and a core peptide. Some precursor peptides additionally contain a follower peptide. The leader- and follower-peptide are usually important for the recognition for many post-translational modification (PTM) enzymes, that modify core peptide by providing binding affinity and is cleaved by the protease before the formation of final products. PTM enzymes are generally encoded in the same gene cluster as the precursor peptide gene, rendering the originally linear core peptides into mature products. As shown in Table 1-1-1, over 40 sub-classes of RiPPs, determined by characteristic PTMs of primary modification enzymes, are currently characterized[2].

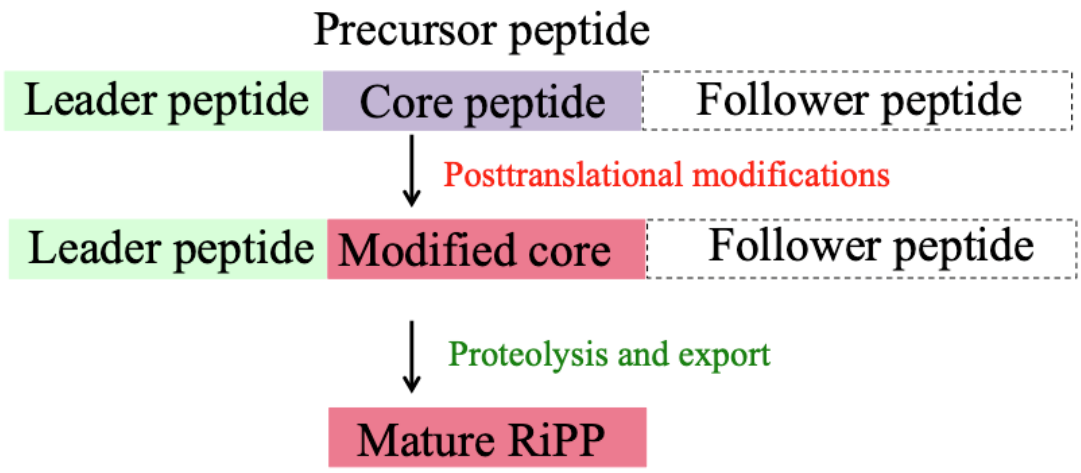


Figure 1-1-1. General biosynthetic pathway for RiPPs.

Table 1-1-1. Currently known RiPP classes and their representative natural products.

AEP, asparaginyl endoprotease; Dha/Dhb, dehydroalanine/dehydrobutyrine; DUF, domain of unknown function; FAS, fatty acid synthase; LP, leader peptide; MT, methyltransferase; PEARL, peptide aminoacyl-tRNA ligase; PKS, polyketide synthase; POP, prolyl oligopeptidase; rSAM, radical SAM.

Class	**Example**	**Class-defining PTM(s) or feature (enzyme responsible)**
Amatoxins	Phalloidin	N-to-C cyclization, Cys–Trp crosslink
Guanidinotides	Pheganomycin	α-Guanidino acid containing peptides (ATP-grasp)
Atropitides	Tryptorubin	Aromatic amino acids crosslinked to give a non-canonical atropisomer
Autoinducing peptides	AIP-I	Cyclic ester or thioester
Bacterial head-to-tail cyclized peptides	Enterocin AS-48	N-to-C cyclization (DUF95 & ATP-grasp)
Borosins	Omphalotin	Amide backbone N-methylation, N-to-C cyclization (POP)
Bottromycins	Bottromycin	Macrolactamidine (YcaO)

	A1	
Cittilins	Cittilin A	Biaryl and aryl–oxygen–aryl ether crosslinks (P450)
ComX	ComX168	Indole cyclization and prenylation
Conopeptides	Conantokin G	Peptides produced by cone snails
Crocagins	Crocagin A	Indole-backbone cyclization
Cyanobactins	Patellamides	N-terminal proteolysis (PatA protease)
Cyclotides	Kalata B1	N-to-C cyclization, disulfide(s) (AEP)
Dikaritins	Ustiloxin	Tyr-Xxx ether crosslink
Epipeptides	YydF	D-amino acids (rSAM)
Glycocins	Sublancin 168	S, O-glycosylation of Ser/Cys
Graspetides	Microviridin J	Macrolactones/lactams (ATP-grasp)
Lanthipeptides	Nisin	(Methyl)lanthionine, labionin
Lasso peptides	Microcin J25	Macrolactam with threaded C-terminal (Asn synthetase homolog)
Linaridins	Cypemycin	Dhb, no lanthionines
Linear azol(in)e-containing peptides	Microcin B17	Cys, Ser, or Thr derived azol(in)es (YcaO)
Lipolanthines	Microvionin	C-terminal labionin/avionin containing peptide and N-terminal FAS/ PKS segment
Lyciumins	Lyciumin A	Pyroglutamate, Trp–Gly crosslink
Streptide	Streptide	Trp–Lys crosslink (rSAM)
Methanobactins	Methanobactin	Oxazolones
Microcin C	Microcin C	Aminoacyl adenylate or cytidylate with a phosphoramidate linkage (ubiquitin E1

		homolog)
Mycofactocin	Mycofactocin	Val–Tyr crosslink (rSAM)
Orbitides	Cyclolinopeptide A	N-to-C cyclization; no disulfides
Pantocins	Pantocin A	Glu–Glu crosslink (PaaA)
Pearlins	Thiaglutamate	aa-tRNA derived (PEARL)
Proteusins	Polytheonamide	Nitrile hydratase LP
Pyrroloquinoline quinones	PQQ	Glu–Tyr crosslink (rSAM)
Ranthipeptides	Freyrasin	Sulfur-to-non-Cα thioether crosslink (rSAM)
Rotapeptides	TQQ	Oxygen-to-α-carbon crosslink
Ryptides	RRR	Arg–Tyr crosslink (rSAM)
Sactipeptides	Subtilosin	Sactionine crosslink (rSAM)
Spliceotides	PlpA	β-Amino acids (rSAM)
Sulfatyrotides	RaxX	Tyrosine sulfation
Thioamitides	Thioviridamide	Backbone thioamide (YcaO)
Thiopeptides	Thiostrepton	[4 + 2] cycloaddition of two Dha
Thyroid hormones	Triiodothyronine	Iodination of Tyr, excised from thyroglobulin

In addition to the primary PTMs that define each subclass, many compounds undergo compound-specific secondary (also referred to as “ancillary” or “tailing”) PTMs. These secondary PTM reactions usually confer higher affinity and selectivity on the corresponding RiPPs to their targets through additional molecular interactions[3]. Furthermore, the stabilities of RiPPs can be improved when the secondary PTM reactions take place at their N- and/or C- termini. Secondary PTM reactions reported in a recent decade include structural rearrangement, methylation, arginine conversions to citrulline and ornithine, phosphorylation, hydroxylation, halogenation, prenylation, N-terminal acylation, glycosylation, D-amino acid introduction, etc.

1.2 Lasso peptide biosynthesis

Bacterial lasso peptides belong to a structurally unique class of RiPPs, characterized by their lariat knot-like three-dimensional structure (Figure 1-2-1) that gives rise to their name. The lasso scaffold is composed of a 7-9 amino acid (aa) macrocyclic ring via an isopeptide bond formed between the N-terminal amino group and the carboxylic acid side chain of a Glu or Asp and C-terminal peptide tail which is threaded through the ring[4]. The tail is locked in the ring by bulky side-chain residues within itself and/or by disulfide linkages formed between cysteine residues. This unusual structural feature confers exceptional stability against thermal and protease treatment on many lasso peptides. After the first discovery of a lasso peptide anantin in 1991, diverse biological activities of lasso peptides were identified by bioactivity-guided screening, including antibacterial, anticancer, antiviral, and enzyme inhibitory activities. Since 2008, the advent of genome mining approach has accelerated the discovery speed of lasso peptides[5]. To date, over 70 members were isolated.

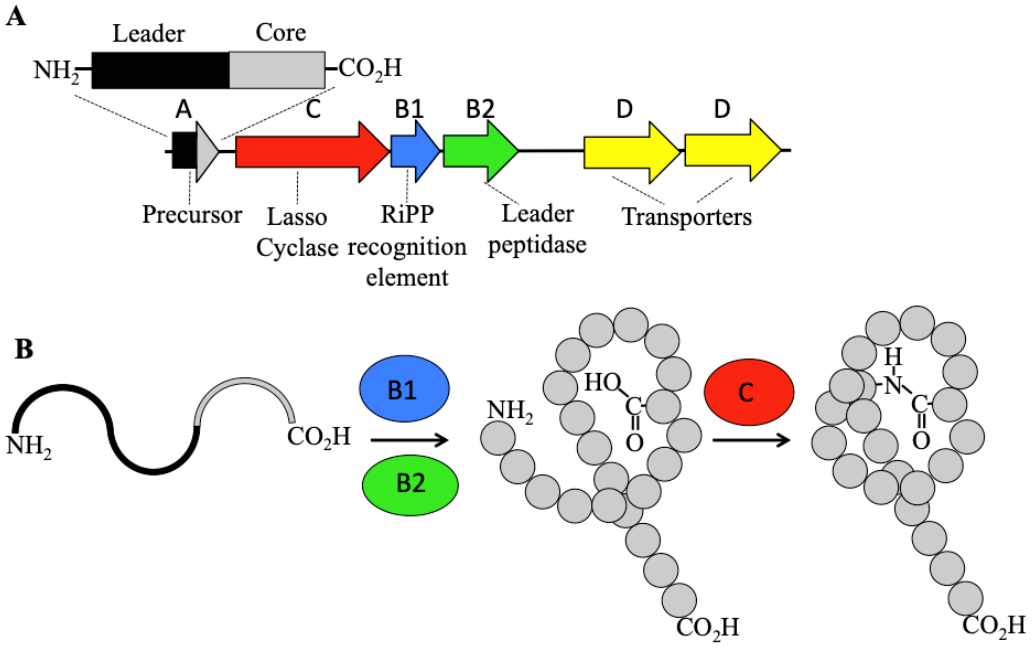


Figure 1-2-1. Biosynthesis of lasso peptides.

(A) A general biosynthetic gene cluster. (B) Process of maturation.

Because lasso peptides belong to RiPPs, the biosynthesis of lasso peptides starts with the translation of a ribosomally synthesized precursor peptide (A protein) which consists of an N-terminal leader region and a C-terminal core region (Figure 1-2-1 B). A widely conserved motif (YxxPxLx3Gx5Tx) within leader peptide is recognized and bound by the precursor peptide recognition element (RRE, B1 protein) which recruits the biosynthetic enzymes for post-translational modifications into lasso scaffold via a two-step procedure, including a leader peptidase, a homolog of transglutaminases (B2 protein), and a lasso cyclase (macrolactam synthetase), a homolog of asparagine synthetases (C protein). Upon the binding of leader region by B1 protein, the B2 protein site-specifically cleaves the leader peptide, liberating the N-terminus of core peptide. The C protein subsequently catalyzes the ATP-dependent isopeptide bond formation via adenylation of a Glu or Asp side chain prior to nucleophilic attack by the N-terminal amine to the carbonyl carbon which finishes the formation of lasso topology. Bioinformatic investigation revealed that B2 proteins in approximately one-third of predicted lasso peptide BGCs were conjugated to the C-terminus of B1 protein resulting in the formation of a fusion protein (B protein)[6]. It was reported that the B protein is presumed to fold the precursor peptide into a pre-lasso conformation before the ring formation[7]. Disulfide bonds are sometimes found in some lasso peptides formed by additional disulfide oxidoreductases encoded in their gene clusters. The physicochemical properties and bioactivities render lasso peptides a promising candidate for peptide drug design[8-10]. Furthermore, the rigid structure can serve as a scaffold for epitope grafting which endows lasso peptides with a novel activity[11, 12].

1.3. RiPPs containing D-amino acid residues

In general, the presence of D-amino acids is commonly the hallmark of peptides

biosynthesized via non-ribosomal peptide synthetases (NRPSs). In contrast, RiPPs generally consist of only L-amino acid due to the ribosomal origin. Although post-translational and site-specific epimerization of L-amino acids into their D-counterpart is chemically challenging, a few interesting cases of D-amino acid-containing RiPPs have recently been reported. The introduction of D-amino acids in RiPPs likely increase the structural diversity of RiPP natural products and endow peptide with various bioactivities as well as improved stability against natural proteases.

One such example is a group of polytheonamides, extremely potent toxins, from the uncultivated marine sponge *Theonella swinhoei*, (Figure 1-3-1 A). Considering their structures, the biosynthesis requires multiple D-amino acid introductions, *C*-methylations, and *N*-methylations that induce the formation of a membrane-inserting β-helix pore and confer the bioactivities. Biosynthetic studies on polytheonamides showed that the introduction of 18 D-amino acids at alternating positions of the core peptide was catalyzed by only one rSAM enzyme PoyD encoded in the gene cluster[13]. By further mutagenesis and labeling experiments, the mechanism of radical SAM peptide epimerase PoyD was proposed in Figure 1-3-1 B[14]. After reducing [4Fe-4S] cluster, PoyD generates a 5′-deoxyadenosyl radical (5′-dA$^{\bullet}$) and abstracts a Cα hydrogen atom of an amino acid residue. A carbon-centered radical is formed and quenched by the thiolate H atom of Cys-372 leading to the formation of a D-amino acid residue. Reduction of the thiyl radical is likely assisted by other amino acid residues from PoyD in similar manner to ribonucleotide reductase or spore photoproduct lyase for the next catalytic cycle. Further studies showed that genes encoding PoyD orthologs were widely distributed in cyanobacteria, rhizobia, and other bacteria and three proteins (AvpD, PlpD, and OspD) from cyanobacteria were characterizaed[15].

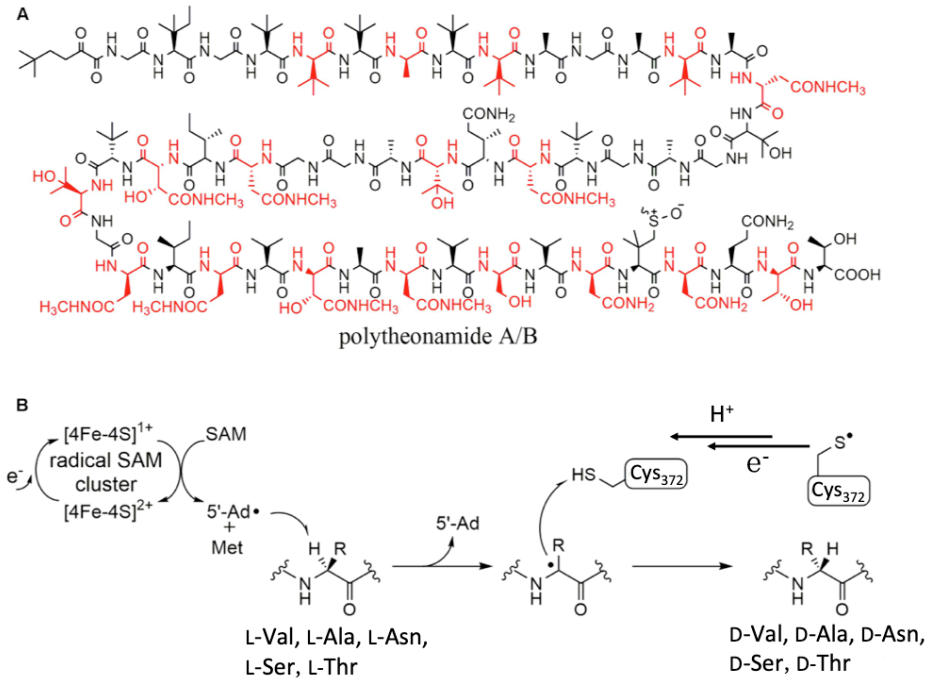


Figure 1-3-1 (A) structures of polytheonamides (A/B only differing in sulfoxide configuration). D-amino acids are shown in red. (B) proposed mechanism for the radical SAM peptide epimerase PoyD.

Epipeptides ($YydF_{33\text{-}49DD}$) were produced by *Bacillus subtills*[16]. The precursor peptide YydF is expressed as a 49-aa peptide and a D-Val at position 36 and a D-*allo*-Ile at 44 were later introduced by epimerization. Protease encoded in the cluster then cleaved the N-terminal 32-aa-leader peptide to form the final product $YydF_{33\text{-}49DD}$ without further modification. Introduction of these two D-amino acid residues were showed to be essential for the activity that induces the expression of a two-component system LiaRS, a major component of the bacterial cell envelope stress-response system. By *in vitro* and labeling experiments, a radical SAM enzyme YydG was characterized to be responsible for above two epimerizations. Its mechanism was shown in Figure 1-3-2. YydG generates a 5′-dA• radical that abstracts Cα hydrogen atom. A carbon-centred radical is generated and quenched by the thiolate hydrogen atom of Cys223, which leads to the formation of a D-amino acid residue. A second [4Fe-4S] cluster

found in SPASM domin probably assists the quenching of the thiyl radical to regenerate the Cys223 for the next catalytic cycle. Radical SAM epimerase YydG and PoyD use a similar mechanism although they belong to two distinct familys by phylogenetic analysis. A basic local alignment search tool (BLAST) search revealed over 200 YydG homologs were present in Gram-positive bacteria mainly related to the human microbiota.

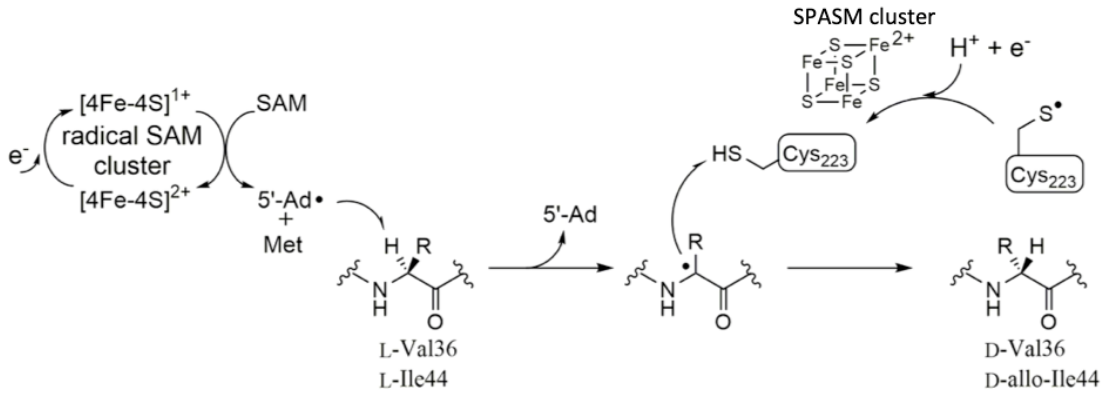


Figure 1-3-2 Proposed mechanism of YydG epimerization.

Bottromycin is a macrocyclic peptide first isolated from the terrestrial bacterium *Streptomyces bottropensis* and exhibits antibacterial activity against Gram-positive bacteria including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE) by binding to the A site of the ribosome and preventing the binding of aminoacyl-*t*RNA. Its structure contains a macrocyclic amidine and thiazole. Intersetingly, a D-Asp is installed at position 7 in bottromycin. An atypical α/β-hydrolase (ABH) fold enzyme BotH, lacking the conserved catalytic residues, was recently found and shown to be responsible for this epimerization (Figure. 1-3-3)[17]. Structral analysis revealed that two residues Val41 and Phe110 of BotH form hydrogen bonds with the thiazoline carboxy group. Furthermore, a water molecule is

trapped within hydrogen-bonding distance of the thiazoline's nitrogen and carboxy groups. The side chain of Asp7 serves as a base to abstract the Cα proton from itself, which triggers enamine formation and leads to proton transfer from the ordered water molecule to the thiazoline nitrogen. The resulting hydroxide ion abstracts the proton and triggers reprotonation of the enamine by abstracting the proton from the side chain of Asp7, resulting the epimerization of Asp7. Spontaneous epimerization of the Asp was also observed. Because D7N mutant peptide, which doesn't contain a free carboxyl group at position 7, was not a substrate of BotH, the BotH epimerization was suggested as an example of substrate-assisted catalysis.

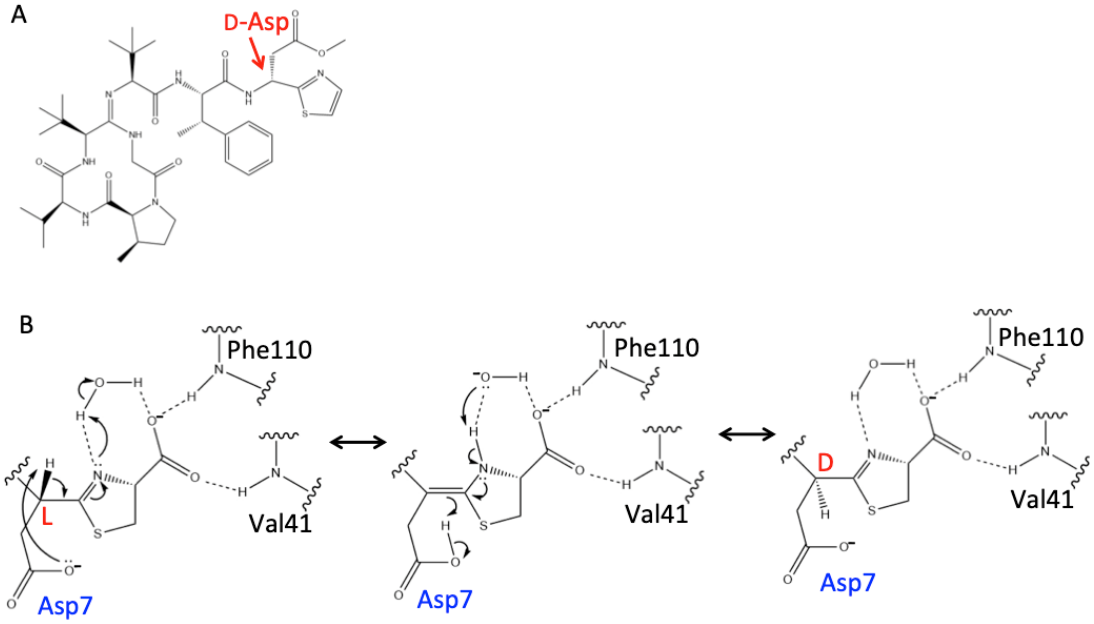


Figure 1-3-3. (A) D-Asp at position 7 in bottromycin A2
(B) Proposed mechanism for the substrate-assisted epimerization in the maturation of bottromycin. BotH residues are in black and substrate Asp in blue.

Lanthipeptides show a wide range of activities ranging including antibacterial，antifungal, antiviral, antinociceptive and actiallodynic activities. Their structures are characterized by the presence of lanthionine and 3-methyllanthionine linkages. Different from direct L-to-D epimerization as mentioned above， D-Ala can be installed

via an indirect route during the maturation of some lanthipeptides. In this route, the zinc-dependent LanJ dehydrogenases catalyze the reduction of dehydroalanine (Dha) into D-Ala in the core peptide (Figure. 1-3-4)[18].

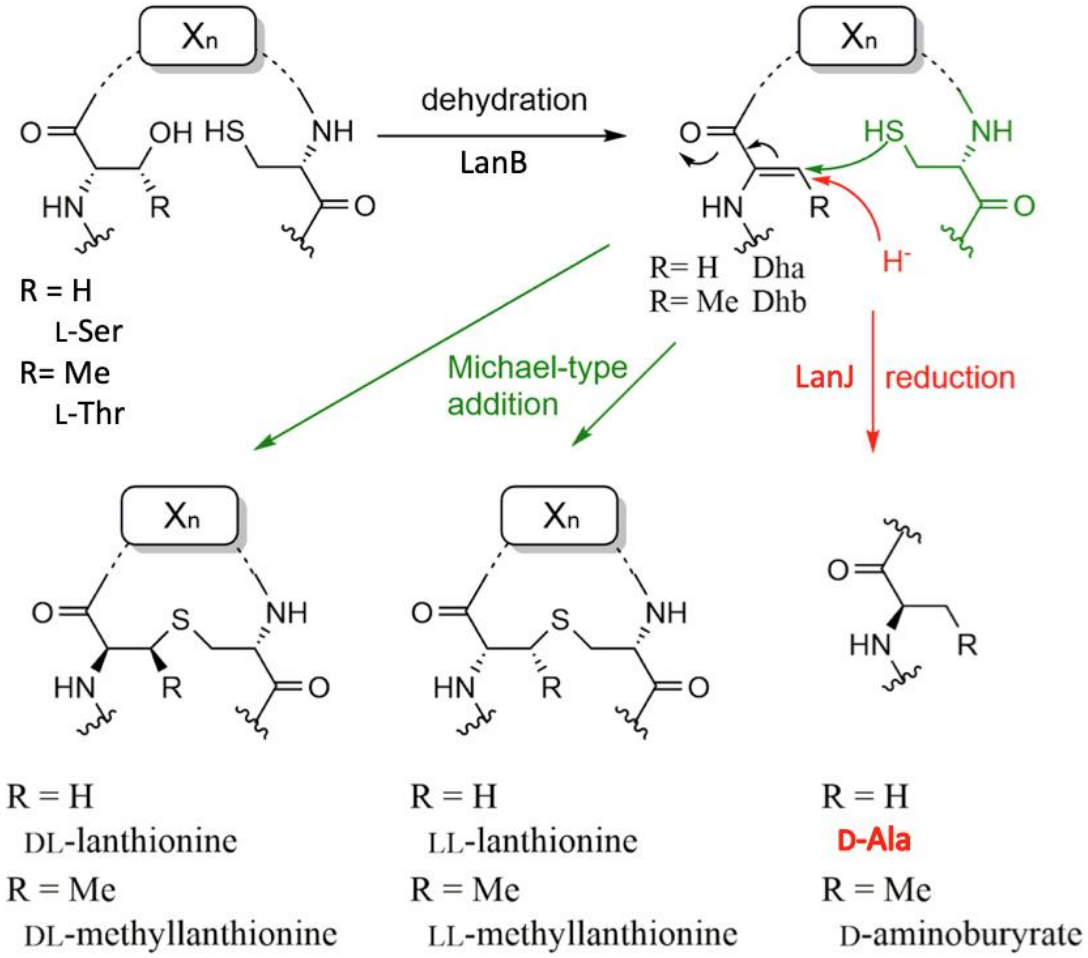


Figure 1-3-4. Formation of D-Ala in the maturation of lanthipeptides.

These are only mechanisms to introduce D-aa into RiPPs natural products reported to date. Besides above examples, there are two more examples whose mechanisms of the D-aa introduction remained unknown. The linaridin subfamily salinipeptins were isolated and identified recently (Figure 1-3-5), whose structures were highly modified[19]. These PTMs included multiple D-aa, *N*,*N*-dimethylalanine-*N*-oxide, dimethylimidazolidin-4-one moiety, C-terminal aminovinyl-cysteine. Although

biosynthetic genes for most PTMs were found in the identified gene cluster, the epimerase gene responsible for D-aa introduction in these peptides remains unknown.

Figure 1-3-5. Structures of the salinipeptins and D-amino acids in red.

A lasso peptide MS-271 was isolated from *Streptomyces* sp. M-271 as a potent inhibitor of calmodulin-activated myosin light chain kinase in 1996[20]. MS-271 also showed antibacterial activity against *Bacillus subtilis*, *Enterococcus faecium* and *Staphulococcus aureus*. Its three-dimensional structure was established by NMR analysis (Figure 1-3-6)[21]. The lasso scaffold of MS-271 is further modified by two disulfide linkages between Cys1 and Cys13, and Cys7 and Cys19. Intriguingly, amino acid analysis revealed that the C-terminal tryptophan (Trp) had a D-configuration while the other 20 amino acids were all L-type. Two possibilities can be considered to explain the biosynthetic origin of C-terminal D-Trp in MS-271 structure. The first one is that a D-Trp is ligated directly to the C-terminus of the 20-aa ribosomal peptide. The second one is that an epimerization of the C-terminal trptophan in the 21-aa peptide takes place.

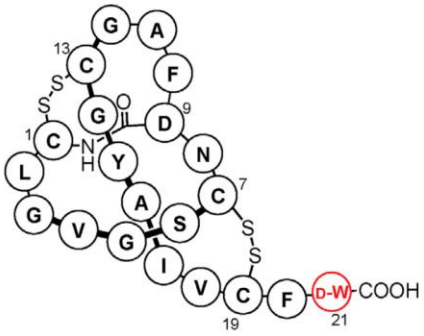


Figure 1-3-6. Structure of MS-271 and a C-terminal D-Trp in red.

In this study, I characterized a novel type of epimerase for D-Trp introduction in the lasso peptide MS-271 biosynthesis. This is the first epimerase that catalyzes epimerization at the Cα center adjacent to a carboxylic acid in a metal- and cofactor-independent manner in the biosynthesis of RiPPs natural product. D-Aa in peptides is generated by chemical synthesis to confer desirable properties in the pharmaceutical industry, such as the improvement of stability to proteases, change of tertiary structure, toxicity elimination of antibacterial peptides toward mammalian cells and alternation of biological activity. This novel enzyme chemistry found in this study, along with further understanding of its catalytic mechanism, will contribute to generating bioactive peptides with further desirable functions by the introduction of D-aa in an environmentally friendly way.

1.4. Other identified peptide epimerization machineries in microorganisms

Besides the D-aa introduction mechanisms into peptide involved in the RiPP biosynthetic pathways, three other types of enzymes have been identified in microorganisms[22]. Two are involved in the primary metabolism and one is in the secondary metabolism.

The first example in bacterial primary metabolisms is enzymes of the enolase superfamily, found in *Escherichia coli* and *Bacillus subtilis*, that catalyze the Mg^{2+}-dependent interconversion between L-Ala-D-Glu and L-Ala-L-Glu. L-Ala-L-Glu is a product of peptidoglycan degradation and can be decomposed by proteases (Figure 1-4-1)[23,24].

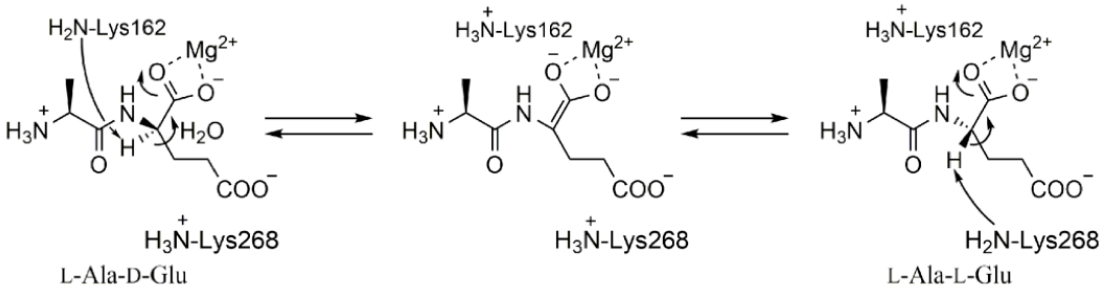


Figure 1-4-1. Reaction of L-Ala-D/L-Glu epimerases.
A two-base reaction mechanism typical for enolase superfamily enzymes was postulated by X-ray crystallography and mutagenesis studies.

The second example is identified by our laboratory in an alternative peptidoglycan biosynthetic pathway operating in *Xanthomonas oryzae* and several other microorganisms[25]. After MurD2 ligates L-Glu to UDP-MurNAc-L-Ala, the peptide epimerase MurL catalyzes the epimerization of the terminal L-Glu of the product (Figure 1-4-2). This unidirectional epimerization is ATP and Mg^{2+} dependent.

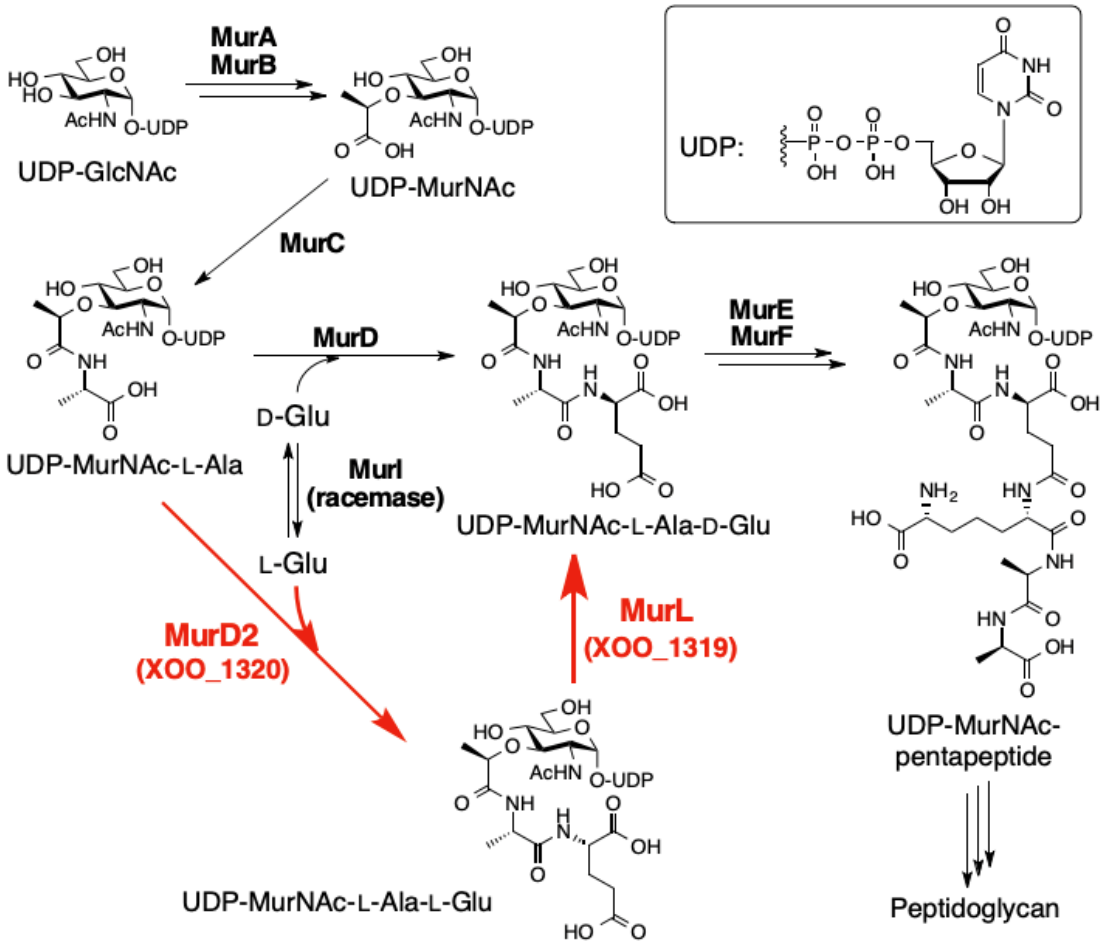


Figure 1-4-2. Peptidoglycan biosynthetic pathways of MurD/MurI and MurD2/MurL.

The third example was non-ribosomal peptide synthetases (NRPSs). The presence of D-amino acids in peptide natural products is generally the hallmark of peptides biosynthesized via NRPSs. Detailed studies using dissected modules of gramicidin synthetase and tyrocidine synthetase revealed that the epimerase (E) domains embedded in the enzymes catalyze the peptide epimerization in a metal- and cofactor-independent way (Figure 1-4-3)[26, 27]. X-ray structural analysis of E domain suggested that a Glu residue, acting as a catalytic base, abstracts a proton from the Cα position of the PCP-bounded Phe. The resulting enolate anion intermediate is electrostatically stabilized by a conserved His residue, a main-chain amide bond of a close Asp residue and positive pole of a α-helix. Proton abstraction form the conserved His residue located on the opposite side was presumed to reprotonate the Cα position to finish the epimerization reaction.

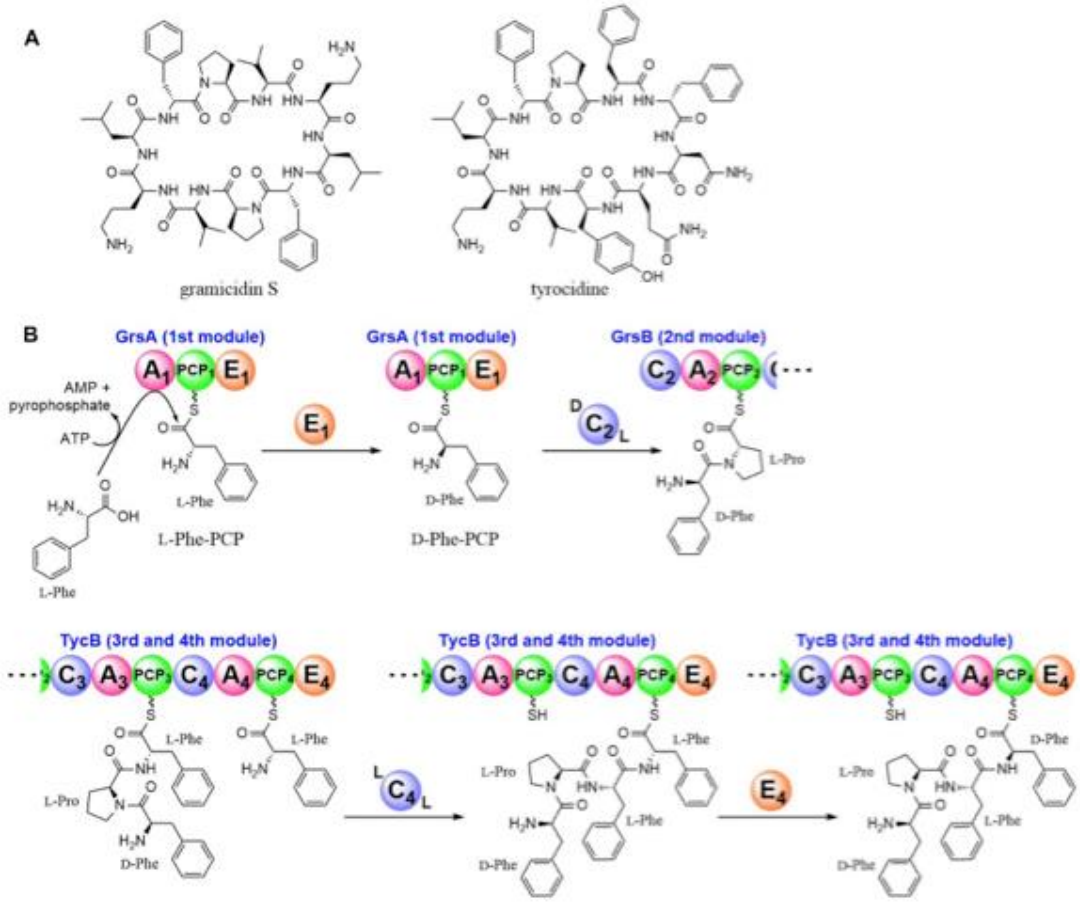


Figure 1-4-3. Epimerization by NRPS.
(A) Structures of gramicidin S and tyrocidine. (B) Epimerization reactions by NRPS E domains.

To the best of my knowledge, these machineries (except D-Ala in lathipeptides introduced by a sequential dehydration and reduction) introduced above are the only examples of peptide epimerization in microorganisms.

References

1. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.*, **30**, 108-160.
2. Montalban-lopez, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A. J.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Grande Burgos, M. J.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martinez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H, G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Sussmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J. K.; van Wezel, G. P.; Zhang, Q.; Zhong, Jin.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. (2020). New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.*, DOI: 10.1039/d0np00027b.
3. Funk, M. A.; van der Donk, W. A. (2017). Ribosomal narural products, tailed to fit. Acc. Chem. Res., **50**, 1577-1586.
4. Hegemann, J. D., Zimmermann, M., Xie, X., Marahiel, M. A. (2015). Lasso peptides: an intriguing class of bacterial natural products. *Acc. Chem. Res.* **48**, 1909.
5. Maksimov, M. O., Link, A. J. (2014). Prospecting genomes for lasso peptides. *J. Ind. Microbiol. Biotechnol.* **41**, 333.
6. Tietz, J., Schwalen, C., Patel, P. S., Maxson, T., Blair, P. M., Tai, H. C., Zakai, U. I., Mitchell, D. A. (2017). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **13**, 470.
7. Ortega M. A.; van der Donk W. A. (2016). New insights into the biosynthetic logic

of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**, 31.

8. Lopez, F. E., Vincent, P. A., Zenoff, A. M., Salomon, R. A., Farías, R. N. (2007). Efficacy of microcin J25 in biomatrices and in a mouse model of salmonella infection. *J. Antimicrob. Chemother*. **59**, 676.
9. Yamamoto, T., Matsui, H., Yamaji, K., Takahashi, T., Overby, A., Nakamura, M., Matsumoto, A., Nonaka, K., Sunazuka, T., Omura, S., et al. (2016). Narrow-spectrum inhibitors targeting an alternative menaquinone biosynthetic pathway of Helicobacter pylori. *J. Infect. Chemother*. **22**, 587.
10. Naimi, S., Zirah, S., Hammami, R., Fernandez, B., Rebuffat, S., Fliss, I. (2018). Fate and biological activity of the antimicrobial lasso peptide Microcin J25 under gastrointestinal tract conditions. *Front. Microbiol.* **9**, 1764.
11. Knappe, T. A., Manzenrieder, F., Mas-Moruno, C., Linne, U., Sasse, F., Kessler, H., Xie, X., Marahiel, M. A. (2011). Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. *Angew. Chem., Int. Ed.* **50**, 8714.
12. Hegemann, J. D., De Simone, M., Zimmermann, M., Knappe, T. A., Xie, X., Di Leva, F. S., Marinelli, L., Novellino, E., Zahler, S., Kessler, H., et al. (2014). Rational improvement of the affinity and selectivity of integrin binding of grafted lasso peptides. *J. Med. Chem.* **57**, 5829.
13. Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H. G.; Matsunaga, S.; Piel, J. (2012). Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science*, **338**, 387.
14. Parent, A.; Benjdia, A.; Guillot, A.; Kubiak, X.; Balty, C.; Lefranc, B.; Leprince, J.; Berteau, O. (2018). Mechanistic investigations of PoyD, a radical S-adenosyl-L-methionine enzyme catalyzing iterative and directional epimerizations in polytheonamide A biosynthesis *J. Am. Chem. Soc.*, **140**, 2469 -2477
15. Morinaka, B. I.; Vagstad, A. L.; Helf, M. J.; Gugger, M.; Kegler, C.; Freeman, M. F.; Piel, J. (2014). Radical S-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. *Angewandte Chemie International Edition*, **53**, 8503-8507.

16. Benjdia, A.; Guillot, A.; Ruffie, P.； Leprince, J.； Berteau, O. (2017). Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in Bacillus subtilis. *Nat. Chem.*, **9**, 698-707
17. Sikandar, A.; Franz, L.; Adam, S.; Santos-Aberturas, J.; Horbal, L.; Luzhetskyy, A.; Truman, A. W.; Kalinina, O. V.; Koehnke, J. (2020). The bottromycin epimerase BotH defines a group of atypical α/β-hydrolase-fold enzyme. *Nat. Chem. Biol.*, **16**, 1013-1018.
18. Cotter, P. D.; O'Connor, P. M.; Draper, L. A.; Lawton, E. M.; Deegan, L. H.; Hill, C.; Ross, R. P. (2005). Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 18584-18589.
19. Shang, Z.; Winter, J. M.; Kauffman, C. A.; Yang, I.; Fenical, W. (2019). Salinipeptins: integrated genomic and chemical approaches reveal unusual D-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a great salt lake Streptomyces sp. *ACS Chem. Biol.* **14**, 415-425.
20. Yano, K.; Toki, S.; Nakanishi, S.; Ochiai, K.; Ando, K.; Yoshida, M.; Matsuda, Y.; Yamasaki, M. (1996). MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from streptomyces sp.-I. Isolation, structure determination and biological properties of MS-271. *Bioorg. Med. Chem*. **4**, 115-120.
21. Katahira, R.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. (1996). MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from streptomyces sp.-II. Solution structure of MS-271: characteristic features of the 'lasso' structure. *Bioorg. Med. Chem*. **4**, 121-129.
22. Ogasawara, Y.; Dairi, T. (2018). Peptide epimerization machineries found in microorganisms. *Frontiers in microbiology*, **9**, 156.
23. Schmidt, D. M.; Hubbard, B. K.; Gerlt, J. A. (2001). Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in Bacillus subtilis and Escherichia coli as L-Ala-D/L-Glu epimerases. *Biochemistry*, **40**, 15707-15715.
24. Gulick, A. M.; Schmidt, D. M.; Gerlt, J. A.; Rayment, I. (2001). Evolution of enzymatic activities in the enolase superfamily: crystal structures of the L-Ala-D/L-Glu epimerases from Escherichia coli and Bacillus subtilis. *Biochemistry*, **40**, 15716-15724.

25. Feng, R.; Satoh, Y.; Ogasawara, Y.; Yoshimura T.; Dairi, T. (2017). A glycopeptidyl-glutamata epimerase for bacterial peptidoglycan biosynthesis. *J. Am. Chem. Soc.*, **139**, 4243-4245.
26. Stachelhaus, T.; Walsh, C. T. (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry*, **39**, 5775-5787.
27. Luo, L.; Kohli, R. M.; Onishi, M.; Linne, U.; Marahiel, M. A.; Walsh, C. T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry*, **41**, 9184-9196.

Chapter 2

Identification of the biosynthetic gene cluster of a D-tryptophan-containing lasso peptide, MS-271

2.1. Introduction

As mentioned in chapter 1, the mechanism of D-Trp introduction in MS-271 structure is of great interest. The biosynthetic gene cluster can be identified based on genome sequencing and homology analysis to the general biosynthetic machinery of lasso peptides. Maturation of lasso peptides from the precursor peptide generally needs three modification enzymes RiPP recognition element (B1 protein), a leader peptidase (B2 protein) and a macrolactam synthetase (C protein). In addition to these minimal set of genes further ancillary modifications of lasso peptides have been reported. For example, methyltransferase (StspM) in lassomycin-like lasso peptides biosynthesis, kinase (ThcoK) in paeninodin biosynthesis, acetyltransferase (AlbT) in albusnodin biosynthesis and hydroxylase (CanE) in canucin A biosynthesis are found to be colocalized in their gene clusters, respectively[1-4].

In this chapter, I identified a putative MS-271 biosynthetic gene cluster (*msl*) and showed that the D-Trp in MS-271 was introduced by epimerization of the ribosomal peptide as a post-translational modification. I then carried out heterologous expression of it in *Streptomyces lividans* and suggested that the *msl* cluster contained all the necessary genes for MS-271 biosynthesis, including a peptide epimerase gene. Furthermore, I performed the gene knockout experiments and showed that MslB1, -B2, -C and -H were indispensable for MS-271 production.

2.2. Results

2.2.1 Identification of the putative MS-271 biosynthetic gene cluster and a gene knockout experiment of a function-unknown gene *mslH*

To gain insight into the introduction of C-terminal D-Trp, the biosynthesis of the lasso peptide MS-271 was investigated. As a first step, I carried out the draft genome sequencing of the producer *Streptomyces* sp. M-271. Due to its ribosomal origin, I next employed a tBlastn search (a search of translated nucleotide databases with a protein query) against the genome databse by using the amino acid sequences of MS-271 as the query (Figure 2-2-1-1. A). As a result, the precursor peptide gene *mslA* was identified (Figure 2-2-1-1. B). It encoded a 42-residue peptide with a leader peptide at its N-terminus and, most importantly, the C-terminal core region contained all 21 amino acids of MS-271, including the C-terminal tryptophan. This suggested that the D-Trp residue in the MS-271 is introduced by epimerization into a ribosomal peptide as a post-translational modification.

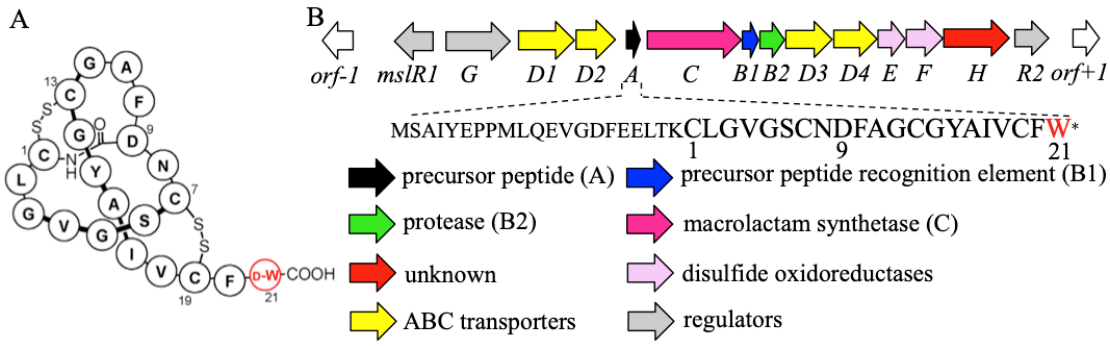


Figure 2-2-1-1 (A) Structure and (B) biosynthetic gene cluster (*msl*) of MS-271

Since genes for biosynthetic enzymes are located in the nearby region of the precursor peptide gene in lasso peptide BGCs, the flanking region of the precursor peptide gene *mslA* was investigated and MS-271 BGC (*msl*) was identified (Figure 2-

2-1-1. B). In the *msl* cluster, genes for modification enzymes such as a macrolactam synthetase (MslC), precursor peptide recognition element (MslB1), cysteine protease (MslB2), disulfide oxidoreductases (MslE， MslF), and a protein of unknown function (MslH) were found (Table 2-2-1). For the biosynthesis of MS-271, MslC, MslB1 and MslB2 would be required to generate the lasso topology. MslE and MslF are likely responsible for the disulfide bond formation. Besides these biosynthetic proteins, MslD1 to -D4 exhibited homology to ABC transporters, which constitute a large family of membrane-associated proteins involved in natural product export and resistance. The *msl* cluster also contained three homologues of known regulators of secondary metabolism. However, no obvious candidate gene for the epimerization of C-terminal tryptophan was found in the cluster. Instead, one gene with unknown function (*mslH*) existed and translational coupling of all genes between *mslC* and *mslH* suggested that they constitute an operon and that MslH is involved in the biosynthesis of MS-271.

BLAST search indicated that MslH exhibited homolog (~52%) to CapA family proteins which were involved in poly- γ -glutamic acid (PGA) biosynthesis. PGA is a natural biopolymer composed of D- and/or L- glutamic acid monomers linked by γ -amide bonds[5]. However, its biosynthetic study was hindered due to the instability of the membrane-bound biosynthetic enzyme complex.

Table 2-2-1. Proposed functions of ORFs in the *msl* cluster

ORF	Length [aa]	Proposed functions	Typical homology, protein accession (% identity/% similarity) [organism]
MslR1	230	LuxR type regulator	DNA-binding-response regulator, WP_067372316 (94/97), [*Streptomyces olivochromogenes*]
MslG	377	sensor histidine kinase	sensor histidine kinase, WP_062022673 (84/90) [*Streptomyces phaeopurpureus*]
MslD1	307	ABC transporter	ABC transporter ATP-binding protein, WP_059005878 (74/84) [*Streptomyces specialis*]
MslD2	250	ABC transporter	ABC transporter ATP-binding protein, WP_059005879 (70/82) [*S. specialis*]
MslA	42	precursor peptide	
MslC	613	lasso peptide isopeptide bond-forming cyclase	lasso peptide isopeptide bond-forming cyclase. WP_086700785 (77/84) [*Streptomyces vinaceus*]
MslB1	87	lasso peptide protease B1 protein	lasso peptide biosynthesis PqqD family chaperone, WP_059005881 (69/82) [*S. specialis*]
MslB2	156	lasso peptide protease B2 protein	lasso peptide biosynthesis B2 protein, WP_059005882 (87/90) [*S. specialis*]
MslD3	321	ABC transporter	ABC transporter ATP-binding protein, WP_059005883 (79/85) [*S. specialis*]
MslD4	289	ABC transporter	ABC transporter ATP-binding protein, WP_059005884 (82/88) [*S. specialis*]
MslE	181	disulfide bond formation	DoxX family membrane protein, WP_075027174 (90/91) [*Streptomyces mirabilis*]
MslF	246	disulfide bond formation	disulfide bond formation protein DsbA, WP_099941913 (94/96) [*Streptomyces* sp. 93]
MslH	440	unknown	CapA family protein WP_051857567 (52/64) [*Streptomyces cellulosae*]
MslR2	224	LuxR type regulator	DNA-binding response regulator, WP_073930679 (90/93) [*Streptomyces* sp. CB02400]

To test the involvement of the function-unknown *mslH* in the biosynthesis of MS-271, I carried out a gene knockout experiment of *mslH*. In-frame deletion of *mslH* was used to avoid the nonsense mutation that will terminate the translation of gene in the downstream region of *mslH*. The plasmid pWHM3-mslH-ko1 containing the upstream region and downstream region of *mslH* was constructed and introduced to the *Streptomyces* sp. M271 protoplast to construct the Δ*mslH* mutant (Figure 2-2-1-2). After the cultivation, 16 colonies were selected for colony PCR and three Δ*mslH* mutants were constructed successfully which were confirmed by agarose gel electrophoresis.

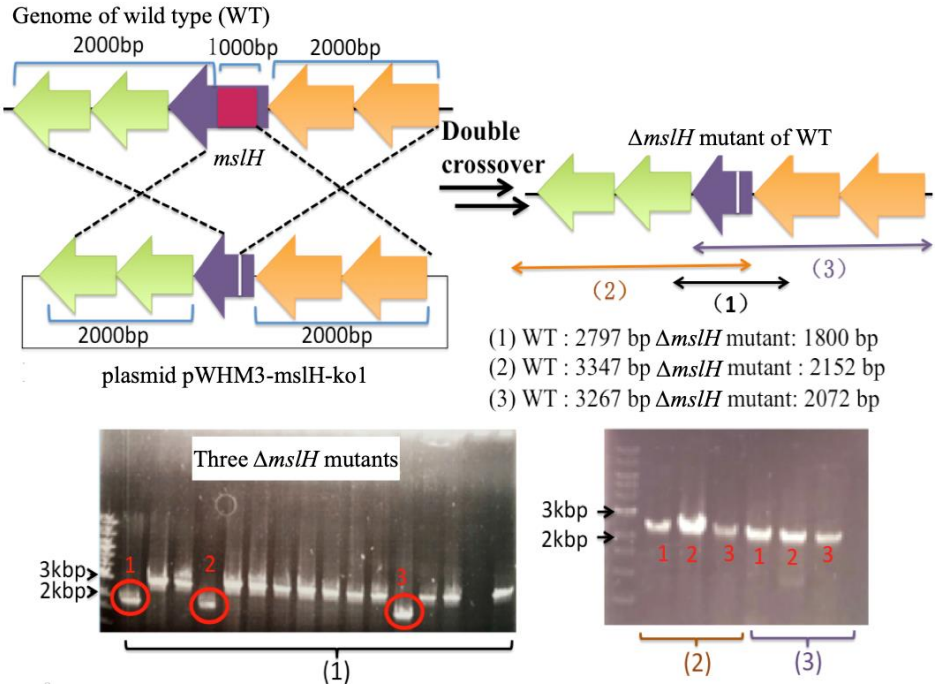


Figure 2-2-1-2. In-frame deletion of a function unknown gene *mslH*.

Then, I examined whether the Δ*mslH* mutants produced the final product MS-271. Three mutants were cultured along with wild-type *Streptomyces* sp. M-271 on agar production plates. The metabolites were extracted and analyzed by LC-MS. As shown in Figure 2-2-1-3, a specific peak with a retention time (RT) at 30.8 min was only detected in the sample of wild type. LC-MS analysis showed the compound in this peak

shared the same molecular weight as that of MS-271 (calculated mass: $[M+H]^+$: *m/z* = 2163.86; observed mass: $[M-H]^-$: *m/z* = 2161.02), suggesting that the compound is MS-271 (Figure 2-2-1-4). Although no intermediates were identified by LC-MS analysis, the abolishment of MS-271 in three Δ*mslH* mutants clearly demonstrated that the function-unknown *mslH* was essential for MS-271 biosynthesis.

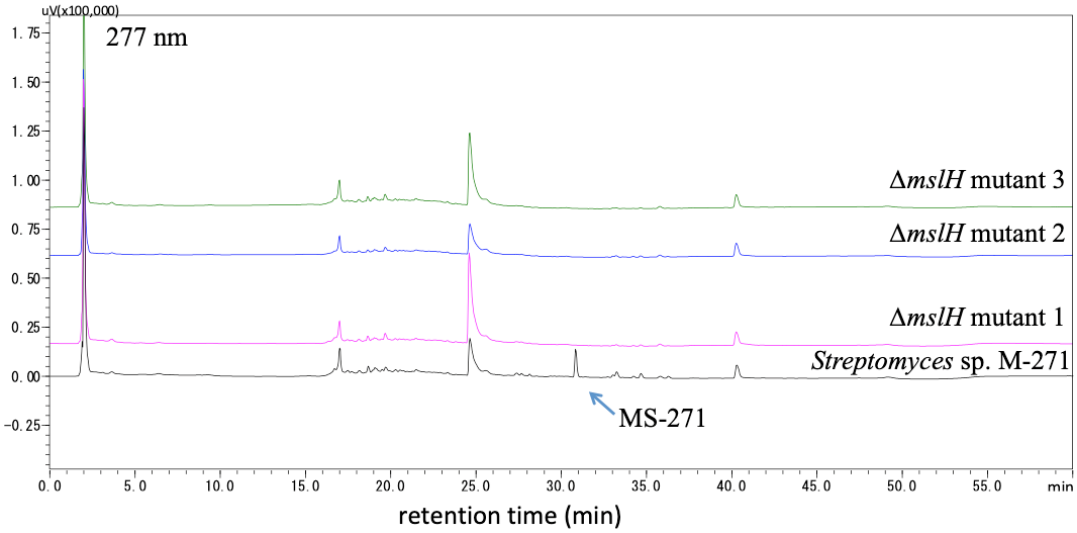


Figure 2-2-1-3. HPLC analysis of the extracts of three Δ*mslH* mutants and wild type.

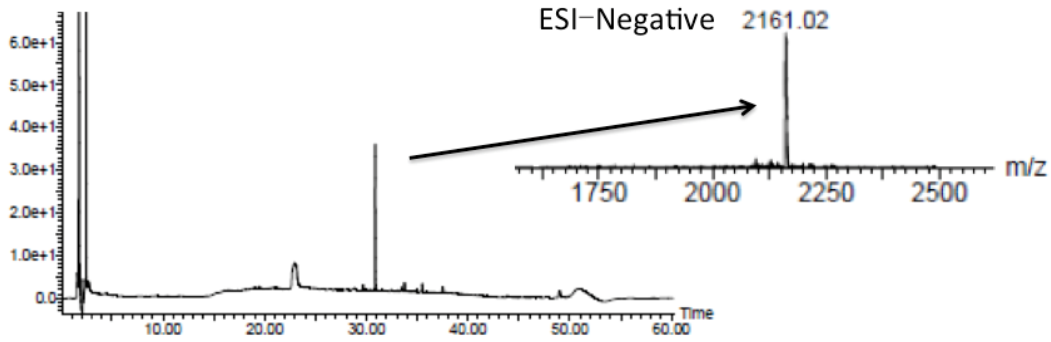


Figure 2-2-1-4. LC-MS analysis of compound in the peak with a RT at 30.8 min**.**

2.2.2 Heterologous expression of the putative MS-271 biosynthetic gene cluster in *Streptomyces lividans*

Although the *msl* cluster lacked a candidate gene for epimerization of the C-terminal tryptophan residue, I examined whether it contained all the necessary genes for MS-271 biosynthesis by heterologous expression, a useful tool for the identification of lasso peptide genes[6]. The *msl* cluster, spanning an approximately 11-kbp region from *mslR1* to *mslH*, was selected and cloned into the expression vector pWHM3 to generate the plasmid pWHM3-msl and expressed in the heterologous host *Streptomyces lividans* TK23.

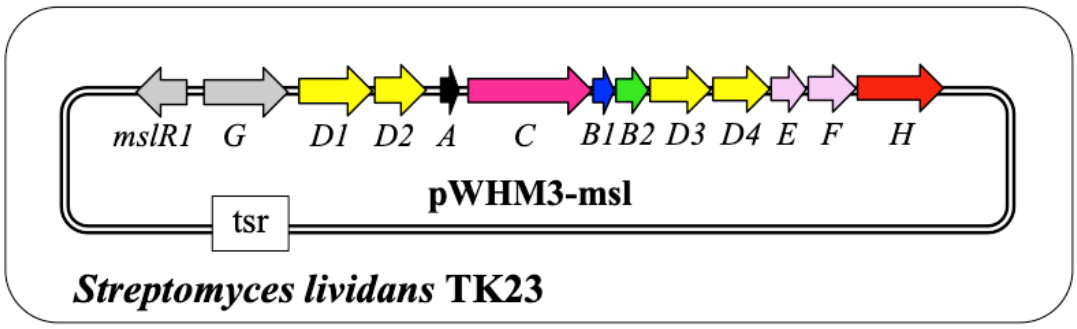


Figure 2-2-2-1 The plasmid containing the putative MS-271 cluster was expressed in *S. lividans*.

Then, I examined whether the constructed strain produced the final product MS-271. The constructed strain (*S. lividans*/pWHM3-msl) was cultured along with wild-type *Streptomyces* sp. M-271 on agar production plates. The metabolites were extracted and analyzed by LC-MS. As shown in Figure 2-2-2-2, a metabolite with the same retention time (at t_R = 23.8 min) as MS-271 was detected in the culture of the constructed strain (trace ii) in a manner similar to the wild-type strain (trace i). Its molecular weight was also same as that of MS-271 (Figure 2-2-2-2) (calculated mass: $[M+H]^+$: m/z = 2163.86; observed mass: $[M+H]^+$: m/z = 2163.59), and its MS/MS data

supported that it had the same tertiary structure as that of MS-271 (Figure 2-2-2-4). Furthermore, the C-terminal tryptophan in the metabolite was confirmed to be D configuration by chiral analysis (Figure 2-2-2-5). These results clearly showed that the product was indeed MS-271 and that the gene cluster contained all the genes responsible for the biosynthesis of MS-271, including a peptide epimerase gene.

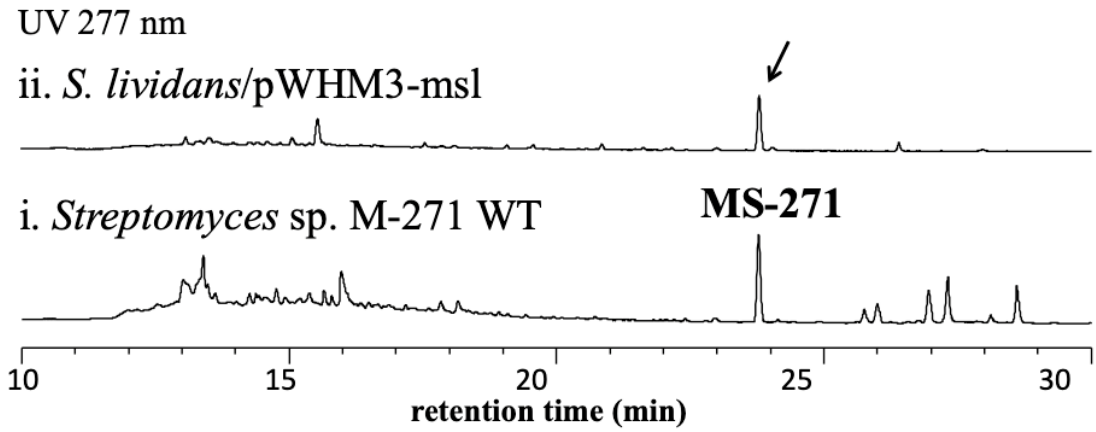


Figure 2-2-2-2 LC-MS analysis of products accumulated in the cultures. i) *Streptomyces* sp. M-271 wild type. ii) *S. lividans*/pWHM3-msl.

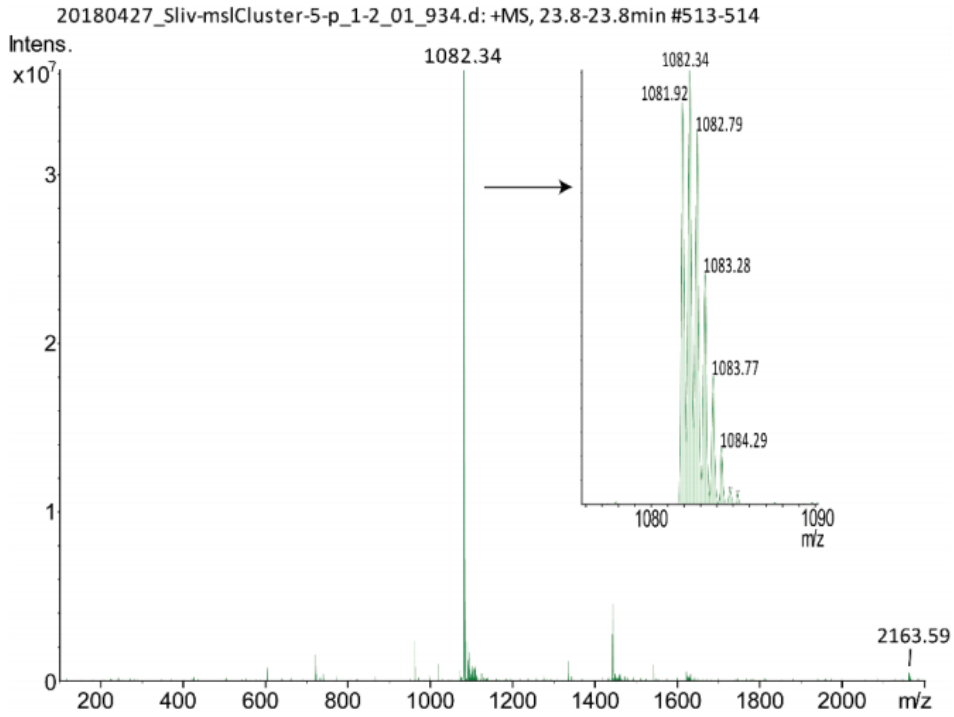


Figure 2-2-2-3 Mass spectrum of MS-271 produced by *S. lividans*/pWHM3-msl (retention time of 23.8 min).

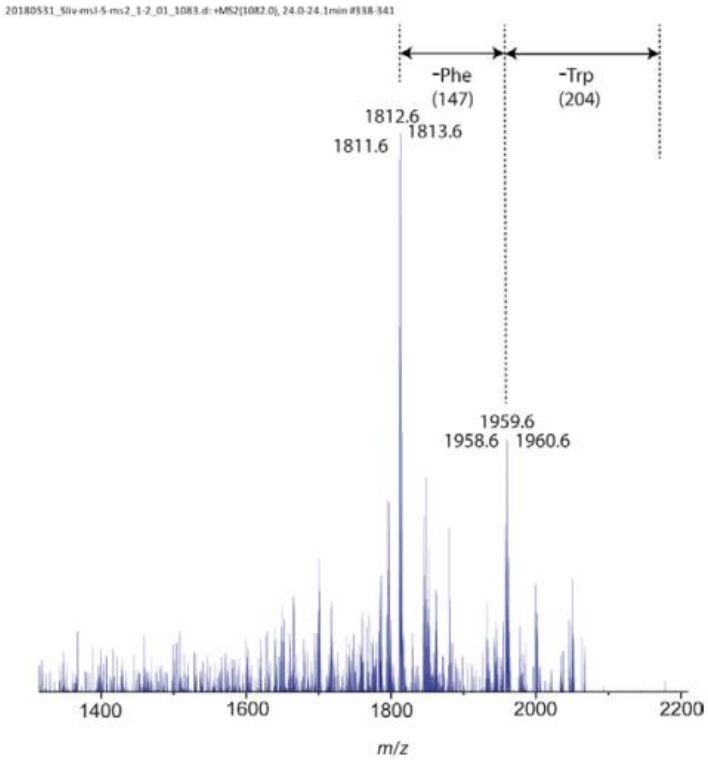


Figure 2-2-2-4 MS/MS spectrum of MS-271 produced by *S. lividans*/pWHM3-msl. A double-charged peak at *m/z* of 1082 (±2) was fragmented.

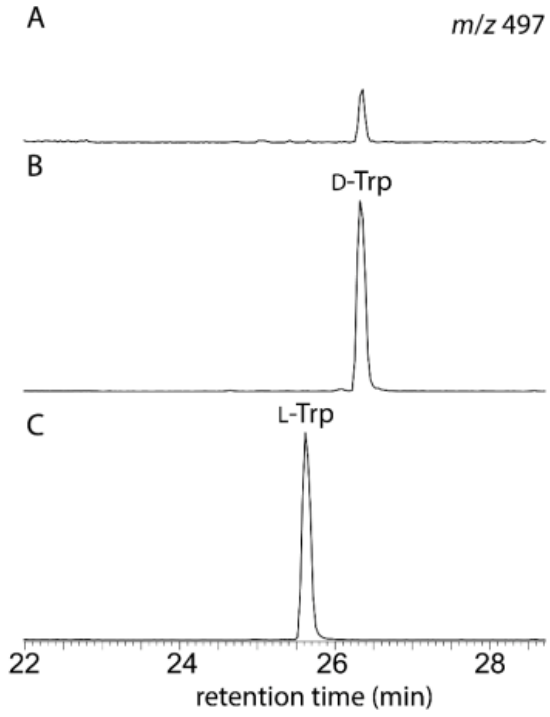


Figure 2-2-2-5 Chiral analysis of the tryptophan using L-FDLA. The LC-MS profile was monitored at m/z 497 ($[M-H]^-$ of FDLA derivatives of Trp) (A) MS-271 from *S. lividans*/pWHM-msl culture, (B) D-Trp standard, and (C) L-Trp standard.

2.2.3 Effects of regulator overexpressions on MS-271 productivity

Since the *msl* cluster contained three regulators including two putative pathway-specific LuxR-type activators, I next investigated whether these regulator genes were involved in the biosynthesis of MS-271. The *mslR1*, *mslG*, and *mslR2* genes were independently cloned into downstream of an *erm*EP* promoter of pTYM18ep[7] and then integrated into the *attB* site[8] on the genome of *S. lividans* TK23 to generate *S. lividans:mslR1*, *S. lividans:mslG* and *S. lividans:mslR2*, respectively (Figure 2-2-3-1). The resulting strains were transformed with pWHM3-msl, and their MS-271 productivities were analyzed by LC-MS.

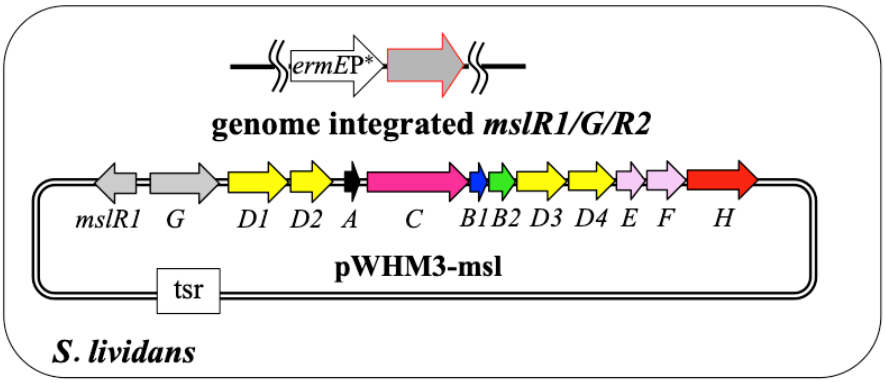


Figure 2-2-3-1 Co-expression of regulator with *msl* cluster.

As shown in Figure 2-2-3-2, the MS-271 yield was increased four-fold with *mslG* (trace iii) and 18-fold (2.1 mg per plate) (traces v) with *mslR2*, respectively. On the contrary, the production of MS-271 decreased with *mslR1* overexpression. These results demonstrated that *mslR2* was a pivotal positive regulator for MS-271 biosynthesis. I thus used *S. lividans:mslR2* as a heterologous host for the following knockout experiments.

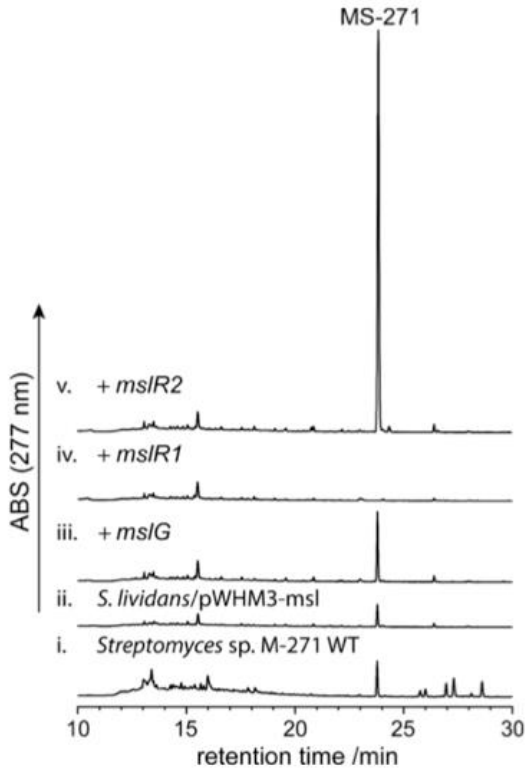


Figure 2-2-3-2 LC-MS analysis of products accumulated in the cultures. i) *Streptomyces* sp. M-271 wild type, ii) *S. lividans*/pWHM3-msl. iii) *S. lividans*:*mslG*/pWHM3-msl, iv) *S. lividans*:*mslR1*/pWHM3-msl, v) *S. lividans*:*mslR2*/pWHM3-msl.

2.2.4 Gene knockout experiments of each modification enzyme gene

To analyze the biosynthetic mechanism of peptide modification, I constructed pWHM3-msl-derived in-frame deletion plasmids lacking *mslB1*, *mslB2*, *mslC*, *mslE*, *mslF*, or *mslH* genes by Red/ET recombination. Each deletion construct was introduced into *S. lividans:mslR2* (Figure 2-2-4-1), and the resulting metabolite profiles were analyzed by LC-MS.

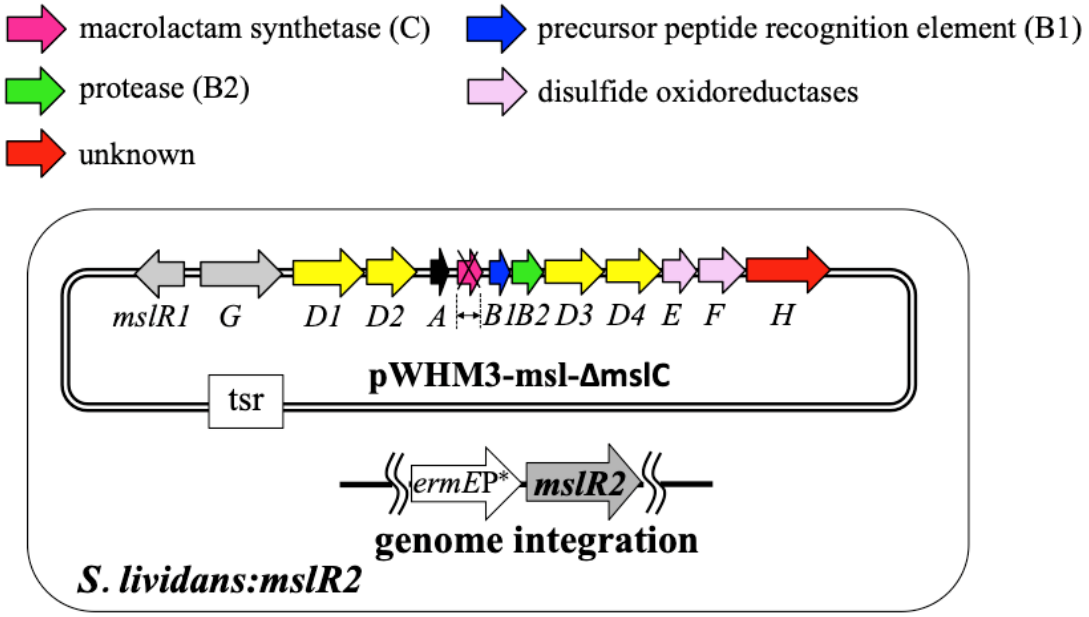


Figure 2-2-4-1 *S. lividans:mslR2* harboring an in-frame deletion plasmid lacking one of modification enzyme genes.

As shown in Figure 2-2-4-2, the mutants deleting *mslE* amd *mslF* showed production of MS-271, although their production levels were reduced (trace v, vi), suggesting that MslE and MslF may assist the formation of correct disulfide bonds in the biosynthesis of MS-271. On the contrary, deletion of *mslB1*, *mslB2*, *mslC*, or *mslH* completely abolished MS-271 production, clearly demonstrating that these genes are indispensable for MS-271 biosynthesis. Furthermore, no intermediate was accumulated in any of the deletion mutants. To date，no linear peptide intermediates have been found in any lasso peptide gene knockout experiments，suggesting that intermediates may be broken down by proteases.

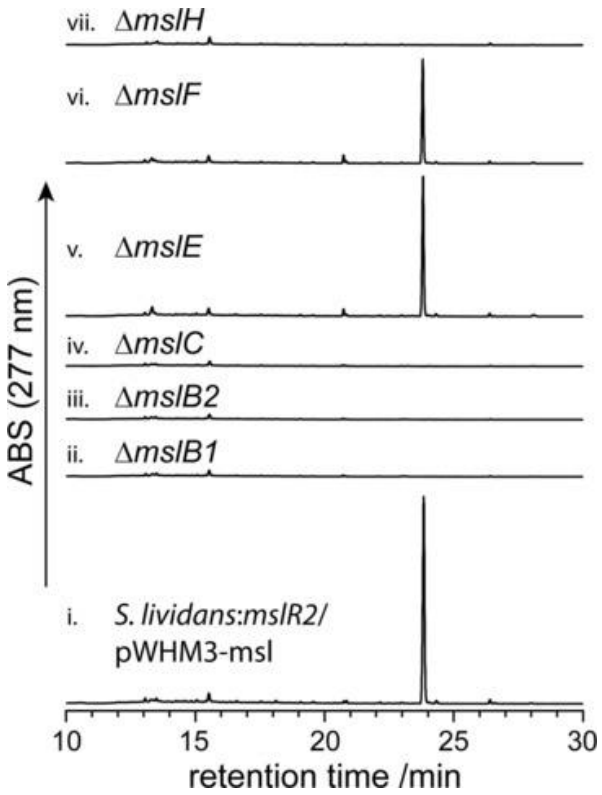
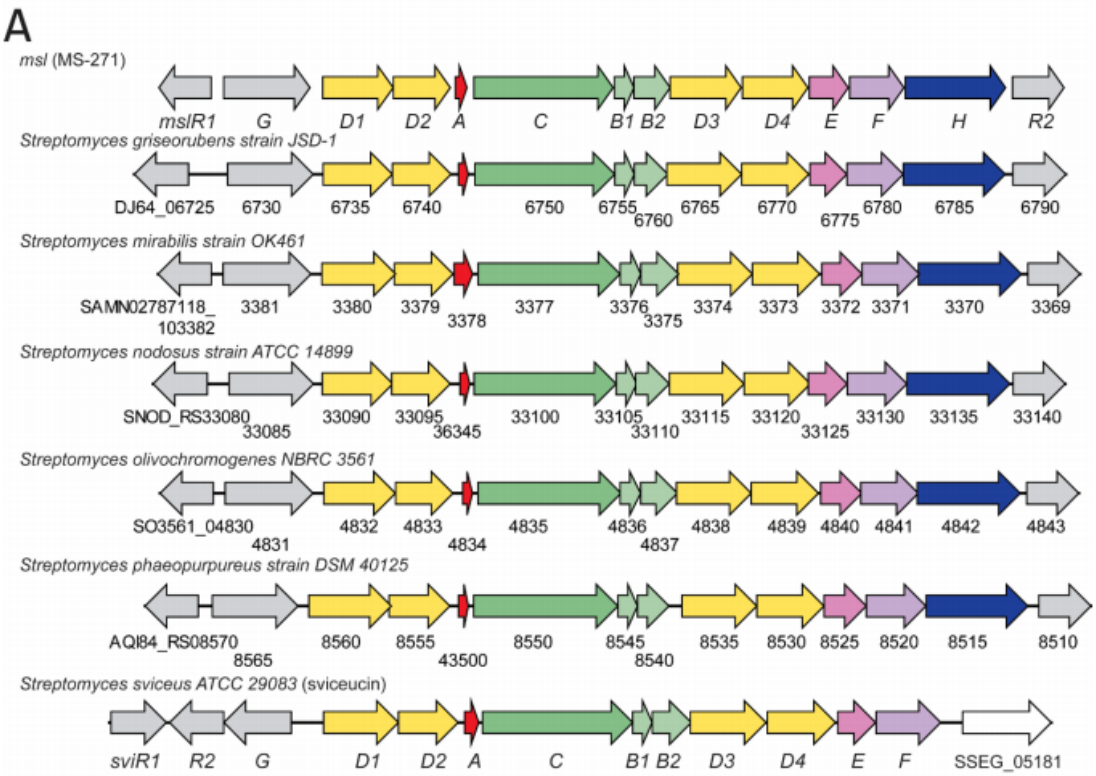


Figure 2-2-4-2 LC-MS analysis of metabolites produced by the mutants i) *S. lividans: mslR2*/pWHM3-msl, ii) /pWHM3-msl-ΔmslB1, iii) /pWHM3-msl-ΔmslB2, iv) /pWHM3-msl-ΔmslC, v) /pWHM3-msl-ΔmslE, vi) /pWHM3-msl-ΔmslF, and vii) /pWHM3-msl-ΔmslH.

To obtain a clue of the function of MslH, I searched for *msl*-like gene clusters by using the core amino acid sequences of MS-271 as the probe. As a result, I identified several such clusters in genome databases (Figure 2-2-4-3 A). All the clusters contained *mslH* orthologues like MS-271 gene cluster, and the core peptides produced by these clusters were almost identical to that of MslA (Figure 2-2-4-3 B). I also searched for gene clusters for lasso peptides possessing two disulfide bonds like MS-271, and found that the sviceucin biosynthetic gene cluster to be the only example. Sviceucin is composed of 20 amino acids with valine at its C-terminus and its sequence was different from those of MS-271. However, the genes in the biosynthetic gene cluster and their organization are the same as those of MS-271 except for the absence of an *mslH* orthologue. These observations suggested that MslH might be a novel epimerase responsible for the epimerization of the C-terminal tryptophan in MS-271 biosynthesis.



B

```
                                                  1
S. sviceus                MTSTDELYEAPELIEIGDYAELTRCVWGGDCTDFLGCGTAWICV*
MS-271                       MSAIYEPPMLQEVGDFEELTKCLGVGSCNDFAGCGYAIVCFW*
S. griseorubens              MTAIYEPPALQEIGDFDELTKCLGIGSCNDFAGCGYAVVCFW*
S. mirabilis                 MSAIYEPPMLQEVGDFEELTKCLGVGSCNDFAGCGYAIVCFW*
S. nodosus                  MSVEIYEPPMLQEIGDFDELTKCLGVGSCNDFAGCGYAIVCFW*
S. olivochromogenes          MSAIYEPPMLQEVGDFEELTKCLGVGSCNDFAGCGYAIVCFW*
S. phaeopurpureus            MSAVYEPPMLQEVGDFDELTKCLGVGSCNDFAGCGYAIVCFW*
```

Figure 2-2-4-3 *msl*-like gene clusters. (A) gene organizations of clusters. Top six clusters have the gene organization identical to that of msl cluster. The bottom cluster is the sviceucin cluster. SSEG_05181 has homology to glycerate kinases and is not related with MslH. (B) Sequence alignment of precursor peptides.

2.3. Discussion

In this chapter, I identified the biosynthetic gene cluster of the D-amino acid-containing lasso peptide MS-271 from *Streptomyces* sp. M-271. Besides the gene for the precursor peptide (MslA), the cluster contained genes encoding enzymes for post-translational modifications such as a macrolactam synthetase (MslC), precursor peptide recognition element (MslB1), cysteine protease (MslB2), disulfide oxidoreductases (MslE, MslF), and a protein of unknown function (MslH). Heterologous expression of the cluster showed that all the necessary genes for MS-271 production were present. Furthermore, geneknock experiment of *mslH* showed that MS-271 production was abolished, suggesting that MslH was indispensable for MS-271 production. These results indicated that MslH was a novel peptide epimerase to introduce D-Trp in MS-271 structure.

In the gene knockout experiments, *mslE* and *mslF* mutants could produce the MS-271 with a reduced amount. MslE had similarity to DoxX family enzymes[9], which were suggested to catalyze thiosulfate oxidation to generate tetrathionate. On the other hand, MslF showed homology to thiol-disulfide oxidoreductases (TDORs) and contained a conserved CXXC motif[10]. Thus, MslE and MslF were predicted to be responsible for the formation of two disulfide bonds in MS-271 structure. Considering that no intermediates such as reduced forms of lasso peptide were detected by LC-MS analysis and that some disulfide-containing lasso peptides clusters identified by genome mining do not contain genes for disulfide bond formation, disulfide bonds in lasso peptides are likely formed spontaneously, and the conversion to correct forms will be guaranteed by MslE and MslF. These speculations were further supported by a previous report that the formation of two disulfide bonds in another lasso peptide sviceucin does not require SviF, the MslF homologue in the sviceucin cluster[11].

References

1. Su Y.; Han M.; Meng X.; Feng Y.; Luo S.; Yu C.; Zheng G.; Zhu S. (2019). Discovery and characterization of a novel C-terminal peptide carboxyl methyltransferase in a lassomycin-like lasso peptide biosynthetic pathway. *Applied Microbiology and Biotechnology.* **103**, 2649.
2. Zhu S.; Hegemann J. D.; Fage C. D.; Zimmermann M.; Xie X.; Linne U.; Marahiel M. A. (2016). Insights into the unique phosphorylation of the lasso peptide paeninodin. *J Biol. Chem.* **291**, 13662.
3. Zong C.; Cheung-Lee W. L.; Elashal H. E.; Raj M.; Link A. J. (2018). Albusnodin: an acetylated lasso peptide from Streptomyces albus. *Chem. Commun*. **54**, 1339.
4. Zhang, C.; Seyedsayamdost, M. R. (2020). CanE, an iron/2-oxoglutarate-dependent lasso peptide hydroxylase from Streptomyces canus. *ACS Chem. Biol.* **15**, 890.
5. Sung, M. H.; Park, C.; Kim, C. J.; Poo, H.; Soda, K.; Ashiuchi, M. (2001). Natural and edible biopolymer poly-gamma-glutamic acid: synthesis, production, and applications. *Chem. Rec.*, **5**, 352-66.
6. a) Zong, C.; Cheung-Lee, W. L.; Elashal, H. E.; Raj, M.; Link, A. J. (2017). Albusnodin: an acetylated lasso peptide from Streptomyces albus. *Chem. Commun*., **54**, 1339. b) Mevaere, J.; Goulard, C.; Schneider, O.; Sekurova, O. N.; Ma, H.; Zirah, S.; Afonso, C.; Rebuffat, S.; Zotchev, S. B.; Li, Y. (2018). An orthogonal system for heterologous expression of actinobacterial lasso peptides in Streptomyces hosts. *Sci. Rep*. **8**, 8232.
7. Onaka, H.; Taniguchi, S.; Ikeda, H.; Igarashi, Y.; Furumai, T. (2003). pTOYAMAcos, pTYM18, and pTYM19, actinomycete-Escherichia coli integrating vectors for heterologous gene expression. *J. Antibiotic.* **56**, 950.
8. Combes, P.; Till, R.; Bee, S.; Smith, M. C. (2002). The Streptomyces gemome contains multiple pseudo-attB sites for the φC31-encoded site-specific recombination system. *J. Bacteriol*. **184**, 5746.
9. Nambi, S.; Long, J. E.; Mishra, B. B.; Baker, R.; Murphy, K. C.; Olive, A. J.; Nguyen, H. P.; Shaffer, S. A.; Sassetti, C. M. (2015). The oxidative stresss network of Mycobacterium tuberculosis reveals coordination between radical detoxification systems. *Cell Host Microbe*, **17**, 829.
10. Kadokura, H.; Katzen F.; Beckwith J. (2003). Protein disulfide bond formation in prokaryptes. *Annu. Rev Biochem*. **72**, 111.

11. Li, Y.; Ducasse, R.; Zirah, S.; Blond, A.; Goulard, C.; Lescop, E.; Giraud, C.; Hartke, A.; Guittet, E.; Pernodet, J. L.; Rebuffat, S. (2015). Characterization of sviceucin from Streptomyces provides insight into enzyme exchangeability and disulfide bond formation in lasso peptide. *ACS Chem. Biol.* **10**, 2641.

Chapter 3

Identification of the peptide epimerase for C-terminal D-tryptophan introduction of ribosomal peptide

3.1 Introduction

In chapter 2, I identified the biosynthetic gene cluster of MS-271 (*msl*) by draft genome analysis. Heterologous expression experiments of the *msl* cluster revealed that all necessary biosynthetic genes were present in spite of lacking an obvious candidate gene for the epimerization. Further gene-knock experiment showed MslH is indispensable for MS-271 biosynthesis. Given that MslH was the only protein of unknown function in the *msl* cluster, I speculated that MslH was a novel peptide epimerase responsible for the epimerization of the C-terminal tryptophan.

To investigate the function of MslH, I first sought to obtain the precursor peptide MslA, a possible substrate. Although its high hydrophobicity prevented it from being acquired by solid-phase synthesis, I succeeded to express it in *E. coli* by adding a small peptide tag (SKIK tag) at its N-terminus. Then I co-expressed the precursor peptide with MslH in *E. coli*. After the purification and acid hydrolysis of the peptide product, D-Trp was observed by a modified Marfey's method. These *in vivo* experiments indicated that MslH catalyzed the epimerization of the nascent precursor peptide to generate its epimer. Additional coexpression of precursor peptide recognition element (*mslB1*) enhanced the formation of MslA. To further investigate the catalytic mechanism of MslH, I prepared the recombinant enzymes for *in vitro* experiments. The results showed that MslH catalyzed the epimerization of MslA in a cofactor-independent manner and that the precursor peptide recognition element MslB1 accelerated the reaction rate probably by forming a ternary complex among MSlA, -B1 and -H. Considering that many modification enzymes involved in RiPP biosynthesis require leader peptides for their substrate recognition, I next investigated whether leader peptide is important for the function of MslH by using a chemically synthesized core peptide as a substrate. However, no obvious epimerization was observed. This result

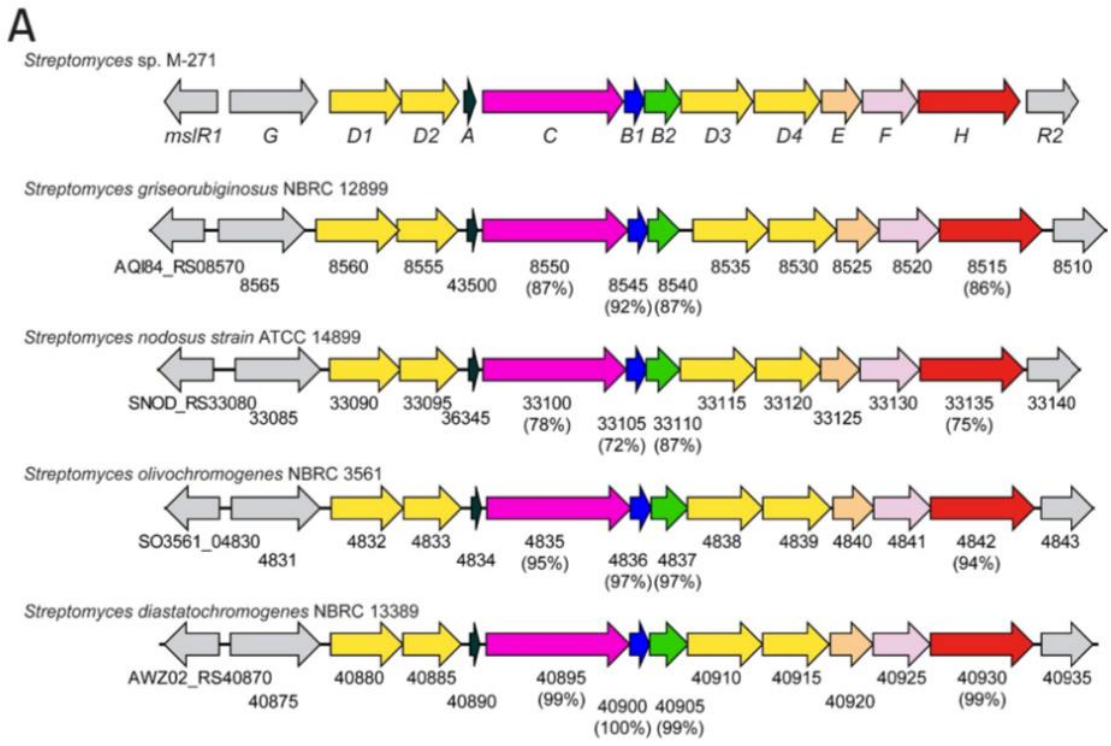
showed that full-length MslA is the substrate of MSlH and the leader peptide is required for substrate recognition. To probe the substrate specificity of the modification enzymes including MslH in MS-271 biosynthesis, I next carried out heterologous expression of the *msl* cluster to produce MS-271 derivatives by altering the core peptide sequences of the *mslA*. Consequently, I demonstrated broad substrate specificity of MslH toward the N-terminal region of the core peptide by heterologous production of unnatural lasso peptides, suggesting it as a potentially useful tool for rational peptide bioengineering.

In this chapter, the details of these experiments are described.

3.2. Results

3.2.1. *in vivo* characterization of MslH as a novel peptide epimerase

To gain initial insight into the epimerization, I used an *in vivo* heterologous expression strategy with *E. coli* BL21(DE3) as the host. To express the proteins, genes of precursor peptide and biosynthetic enzymes were cloned into compatible expression vectors pRSFDuet-1, pETDuet-1, pCDFDuet-1 and pACYCDuet-1, resulting into the construction of pRSF-mslA-WT (WT means wild type), pET-mslB1-mslB2-WT, pCDF-mslE-mslF-WT and pACYC-mslC-mslH-WT. However, only the expression of MslB1 was observed in SDS-PAGE analysis. To obtain the recombinant biosynthetic proteins, I next searched for *msl*-like clusters by using the amino acid sequence of core peptide as the probe. As shown in Figure 3-2-1-1, I identified several clusters where all biosynthetic enzymes showed high identities to proteins encoded in the cluster and precursor peptides showed little difference from each other in the leader peptide sequence.



B

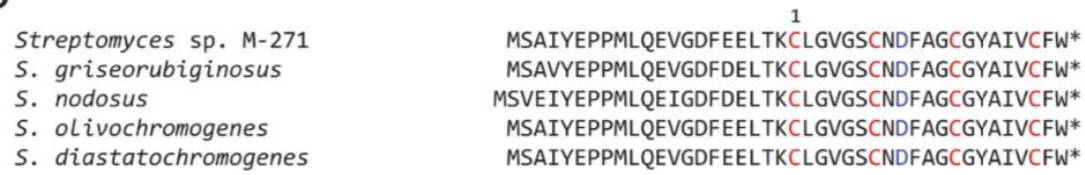


Figure 3-2-1-1. *Msl*-like clusters. (A) The gene organizations of the clusters. Identities to Msl proteins from *Streptomyces* sp. M-271 are shown in parentheses. (B) Sequence alignment of precursor peptides.

To confirm the activities of biosynthetic enzymes, *Streptomyces diastatochromogenes* NBRC 13389, *Streptomyces griseorubiginosus* NBRC 12899, *Streptomyces nodosus* strain ATCC 14899 and *Streptomyces olivochromogenes* NBRC 3561 were cultured along with *Streptomyces* sp. M-271 and their metabolites were analyzed by HPLC. As shown in Figure 3-2-1-2, *Streptomyces griseorubiginosus* NBRC 12899 possessing the same enzymes (86%-92% identities) produced a large amount of MS-271, indicating the expression amounts of biosynthetic proteins were higher than those from *Streptomyces* sp. M-271. Thus, I tried to express the recombinant proteins by using the genomic DNA of *S. griseorubiginosus* NBRC 12899.

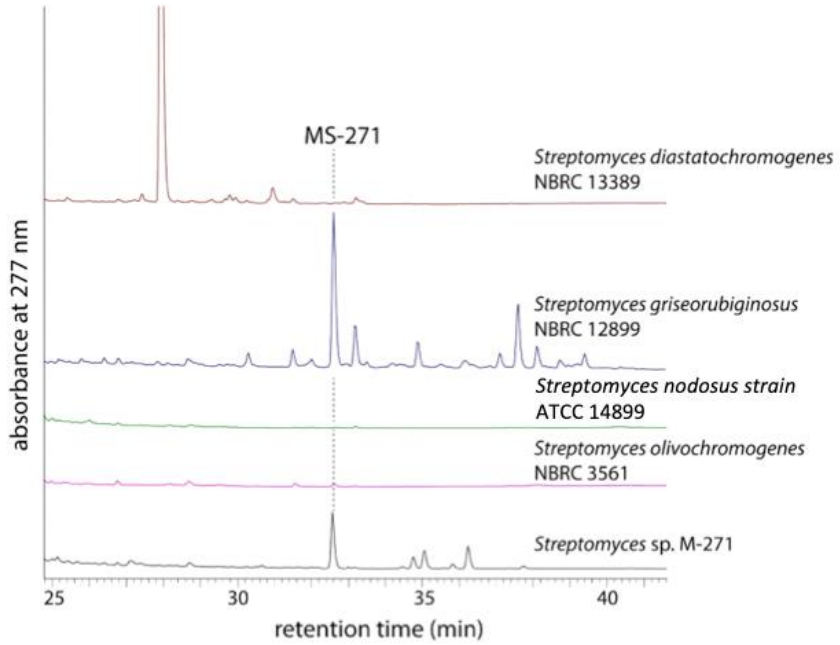


Figure 3-2-1-2 Production of MS-271 by *Streptomyces* strains harboring *msl*-like gene clusters.

Considering that many modification enzymes involved in RiPP biosynthesis require their leader peptides for substrate recognition, I postulated that epimerization may occur on the full-length MslA prior to the cleavage of the precursor peptide by MslB2[1]. As a first step, I tried to obtain the MslA which cannot be chemically synthesized due to its high hydrophobicity. Accordingly, the genomic DNA of *S. griseorubiginosus* NBRC 12899 was used and *mslA* was cloned into pET24b expression vector. The constructed plasmids pET24-mslA were introduced into *E. coli* BL21(DE3) to express the recombinant MslA. However, no MslA band was observed on SDS-PAGE analysis. In addition, I used a cell-free transcription/translation system to synthesize MslA by using the PURExpress In Vitro Protein Synthesis Kit (NEB #E6800) and the fragment between promotor and terminator region of pET24-mslA as the template. However, no MslA band was observed. Ojima-Kato et al. recently reported that the addition of a DNA sequence coding for Ser-Lys-Ile-Lys (SKIK tag) just after the start codon

significantly improved the expression of recombinant proteins in *E. coli* and *Saccharomyces cerevisiae* though its mechanism remained unclear[2]. Using this strategy, I constructed the pET24-SKIK-mslA and pRSF-SKIK-mslA to express the recombinant MslA fused with an SKIK tag at its N-terminus. In this case, the recombinant MslA (5.1 kDa) was observed in the insoluble fractions (Figure 3-2-1-3 lane 6). I also constructed the pET24-SKIKHis-mslA and pRSF-SKIKHis-mslA to express the MslA fused with a His tag sequence followed by an SKIK tag sequence. As the results, relatively larger amount of SKIK-His-tagged MslA was observed with pET24-SKIKHis-mslA (Figure 3-2-1-3 lane 3, 4), while a small amount of SKIK-His-tagged MslA (5.8 kDa) appeared both in the soluble and insoluble fractions with pRSF-SKIKHis-mslA (Figure 3-2-1-3 lane 9, 10). Then, I attempted to increase the MslA by changing the expression condition. However, all attempts including changing the induction temperature and timing failed.

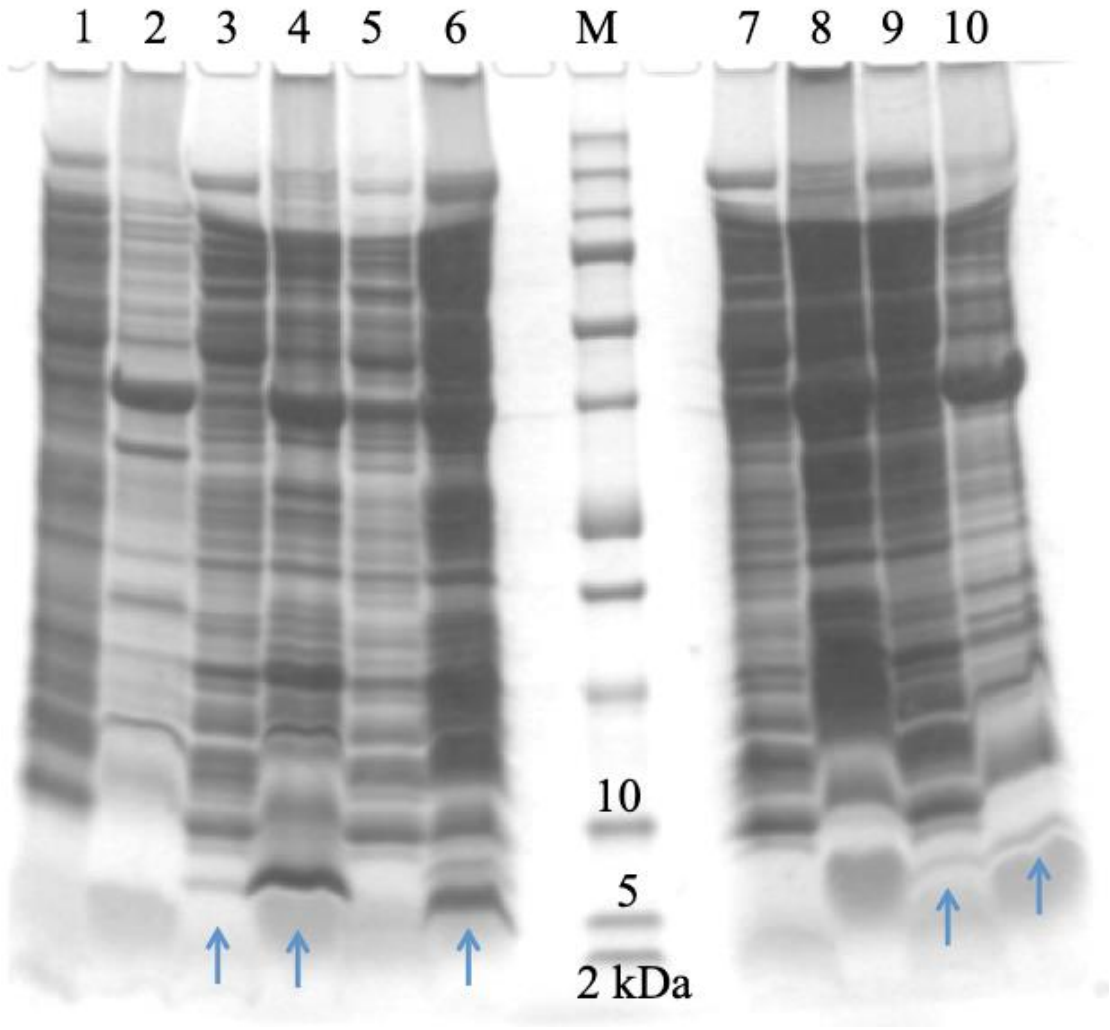


Figure 3-2-1-3 Tricine-SDS-PAGE analysis of recombinant MslA. Soluble fractions of *E. coli* BL21(DE3) harboring (1) empty pET24b as a control; (3) pET24-SKIKHis-mslA; (5) pET24-SKIK-mslA; (7) pRSF-SKIK-mslA; (9) pRSF-SKIKHis-MslA. Insoluble fractions of *E. coli* BL21(DE3) harboring (2) empty pET24b as a control; (4) pET24-SKIKHis-mslA; (6) pET24-SKIK-mslA; (8) pRSF-SKIK-mslA; (10) pRSF-SKIKHis-mslA. M, marker. After IPTG was added, induction was performed at 16 ˚C overnight.

Since MslA was obtained, I next carried out coexpresstion of MslA with MslH in *E. coli*. to investigate the function of the MslH. I cloned *mslH* into the vector pCDF-Duet-1 and constructed pCDF-His-mslH. This plasmid was introduced into *E. coli* BL21(DE3) harboring the pET24-SKIKHis-mslA or pRSF-SKIKHis-mslA. However, no MslA was observed in both sets of plasmids though His-tagged MslH (50.6 kDa) was expressed (Figure 3-2-1-4 lane 4, 5, 6, 4', 5', 6', 7). In the control experiment using the transformant harboring pET24-SKIKHis-mslA and pCDFDuet-1, MslA was observed in the insoluble fractions (Figure 3-2-1-4 lane 1, 2, 3). I also cloned *mslH* into the vector pACYC-Duet-1 vector and constructed pACYC-His-mslH but it again resulted no expression of MslA.

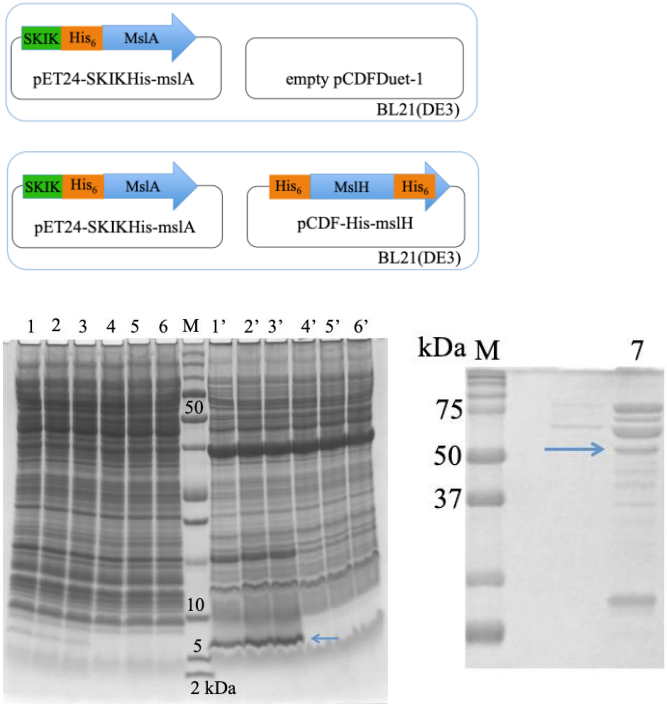


Figure 3-2-1-4 Plasmid sets and Tricine-SDS-PAGE analysis of recombinant MslA and MslH. Soluble/insoluble fractions of *E. coli* BL21(DE3) harboring pET24-SKIKHis-mslA and empty pCDFDuet-1 vector as a control experiment under different culture conditions: (1/1') OD_{600}= 0.7, 0.5 mM IPTG, 16°C, 180rpm, O/N; (2/2') OD_{600}= 1.2, 0.5 mM IPTG, 16°C, 180 rpm, O/N; (3/3') OD_{600}= 1.2, 1 mM IPTG, 16°C, 180 rpm, O/N; Soluble/insoluble fractions of *E. coli* BL21(DE3) harboring pET24-SKIKHis-

mslA and pCDF-His-mslH under different culture conditions: (4/4'): OD_{600}= 0.7, 0.5 mM IPTG, 16°C, 180 rpm, O/N; (5/5') OD_{600}= 1.2, 0.5 mM IPTG, 16°C, 180 rpm, O/N; (6/6') OD_{600}= 1.2, 1.0 mM IPTG, 16°C, 180 rpm, O/N. (7) His_6-tagged MslH at both termini (50.6 kDa). Number with a "'" symbol represents the insoluble fraction of each experiment.

During optimization, both MslA and MslH were successfully expressed using the plasmid combination of pET24-SKIKHis-mslA and pSTV-His-mslH, which was constructed by inserting *mslH* into pSTV28 vector, though the MslA were insoluble (Figure 3-2-1-5).

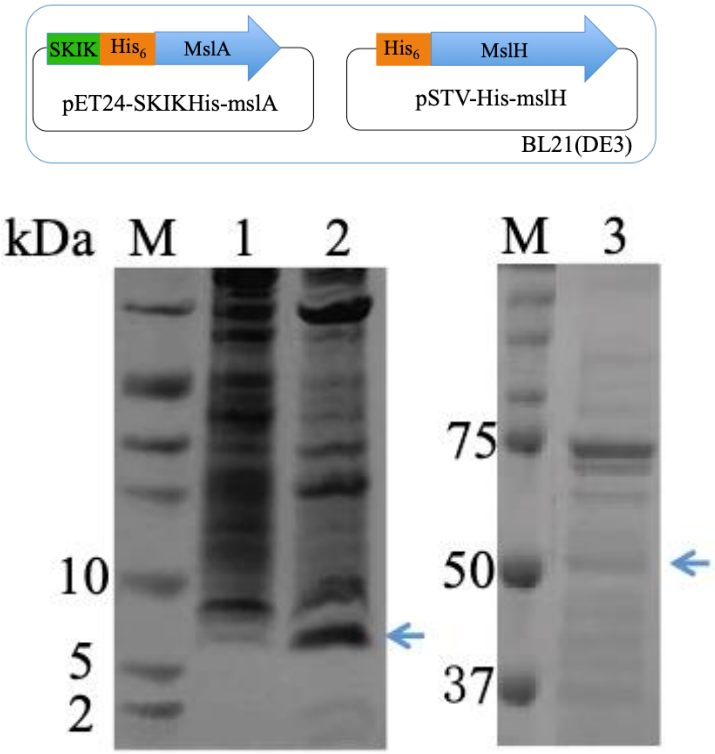


Figure 3-2-1-5 Plasmid set and Tricine-SDS-PAGE analysis of recombinant MslA and MslH. 1, soluble fractions; 2, insoluble fractions containing MslA (5.8 kDa); 3, purified MslH (49.1 kDa) with a low purity.

Considering that MslB1 recognizing the leader peptide may be required for the substrate recognition by the epimerase, I next sought to co-express MslA with MslH and MslB1. I cloned *mslB1* into pCDFDuet-1 and constructed pCDF-His-mslB1.

pCDF-His-mslB1 was then introduced to *E. coli* BL21(DE3) harboring pET24-SKIKHis-MslA and pSTV-His-mslH. In this case, MslA, MslH and MslB1 were found in the soluble fraction (Figure 3-2-1-6). Intriguingly, hydrophobic MslA was expressed as a soluble peptide in the presence of MslB1 in contrast to when it was expressed alone, probably due to the interaction between them.

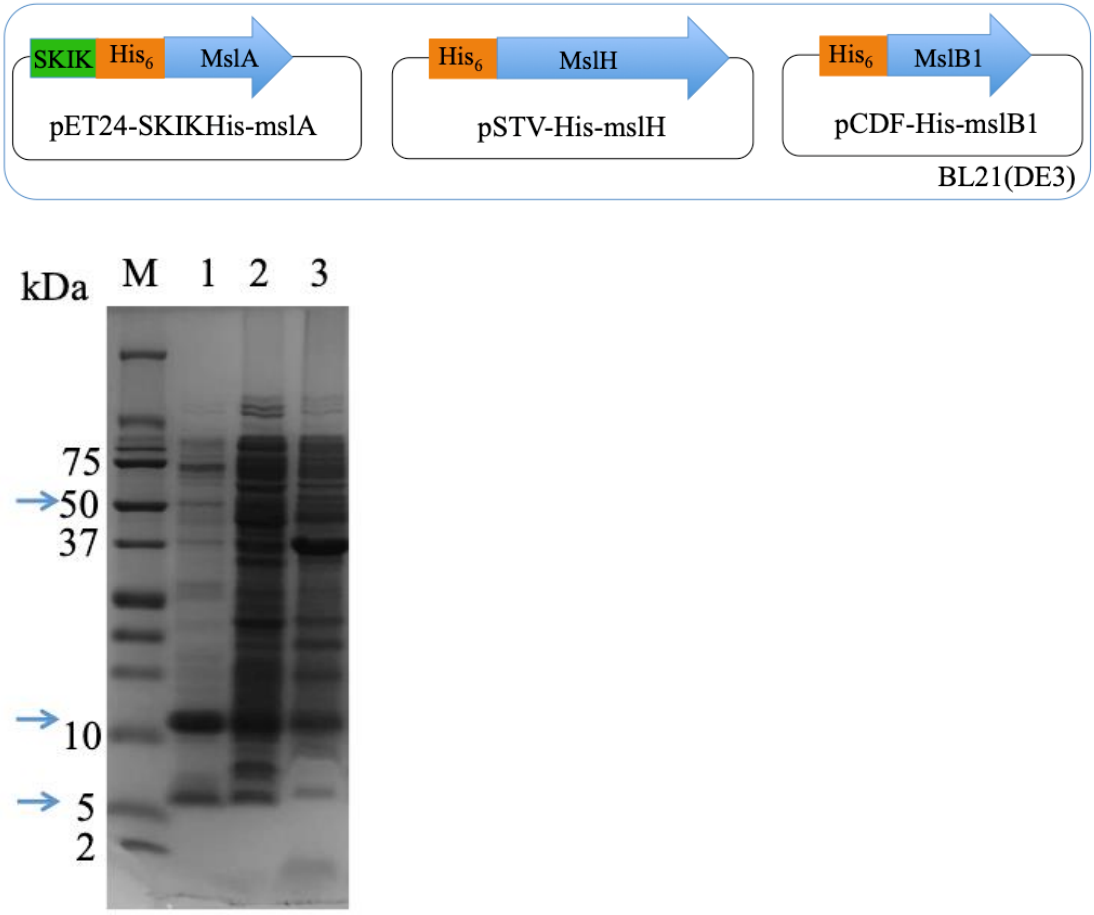


Figure 3-2-1-6 Plasmid set and Tricine-SDS-PAGE analysis of recombinant MslA (5.8 kDa), MslH (49.1 kDa) and MslB1 (10.4 kDa). M, marker; 1, purified His-tagged MslA, MslH and MslB1; 2, soluble fractions; 3, insoluble fractions.

To examine whether epimerization of the peptide product occurred, the MslA product was purified by Tricine-SDS-PAGE (16% gel) and the chirality of the C-terminal Trp, which is the sole Trp residue in MslA, was analyzed by a modified Marfey's method after acid hydrolysis followed by HPLC purification of Trp. As shown

in Figure 3-2-1-7, when *mslH* was co-expressed with *mslA* and *mslB1*, D-Trp was observed. The formation of D-Trp was also observed, albeit a small amount, when *mslB1* was omitted. In contrast, D-Trp was not detected in the purified MslA products prepared without expression of *mslH*. These results clearly indicated that MslH catalyzed the epimerization of the nascent precursor peptide to generate its epimer, epi-MslA, and that MslB1 enhanced the activity of MslH probably by forming a ternary complex with MslA.

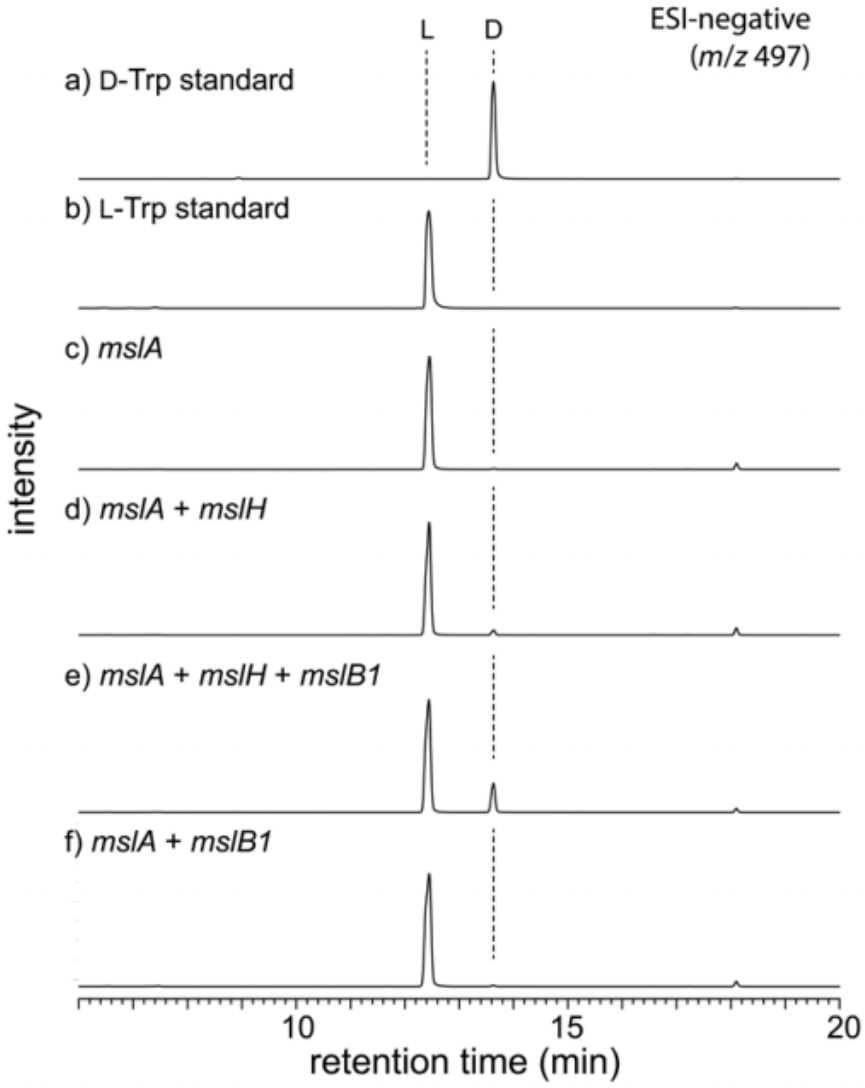


Figure 3-2-1-7 LC-MS analysis (ESI negative ion mode, monitored at *m/z* 497) of *in vivo* assay. L-FDLA derivatives of a) D-Trp standard and b) L-Trp standard. L-FDLA derivatives of Trp in MslA by heterologous expression of c) *mslA* alone, d) *mslA* and *mslH*, e) *mslA*, *mslH*, and *mslB1*, and f) *mslA* and *mslB1*.

In addition, I examined whether expression of MslB2 or MslC, which would be involved in the lasso formation, affected the epimerization activity. I first cloned *mslB2* into pColaDuet-1 to construct pCola-His-mslB2 but no expression of MslB2 was observed. However, I finally succeeded in expressing MslB2 by adding an MBP tag (Figure 3-2-1-8 lane 1) at N-terminus of MslB2. I also cloned *mslC* into pColaDuet-1 to construct pCola-His-mslC but recombinant MslC formed inclusion body. However, MslC was also expressed as a soluble form by adding an MBP tag (Figure 3-2-1-8 lane 4). However, expression of MslB2 or MslC resulted in no effects on the epimerization activity (Figure 3-2-1-9). Tricine-SDS-PAGE analysis of protein expressions for *in vivo* analysis was summarized in Figure 3-2-1-10.

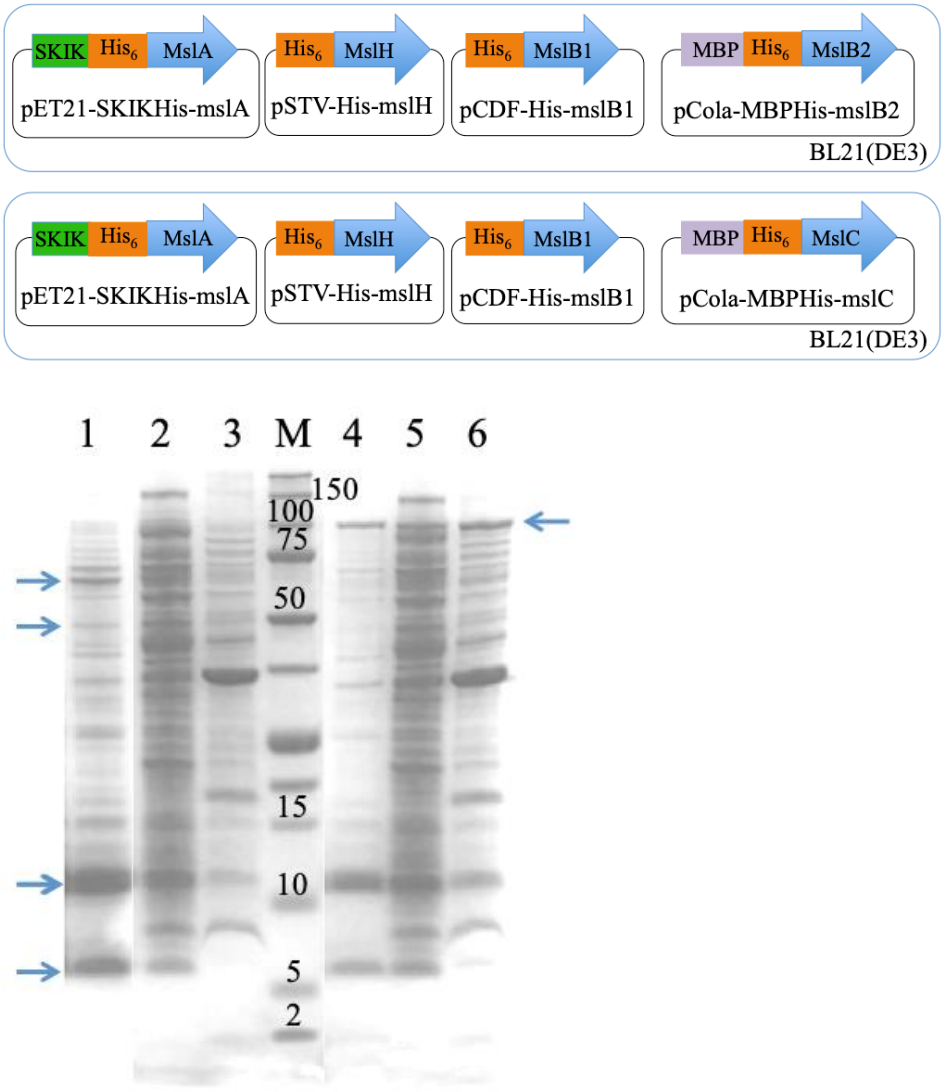


Figure 3-2-1-8 Plasmid sets and Tricine-SDS-PAGE analysis of recombinant proteins (MslA 5.8 kDa, MslH 49.1 kDa, MslB1 10.4 kDa and MslB2 64.5 kDa or MslC 109.8 kDa). 1, purified fractions of MslA, -H, -B1 and -B2 coexpression; 2, soluble fractions of MslA, -H, -B1 and -B2 coexpression; 3, insoluble fractions of MslA, -H, -B1 and -B2 coexpression; M, marker; 4, purified fraction of MslA, -H, -B1 and -C coexpression; 5, soluble fractions of MslA, -H, -B1 and -C coexpression; 6, insoluble fractions of MslA, -H, -B1 and -C coexpression.

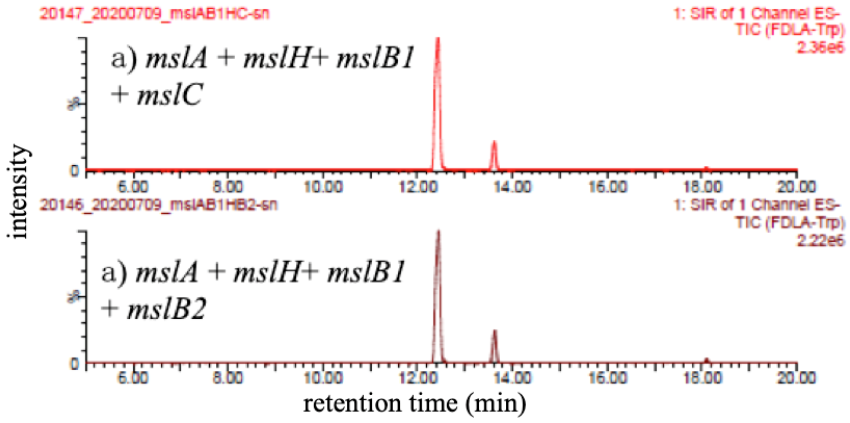


Figure 3-2-1-9 LC-MS analysis (ESI negative ion mode, monitored at *m/z* 497) of *in vivo* assay. L-FDLA derivatives of Trp in MslA by heterologous expression of a) *mslA*, *mslH*, *mslB1* and *mslC*, and b) *mslA*, *mslH*, *mslB1*and *mslB2*.

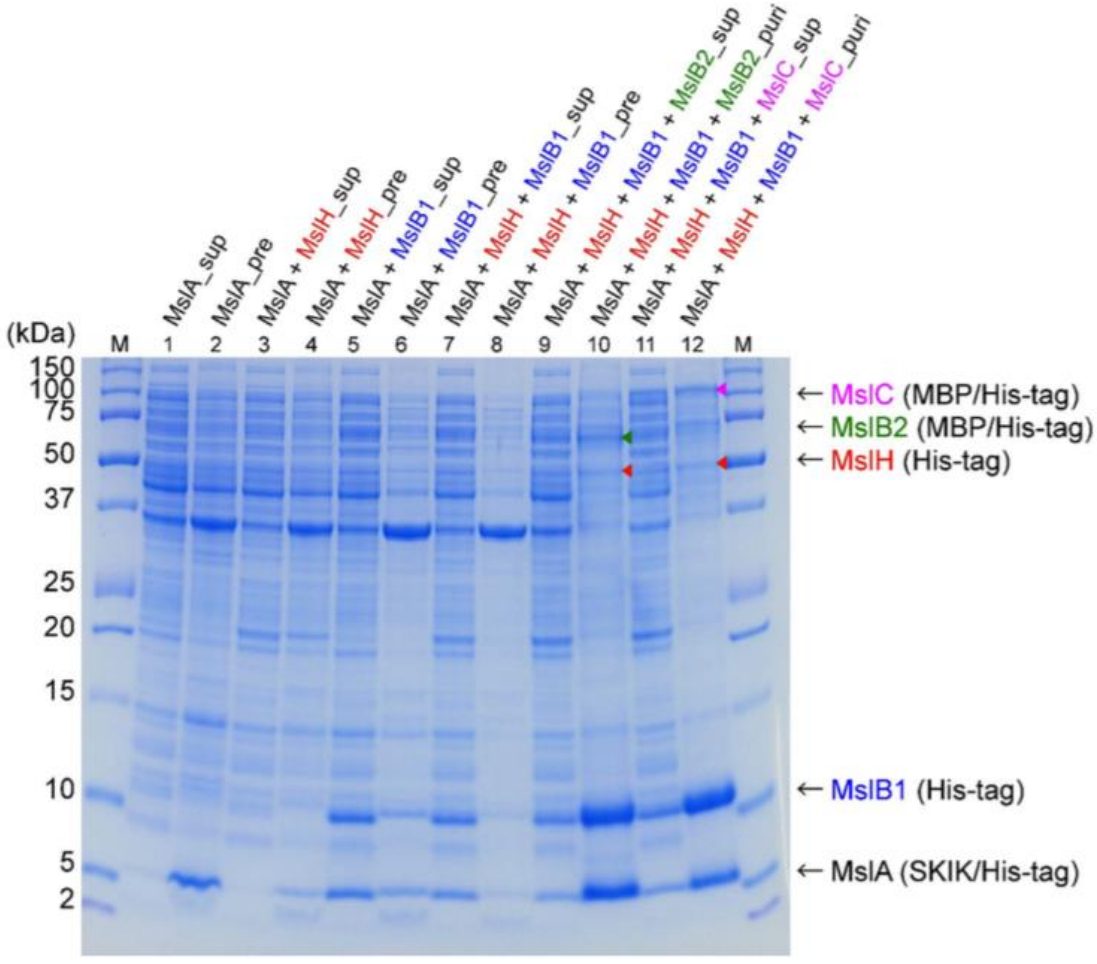


Figure 3-2-1-10 Tricine-SDS-PAGE analysis of recombinant MslA (5.8 kDa), MslB1 (10.4 kDa), MslH (49 kDa), MslB2 (65 kDa), and MslC (110 kDa) in *in vivo* expressions. The odd and even lane numbers show supernatant (sup) and precipitate (pre) fractions of cell extracts of the transformants expressing MslA (Lanes 1 and 2),

MslA + MslH (Lanes 3 and 4), MslA + MslB1 (Lanes 5 and 6), MslA + MslH + MslB1 (Lanes 7 and 8), MslA + MslH + MslB1 + MslB2 (Lane 9), and MslA + MslH + MslB1 + MslC (Lane 11). To confirm protein expression of MslH, MslB2, and MslC, purified fractions (puri) with Ni-NTA column of cell extracts of the transformants expressing MslA + MslH + MslB1 + MslB2 (Lane 10) and MslA + MslH + MslB1 + MslC (Lane 12) were analyzed (MslH, MslB2 and MslC are marked with triangles.).

3.2.2. *in vitro* characterization of MslH as a metal- and cofactor-independent epimerase

To further investigate the MslH-catalyzed epimerization, I next carried out the *in vitro* experiments. Because MslH has no specific motifs such as cofactor binding domains, I did not know what cofactors were needed for the epimerization. So, I first tested MslH catalyzing epimerization in the presence of *E. coli* cell lysate. Because substrate MslA was insoluble when expressed alone, I prepared the recombinant His-tagged protein mixture of MslA and MslB1 from *E. coli* BL21(DE3) using pET21-SKIKHis-mslA and pCDF-His-mslB1 and purified them with Ni-NTA agarose affinity chromatography. Recombinant His-tagged MslH was purified from *E. coli* BL21(DE3) harboring pSTV28-His-mslH. The purified His-tagged MslH was added to the purified protein mixture of His-tagged MslA and MslB1 with (Figure 3-2-2-1 b) or without (Figure 3-2-2-1 a) *E. coli* cell lysate. The cell lysate expressing MslH was also directly added to the MslA and MslB1 mixture (Figure 3-2-2-1 c). After incubation at 30°C for 1 hour, the chirality of the C-terminal Trp was analyzed using the same method. Unexpectedly, D-Trp was detected in all experiments and the conversion rate was similar. These results suggested that MslH catalyzed the epimerization in a cofactor-independent manner.

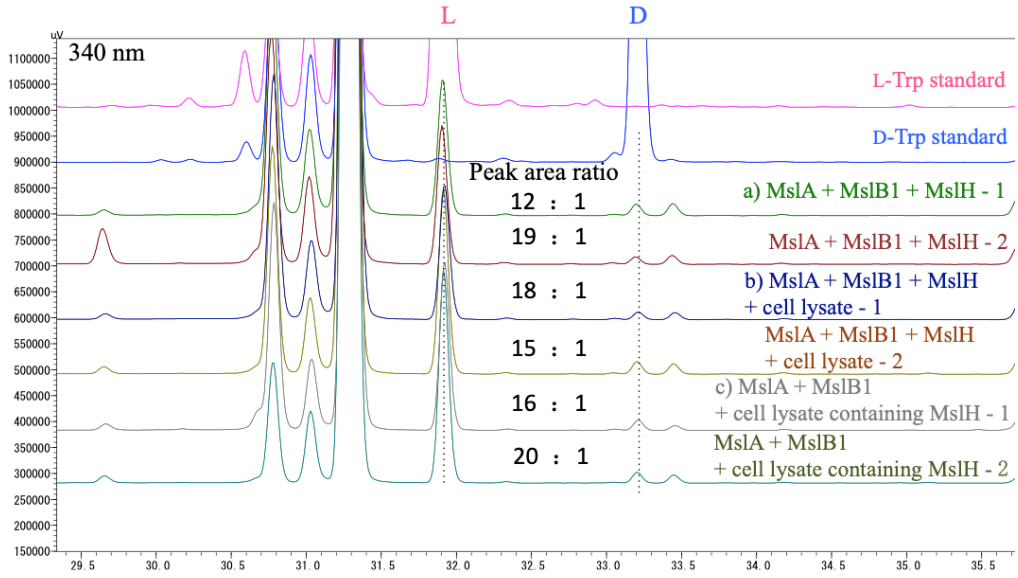


Figure 3-2-2-1 HPLC analysis of MslA, MslB1, MslH reaction with cell lysate. Each experiment was performed in duplicate.

I next carried out the *in vitro* experiments to confirm MslH as a cofactor-independent peptide epimerase. To prepare MslA and MslB1 independently, I cloned MslB1 into pCDFDuet-1 vector to construct pCDF-MslB1 without a His tag and it was introduced to *E. coli* BL21(DE3) harboring pET21-SKIKHis-mslA. I then purified MslA by Ni-NTA agarose affinity chromatography. Unexpectedly, MslB1 was co-purified with MslA. To remove the MslB1, I tried washing the Ni-NTA column with high salt buffer (wash buffer I containing 2 M NaCl) before elution but MslB1 was still coeluted, indicating tight binding of MslB1 to MslA (Figure 3-2-2-2 lane 2). I thus tried to remove MslB1 under denaturing conditions. By washing the Ni-NTA column with the wash buffer containing 8M urea, highly purified MslA was obtained (Figure 3-2-2-2 lane 3). After the concentration, MslA solution became opaque, indicating the precipitation of hydrophobic MslA. His-tagged MslB1 and MslH was purified and obtained independently (Figure 3-2-2-2 lane 4, 5). When MslH (5 μM) was incubated with MslA (220 μM), the formation of D-Trp (~7%) in MslA was observed after

incubated at 30°C for 16 hours in 100 mM Tris buffer (pH 8.0) (Figure 3-2-2-3 c). Similar to the *in vivo* results, the formation of D-Trp was increased to ~19% upon addition of MslB1 (220 μM) (Figure 3-2-2-3 d). These results showed that MslH is a cofactor-independent peptide epimerase which catalyzed MslA into its epimer epi-MslA. Interestingly, the reaction fluid became transparent in about 10 minutes after the addition of MslB1 to the insoluble MslA (Figure 3-2-2-3). This once again suggested that the hydrophobic MslA was solubilized through the interaction with MslB1. Since MslH was identified as a cofactor-independent peptide epimerase by *in vitro* experiments, the epimerization reaction would be bidirectional. Consequently, it remained confusing why the conversion rate was not near 50% after overnight incubation beacuse the free energy of two isomers (42-aa MslA and epi-MslA) would be similar.

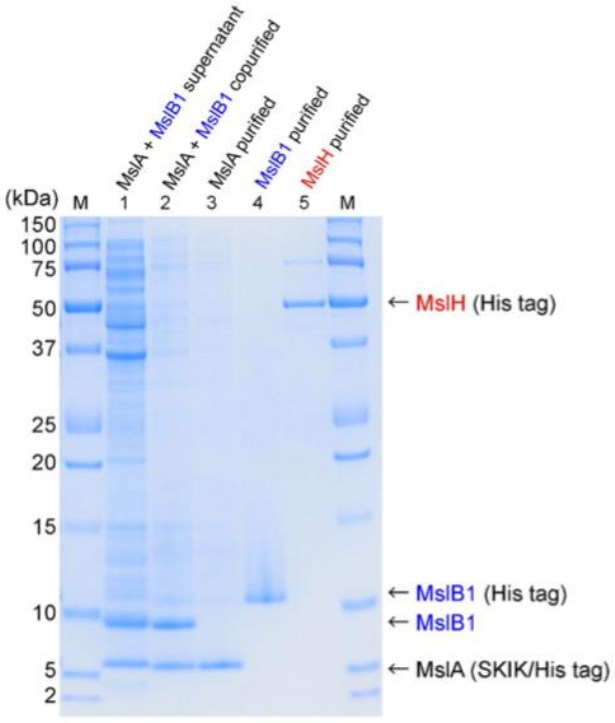


Figure 3-2-2-2 Tricine-SDS-PAGE analysis of purified proteins for *in vitro* assay. Lane 1; Cell-free extract from the *E. coli* BL21(DE3) transformant harboring pET21-SKIK-His-mslA and pCDF-mslB1. Lane 2; Elution fraction of His-tagged MslA (5.8 kDa)

after the Ni-NTA column was washed with high salt buffer (wash buffer I containing 2 M NaCl). MslB1 (9.2 kDa) was copurified with His-tagged MslA. Lane 3; Elution fraction of His-tagged MslA using denaturing conditions. Lane 4; Purified His-tagged MslB1 (10.4 kDa). Lane 5; Purified His-tagged MslH (49 kDa).

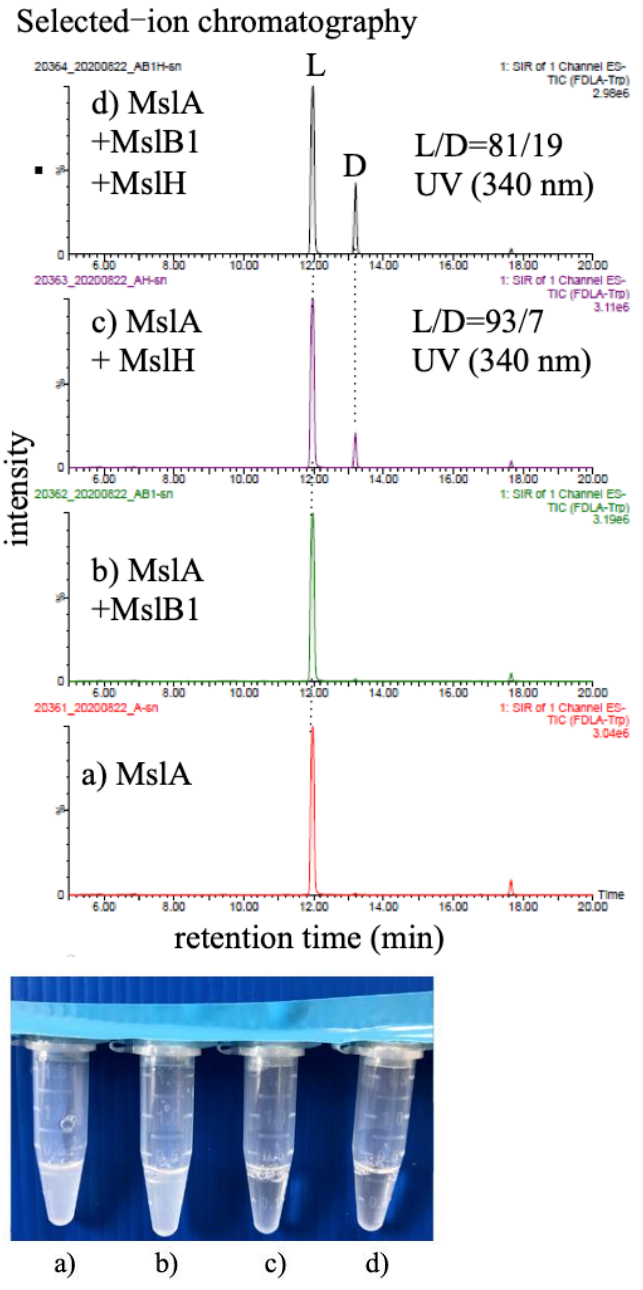


Figure 3-2-2-3 Left: LC-MS analysis (ESI negative ion mode, monitored at *m/z* 497) of *in vitro* assay. L-FDLA derivatives of Trp in MslA by in vitro reaction of a) *mslA*, b) *mslA* and *mslB1*, c) *mslA* and *mslH*, d) *mslA*, *mslB1*and *mslH*. Right: corresponding appearance of in vitro reactions.

Considering the low conversion rate of epi-MslA may be due to the inactivation of MslH, I thus carried out an *in vitro* reaction using MslH which was put in ice for a week. However, epimerization of MslA was still observed and the conversion rate did not change, suggesting MslH was a stable protein. Because MslA contains four cysteine residues, intermolecular or intramolecular disulfide bonds can be formed by oxidation, resulting into the formation of cyclic MslA (5.8 kDa) or MslA dimer (11.6 kDa). Indeed, I observed a possible band of MslA dimer on the gel after loading a increased amount of MslA (Figure 3-2-2-4). Thus, I speculated that MslH could not catalyze the oxidized forms of MslA. To ensure MslA in its reduced form, I next added the reducing agent dithiothreitol (DTT) to the reaction mixture.

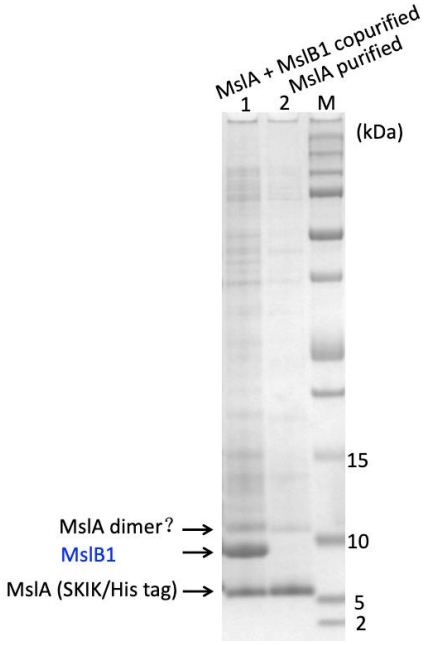


Figure 3-2-2-4 Tricine-SDS-PAGE analysis. Lane 1; Elution fraction of His-tagged MslA (5.8 kDa) after the Ni-NTA column was washed with high salt buffer (wash buffer I containing 2 M NaCl) from the *E. coli* BL21 (DE3) transformant harboring pET21-SKIK-His-mslA and pCDF-mslB. MslB1 (9.2 kDa) was copurified with His-tagged MslA. MslA dimer band (11.6 kDa) may be observed. Lane 2; Elution fraction of His-tagged MslA using denaturing conditions.

As shown in Figure 3-2-2-5, when MslH (5 μM) was incubated with MslA (470 μM), the formation of D-Trp in MslA was observed in a time-dependent manner, as expected, the reaction was accelerated upon the addition of MslB1 under the reducing condition (10 mM DTT). Prolonged incubation (16 h) resulted in about 50% conversion of MslA to epi-MslA, indicating that the MslH catalyzing the epimerization generated an equilibrium mixture of the epimers (Figure 3-2-2-6). Since I have never identified

the C-terminal L-Trp-containing MS-271 derivatives in the producers, the following enzyme of the pathway possibly only recognizes epi-MslA as its substrate.

To test whether leader peptide was important for the core peptide recognition by MslH, I next carried out an *in vitro* experiments of MslH using a chemically synthesized core peptide of MslA (MslA-core) as a substrate. I incubated the MslH (10 μM) with MslA-core (450 μM) for 1 h at 30°C under reducing condition. MslB1 (600 μM) was also added as a control experiment. At this time, only a small amount of D-Trp was observed after the reaction (Figure 3-2-2-5 F-I). These results confirmed that Msl-core is not the substrate of MslH and suggesting that the leader peptide is important for substrate recognition by MslH.

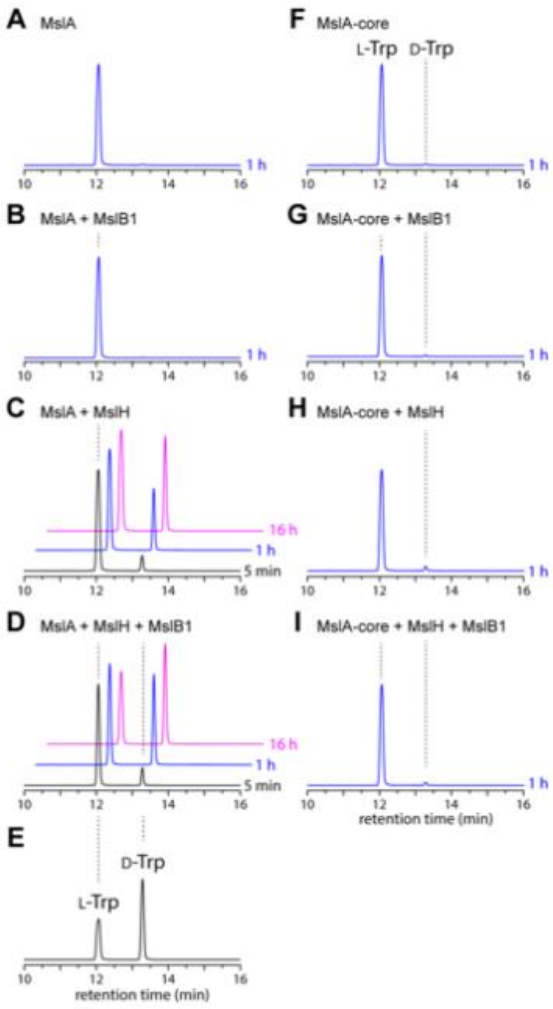


Figure 3-2-2-5 LC-MS analysis (ESI negative ion mode, monitored at *m*/*z* 497) of L- FDLA-Trp derived from *in vitro* reactions. The reaction time is shown next to each trace. (A) MslA only, (B) MslA + MslB1, (C) MslA + MslH, (D) MslA + MslH + MslB1, (E) L- and D-Trp standards, (F) MslA-core only, (G) MslA-core + MslB1, (H) MslA-core + MslH, and (I) MslA-core + MslH + MslB1.

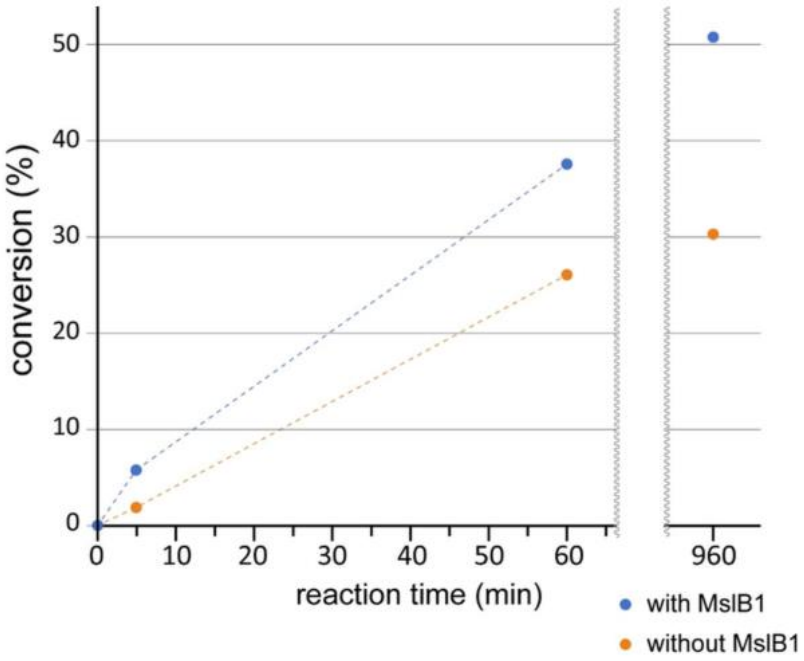


Figure 3-2-2-6 Plot of reaction time (5 min, 60 min, and 16 h) versus conversion from MslA to its epimer by MslH reaction in the presence and absence of MslB1. The conversion (%) was calculated based on the peak area of chromatograms of UV absorbance monitored at 340 nm.

Since web-based programs predicting protein structure (including Phyre[3] and I-TASSER[4]) suggested that MslH belongs to the metallo-dependent phosphatase family. Thus, I probed the metal dependency of MslH (20 μM). MslH was incubated with MslA (430 μM) further supplemented with 5 mM divalent metal ions (Mg^{2+}, Zn^{2+}, Co^{2+}, Fe^{2+}, Mn^{2+}, and Ca^{2+}) or ethylenediaminetetraacetic acid (EDTA), respectively. LC-MS analysis after incubation for 2 h at 30°C in 100 mM Tris (pH 8.0) with various ion or EDTA showed that the activity of MslH was not influenced, indicating that a divalent metal is not required for MslH activity (Figure 3-2-2-7).

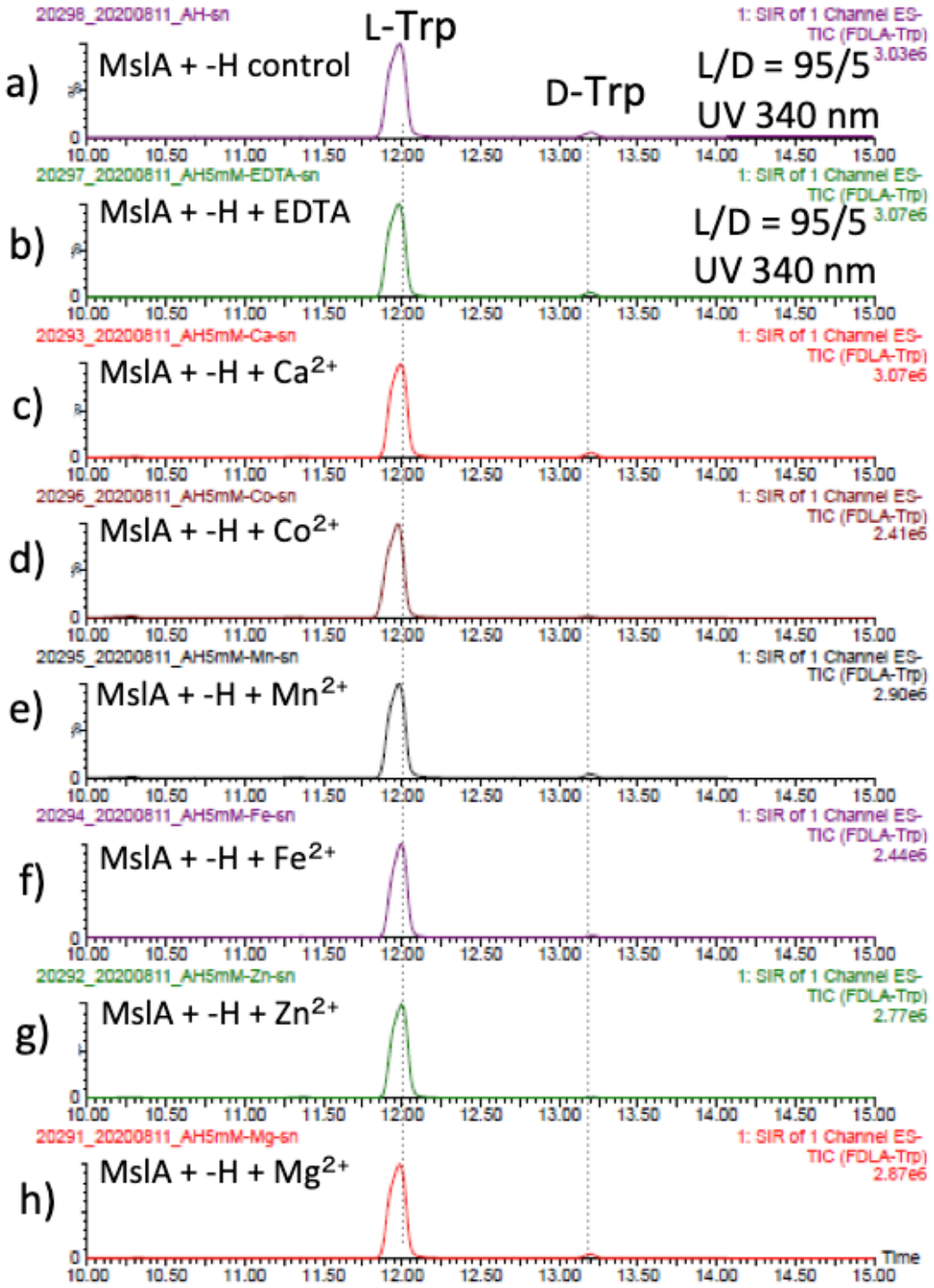


Figure 3-2-2-7 LC-MS analysis (ESI negative ion mode, monitored at *m/z* 497) of *in vitro* assay. L-FDLA derivatives of Trp in MslA by *in vitro* reaction of MslH further supplemented with b) EDTA, c) Ca^{2+}, d) Co^{2+}, e) Mn^{2+}, f) Fe^{2+}, g) Zn^{2+}, h) Mg^{2+}. a) is a control experiment.

3.2.3 Heterologous production of D-amino acid containing unnatural lasso peptides

To date, more than 70 lasso peptides have been isolated. Among them, only MS-271 (also known as Siamycin I) is featured by a D-Aa. Considering that some lasso peptides were reported to have potentials as novel antibiotics due to their different mechanisms of action, I thought that the addition of a C-terminal D-amino acid would very likely improve their stabilities though the exploration and understanding of substrate tolerance of the biosynthetic enzymes are limited. To probe the substrate specificity of the biosynthetic enzymes involved in MS-271 biosynthesis, I next carried out heterologous expression of the *msl* cluster to produce MS-271 derivatives by altering the core peptide sequences of the *mslA* gene using the same method described in chapter 2. The constructed plasmids were individually introduced into *Streptomyces lividans*:*mslR2* and the metabolites were analyzed by HPLC and LC-MS. First, to test whether MslH can recognize and catalyze other amino acids at the C-terminus, the C-terminal tryptophan was exchanged with the other two aromatic amino acids Phe or Tyr. The corresponding MS-271 derivatives, MS-271-W21F and MS-271-W21Y, were produced with relatively large amounts (Figure 3-2-3-1, Figure 3-2-3-2). Chiral analysis by a modified Marfey's method confirmed that MS-271-W21F and MS-271-W21Y contained D-Phe and D-Tyr, respectively (Figure 3-2-3-3). In contrast, a small amount of MS-271-W21V was detected by LC-MS analysis (Figure 3-2-3-1, Figure 3-2-3-2). These results suggested that an aromatic amino acid residue at the C-terminus of the precursor peptide is important for the epimerization reaction.

I next examined the production of derivatives with 20 amino acid residues by eliminating the C-terminal Trp (MS-271-ΔW21) or an internal Ile (MS271-ΔI17) in MslA. However, production of the expected peptides was not observed in either case.

Considering that the C-terminal four amino acid sequence of MS271-ΔI17 was identical to that of MS-271, the length of the core peptide is also likely crucial.

Furthermore, I investigated the MS-271 biosynthetic tolerance using sviceucin, a non-cognate lasso peptide with 20 amino acid residues, as a core peptide. The structure of sviceucin contains two disulfide bridges, Cys1–Cys13 and Cys7–Cys19, in the same manner as MS-271, though its amino acid sequence is quite different from that of MS-271 (Figure 3-2-3-1)[5]. At first, I prepared a chimeric precursor peptide gene by fusing the leader peptide of MslA to the core peptide of sviceucin. Consistent with the above speculation that the length of core peptides is important for production, the native 20-aa sviceucin was produced in a minute quantity (Figure 3-2-3-1). Next I sought to add a Trp to the C-terminus, rendering a 21-aa sviceucin derivative svi-W. Interestingly, a specific peak of svi-W could be detected by HPLC analysis despite a small amount. Svi-W was purified from the fermentation on fourty 20-mL agar plates. Chiral analysis revealed that the C-terminal Trp was D type (Figure 3-2-3-3). Next, I replaced the sequence of the C-terminal region in sviceucin core peptide with MS-271 to further generate a series of sviceucin derivatives svi-CFW, svi-VCFW, and svi-AIVCFW ("AIVCFW" is the tail portion of MS-271). Excitingly, HPLC analysis revealed successful production of these three MS-271/sviceucin hybrid lasso peptides (Figure 3-2-3-1, Figure 3-2-3-2) whose C-terminal Trp all had the D-configuration (Figure 3-2-3-3). These results indicated that the MS-271 biosynthetic enzymes, including MslH, exhibit broad substrate specificities toward the N-terminal region of core peptides while the C-terminal "CFW" sequence is important for substrate recognition. Additionally, when I deleted the C-terminal Trp (W) of svi-CFW, the production of the resulting 20-aa lasso peptides was abolished (data not shown), consisting with the above speculation that the length of core peptide is important for production.

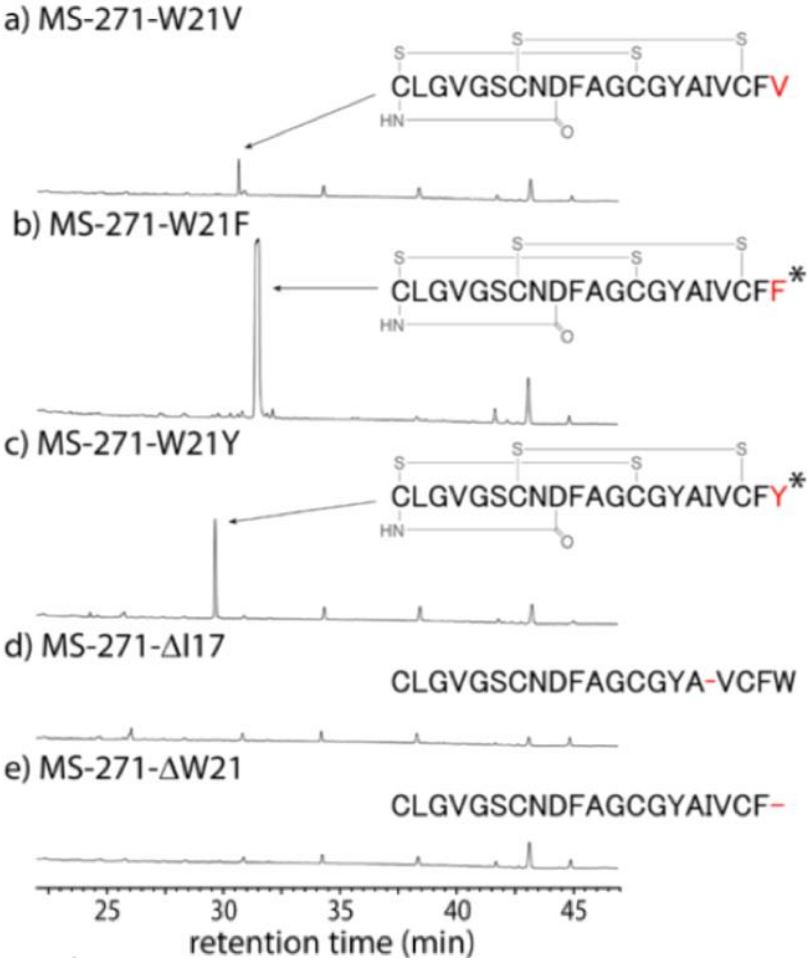


Figure 3-2-3-1 LC-MS analysis (UV absorbance monitored at 210 nm) of metabolites produced by heterologous production of the *msl* cluster with *mslA* mutant plasmids, a) pWHM3-msl-MS-271-W21V, b) pWHM3-msl-MS-271-W21F, c) pWHM3-msl-MS-271-W21Y, d) pWHM3-msl-MS-271-ΔI17, e) pWHM3-msl- MS-271-ΔW21, f) pWHM3-msl-svi, g) pWHM3-msl-svi-W, h) pWHM3-msl-svi- CFW, i) pWHM3-msl-svi-VCFW, and j) pWHM3-msl-svi-AIVCFW. The mutated amino acids in MS-271 derivative are shown in red. Sviceucin- and MS-271-derived sequences in the hybrid lasso peptides are shown in blue and red, respectively. The D-amino acid-containing products confirmed by chiral analysis are marked with an asterisk.

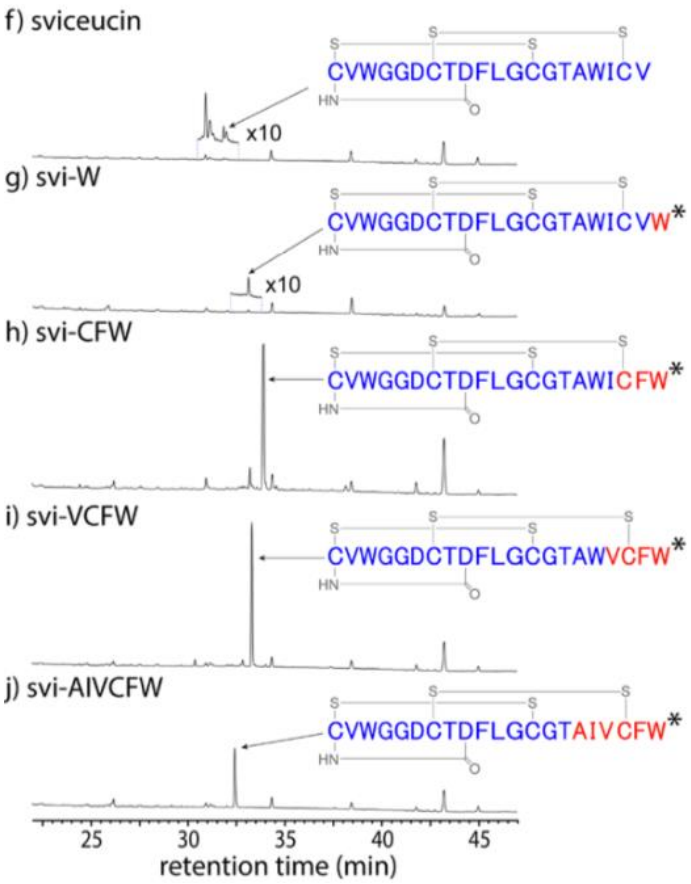


Figure 3-2-3-1 Continued.

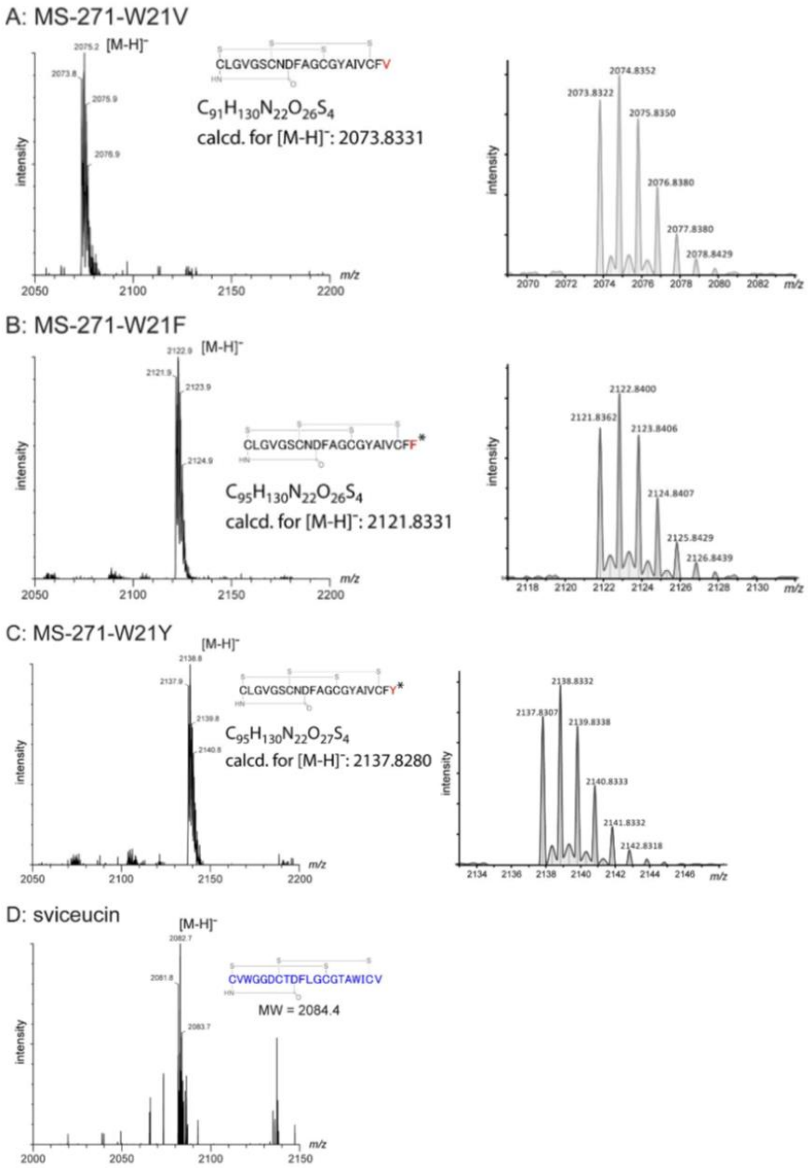


Figure 3-2-3-2 Mass spectra (ESI-negative, low-resolution [left] and high-resolution [right] data) of MS-271 derivatives: (A) MS-271-W21V, (B) MS-271-W21F, (C) MS-271-W21Y, (D) sviceucin, (E) svi-W, (F) svi-CFW, (G) svi-VCFW, and (H) svi-AIVCFW. D-Amino acid-containing products confirmed by chiral analysis are marked with an asterisk.

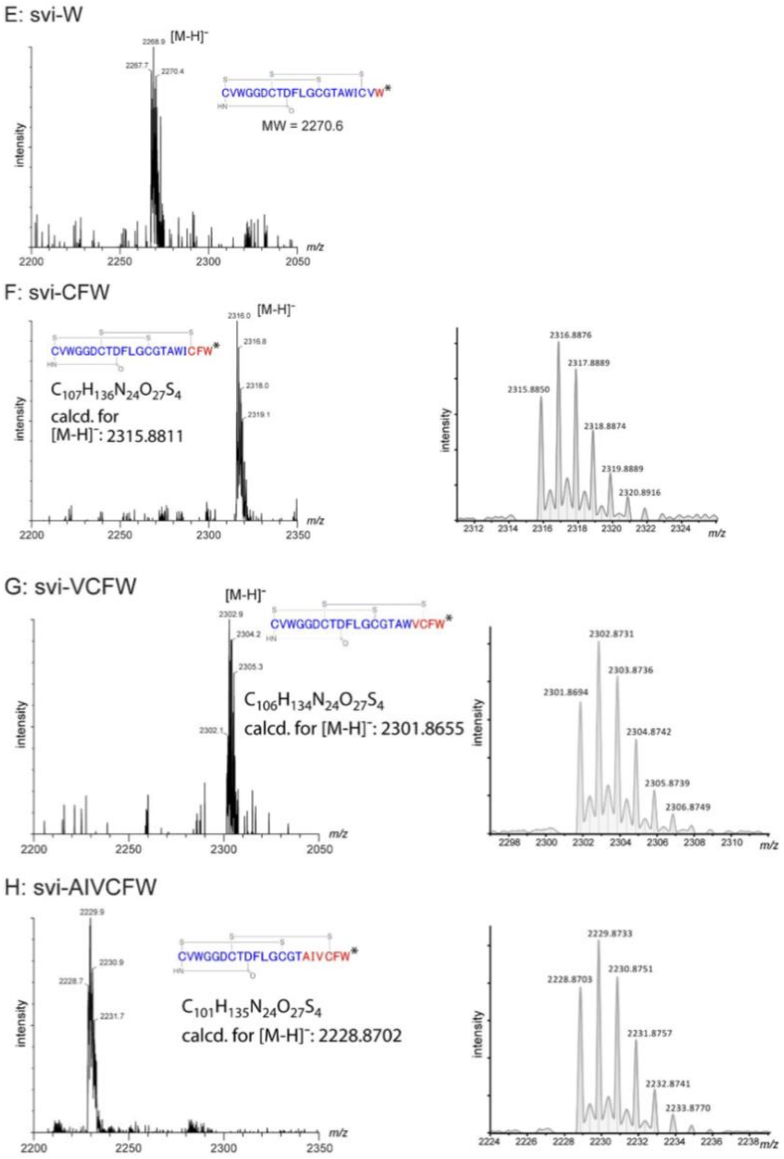


Figure 3-2-3-2 continued.

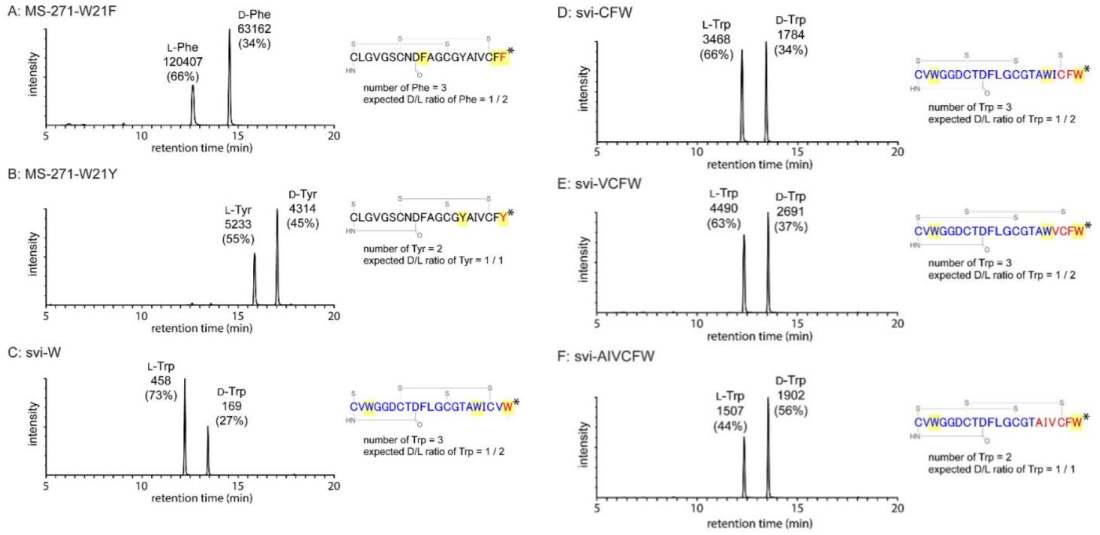


Figure 3-2-3-3 Chiral analysis of MS-271 derivatives. (A) Phe in MS-271-W21F, (B) Tyr in MS-271- W21Y, (C) Trp in svi-W, (D) Trp in svi-CFW, (E) Trp in svi-VCFW, and (F) Trp in svi-AIVCFW. The LC-MS chromatograms of L-FDLA derivatives of Phe, Tyr, and Trp were monitored at *m*/*z* 458, 474, and 497, respectively. Peak areas in UV traces at 340 nm are shown in the chromatograms. The amino acids subjected to chiral analysis are highlighted in yellow in the structures.

3.3. Discussion

In this chapter, I confirmed MslH as a novel peptide epimerase, responsible for the introduction of the C-terminal D-Trp by epimerization of the ribosomal peptide MslA in the biosynthesis of lasso peptide MS-271. Combined with the results of the *in vivo* and *in vitro* experiments, MslH was revealed to catalyze the epimerization of full-length MslA in a metal- and cofactor-independent epimerase. In addition, the epimerization reaction of MslA by MslH was accelerated in the presence of the precursor peptide recognition element MslB1, indicating a ternary protein complex formed by them. Prolonged incubation resulted in about 50% conversion of MslA to its epimer epi-MslA, indicating that MslH was capable of generating an equilibrium mixture of two epimers. Furthermore, I demonstrated that MslH, along with other biosynthetic enzymes in MS-271 biosynthesis, exhibits broad substrate specificity toward the N-terminal region of the core peptide and the heterologous production of MS-271/sviceucin hybrid lasso peptides with a C-terminal D-aa was successful.

As far as I know, MslH is the first epimerase that catalyzes epimerization at the C_α center adjacent to a carboxylic acid in a metal- and cofactor-independent manner. Our lab recently demonstrated the involvement of peptide epimerization in PGA biosynthesis by isotope tracer experiments.[6] Our characterization of MslH suggests that CapA (~52% identity) is an epimerase that introduces D-Glu residues into a homopolymer of L-Glu in PGA biosynthesis. Future functional characterization of CapA will provide further insight into MslH-type epimerases.

These results also provided insight into the timing of modification reactions in lasso peptide biosynthesis. The MslH-catalyzed epimerization occurs on the full-length MslA prior to proteolytic cleavage by MslB2 (Figure 3-3-1) and the leader peptide helps the recognition of MslH. This observation is common for all lasso peptide

modification enzymes characterized *in vitro*, including a methyltransferase (StspM)[7], kinases (ThcoK and SyanK)[8,9], and an iron/2-oxoglutarate-dependent hydroxylase (CanE)[10], suggesting the importance of leader peptide in the recognition of these modification enzymes. This feature is the same as many RiPP biosynthetic proteins, modifying their substrates by binding to the motif located on the leader peptide. It was reported that non-natural hybrid RiPP products were generated successfully by constructing chimeric leader peptides that can be recognized by different biosynthetic proteins from unrelated RiPP biosynthetic pathway[11]. Together with the fact that MslH exhibited broad substrate specificity toward the N-terminal region of the core peptide, this feature is potentially useful in peptide bioengineering.

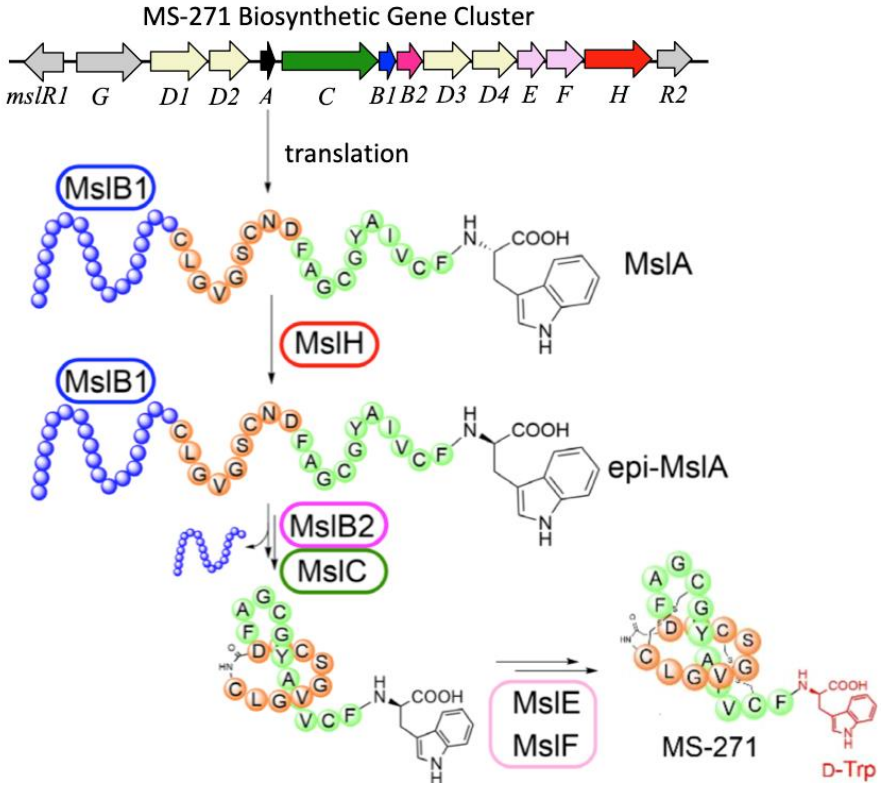


Figure 3-3-1 Biosynthesis of MS-271. MslB1 tightly binds to the leader peptide region (blue) in MslA and enhances the MslH reaction.

References

1. P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*, 2013, **30**, 108-160.
2. T. Ojima-Kato, S. Nagai and H. Nakano, N-terminal SKIK peptide tag markedly improves expression of difficult-to-express proteins in Escherichia coli and Saccharomyces cerevisiae. *J Biosci Bioeng*, 2017, **123**, 540-546.
3. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 2015, **10**, 845-858.
4. J. Yang and Y. Zhang, I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*, 2015, **43**, (W1), 174-181.
5. Y. Li, R. Ducasse, S. Zirah, A. Blond, C. Goulard, E. Lescop, C. Giraud, A. Hartke, E. Guittet, J. L. Pernodet and S. Rebuffat, Characterization of sviceucin from Streptomyces provides insight into enzyme exchangeablility and disulfide bond formation in lasso peptides. *ACS Chem Biol*, 2015, **10**, 2641-2649.
6. Y. Ogasawara, M. Shigematsu, S. Sato, H. Kato and T. Dairi, Involvement of peptide epimerization in poly-γ-glutamic acid biosynthesis. *Org Lett*, 2019, **21**, 3972-3975.
7. Y. Su, M. Han, X. Meng, Y. Feng, S. Luo, C. Yu, G. Zheng and S. Zhu, Discovery and characterization of a novel C-terminal peptide carboxyl methltransferase in a

lassomycin-like lasso peptide biosynthetic pathway. *Appl Microbiol Biotechnol*, 2019, **103**, 2649-2664.

8. S. Zhu, C. D. Fage, J. D. Hegemann, D. Yan and M. A. Marahiel, Dual substrate-controlled kinase activity leads to polyphosphorylated lasso peptides. *FEBS Lett*, 2016, **590**, 3323-3334.
9. S. Zhu, J. D. Hegemann, C. D. Fage, M. Zimmermann, X. Xie, U. Linne and M. A. Marahiel, Insights into the unique phosphorylation of the lasso peptide paeninodin. *J Biol Chem*, 2016, **291**, 13662-13678.
10. C. Zhang and M. R. Seyedsayamdost, CanE, an iron/2-oxoglutarate-dependent lasso peptide hydroxylase from Streptomyces canus. *ACS Chem Biol*, 2020, **15**, 890-894.
11. B. J. Burkhart, N. Kakkar, G. A. Hudson, W. A. van der Donk, D, A, Mitchell. Chimeric leader peptides for the generation of non-natural hybrid RiPP products. *ACS Cent Sci*. 2017, **3**, 629-638.

Chapter 4

Conclusion

MS-271 was isolated from *Streptomyces* sp. M-271 as a potent inhibitor of calmodulin-activated myosin light chain kinase in 1996. It is a lasso peptide natural product consisting of 21 amino acid residues with a D-Trp at its C-terminus. Lasso peptides belong to a subclass of RiPPs and feature an N-terminal macrocyclic ring with a C-terminal tail threaded through the ring. However, the mechanism of the D-Trp introduction remained unknown.

In this study, I investigated the mechanism of the D-Trp introduction in MS-271. In chapter 2, I identified a biosynthetic gene cluster of MS-271 (*msl*) by genome sequencing of the producer. By detailed analysis of the precursor peptide gene (*mslA*), I found that the D-Trp residue was introduced by epimerization of the ribosomal peptide as a post-translational modification. Besides the *mslA*, genes encoded for modification enzymes and a function-unknown enzyme (*mslH*) were present in the *msl* cluster but obvious epimerase gene was absent. By the heterologous expression of *msl* cluster in *Streptomyces lividans*, I showed that it contained all the necessary genes for MS-271 production including a peptide epimerase gene. Further gene knockout experiment of the function-unknown *mslH* showed that it was indispensable for MS-271 production. Based on these results, I hypothesized that *mslH* encoded a novel peptide epimerase responsible for the D-Trp introduction in MS-271 structure.

In chapter 3, I investigate the function of MslH by expressing *mslA* in the presence or absence of *mslH* in *E. coli*. The results showed that D-Trp was observed when *mslH* was present, indicating that MslH catalyzed the epimerization of the nascent precursor peptide (MslA) to generate its epimer. By *in vitro* experiments, I showed that MslH is a novel metal- and cofactor-independent epimerase. Furthermore, the leader peptide within MslA was shown to be indispensable for the activity of MslH. I also demonstrated broad substrate specificity of MslH toward the N-terminal region of the

core peptide by heterologous production of unnatural hybrid lasso peptides. MslH was characterized as the first epimerase that catalyzes epimerization at the Cα center adjacent to a carboxylic acid in a metal- and cofactor-independent manner in the biosynthesis of RiPPs natural product. Although further studies such as crystal structure analysis are necessary to understand the reaction mechanism, the unprecedented enzyme chemistry of MslH, along with its dependence on leader peptide and broad substrate specificity, renders MslH-type epimerases a potentially useful tool to design novel peptide products with improved biological activity and stability.

Experimental section

1. General

All chemicals were purchased from Sigma-Aldrich Japan, Tokyo Chemical Industry Co. Ltd. Japan, or FUJIFILM Wako Pure Chemical Corporation unless specified otherwise. Oligo nucleotides for PCR were obtained from Fasmac (Atsugi, Japan). Enzyme, molecular weight standards and kits for DNA manipulation were purchased from Takara Bio, Promega or New England Biolabs. The custom synthetic peptide (Msl-core 21: CLGVGSCNDFAGCGYAIVCFW) was purchased from Sigma-Aldrich Japan. PCR reactions were carried out using a GeneAmp PCR System 9700 or SimpliAmp thermal cyclers with Tks Gflex DNA polymerase (Takara Bio. Shiga, Japan). Sanger sequences were performed by a commercial company (Fasmac). General genetic manipulation of *E. coli* and *Streptomyces lividans* TK23 were performed according to standard protocols[1]. Plasmids in *E. coli* were maintained using appropriate antibiotics with the following concentrations; ampicillin (100 μg/mL), kanamycin (25 μg/mL), chloramphenicol (30 μg/mL), and/or streptomycin (20 μg/mL). High resolution MS and MS/MS data were recorded using an LTQ-orbitrap XL mass spectrometer (Thermo Scientific). An expression vector pTYM18ep was constructed from integrating vector pTYM18[2] by introducing a DNA fragment containing a constitutive promoter *erm*E*p and a multiple cloning site of pTYM19ep[3]. The MS-271-producing bacterium, *Streptomyces* sp. M-271, was kindly provided by Kyowa Hakko Bio Co., Ltd. *Streptomyces olivochromogenes* NBRC 3561, *Streptomyces nodosus* NBRC 12895, *S. griseorubiginosus* NBRC 12899 (=*Streptomyces phaeopurpureus* DSM 40125) and *Streptomyces diastatochromogenes* NBRC 13389 were obtained from NITE Biological Resource Center, the National Institute of Technology and Evaluation (Tokyo, Japan).

2. Production of MS-271 and its derivatives.

To produce MS-271, *Streptomyces* sp. M-271, *olivochromogenes*, *S. nodosus*, *S. griseorubiginosus* or *S. diastatochromogenes* was first grown in 10 ml of seed medium (1% glucose, 1% soluble starch, 0.5% BD Bacto tryptone, 0.5% BD Bacto yeast extract, 0.3% Ehrlich's fish extract (Kyokuto, Tokyo, Japan), 0.5% $CaCO_3$, pH 7.2 by NaOH) for 2 days at 30°C on a rotary shaker (200 rpm). A 1-mL portion of the culture was inoculated on an agar plate containing 25 mL of production medium (4% soluble starch,

1% soybean meal, 0.5% corn steep liquor (Sigma-Aldrich), 0.5% dry yeast, 0.05% KH_2PO_4, 0.05% $Mg_3(PO_4)_2 \cdot 8H_2O$, 0.001% $ZnSO_4 \cdot 7H_2O$, 0.0001% $CoCl_2 \cdot 6H_2O$, 0.0001% $NiSO_4$, 1.5% agar, pH 7.0 by NaOH) and the culture was incubated at 30°C for 5 days. The cells on the agar plate were collected by spatula and transferred into a vial containing 10 mL MeOH. After soaking the cells in MeOH at 4°C for 16 h, the MeOH extract was concentrated and re-dissolved in 80% aqueous MeOH (1 mL) for analysis. LC-MS analysis of MS-271 production from *Streptomyces* sp. M-271 was performed with a Shimadzu Prominence HPLC system equipped with a Bruker amazon SL Ion Trap and a PDA Detector. The analytical conditions were as follows: Mightysil RP-18GP Aqua column (150 × 2.1 mm); column temperature, 40°C; detection, ESI-positive and ESI-negative mode, and PDA (190–350 nm); elution, A: water with 0.1% formic acid (FA), B: acetonitrile with 0.1% FA, 1% solvent B for 0–7 min and a linear gradient to 95% solvent B for 7–30 min; flow rate, 0.2 mL/min. HPLC analysis of the metabolites from *olivochromogenes*, *S. nodosus*, *S. griseorubiginosus* or *S. diastatochromogenes* was performed with a Shimadzu Prominence HPLC system equipped with a photo diode array (PDA) Detector. The analytical conditions were as follows: column, Mightysil RP-18GP Aqua column (250 × 4.6 mm, 5 mm, Kanto Chemical); column temperature, 40°C; detection, PDA (190–350 nm); mobile phase, A: water with 0.05% trifluoroacetic acid (TFA), B: acetonitrile with 0.05% TFA, 100% solvent A for 0–10 min, a linear gradient to 80% solvent B for 10–35 min, and 80% solvent B for 35–45 min; flow rate, 1.0 mL/min. Heterologous production of MS-271 derivatives and MS-271/sviceucin hybrid lasso peptides were performed using the plasmids listed in Table 2. The resulting metabolites were analyzed by LC-MS (Waters ACQUITY UPLC system equipped with a SQ Detector2) under the following conditions: column: Mightysil RP-18GP Aqua column (150 × 2.0 mm, 3 mm); column temperature, 40°C; detection, ESI-negative mode and PDA; mobile phase, A: water with 0.05% TFA, B: acetonitrile with 0.05% TFA, 5% solvent B for 0–10 min and a linear gradient to 85% solvent B for an additional 30 min; flow rate, 0.2 mL/min. High resolution (HR)- MS was recorded on Bruker microTOF-HS at the Open Facility, Global Facility Center, Creative Research Institution, Hokkaido University.

3. Identification of the MS-271 biosynthetic gene cluster.

Draft genome sequencing of *Streptomyces* sp. M-271 was carried out by a commercial company (Hokkaido System Science, Hokkaido, Japan) using an Illumina HiSeq instrument (Illumina, San Diego, CA, USA). A genomic DNA library (350 bp insert) was constructed with the TruSeq Nano DNA LT Sample Prep Kit (Illumina). Paired-end data (2 × 100 bp) were assembled using Velvet (version 1.2.08)[4]. Gene prediction and annotation was carried out using MiGAP (Microbial Genome Annotation Pipeline). Manual assignments of open reading frames (ORFs) were made with the assistance of FramePlot and the SnapGene software. The nucleotide sequence of the MS-271 biosynthetic gene cluster in *Streptomyces* sp. M-271 has been deposited in the DDBJ/GenBank database (accession no. LC381634).

4. gene knockout experiment of *mslH* in the wild-type *Streptomyces* sp. M-271

The genomic DNA of *Streptomyces* sp. M271 was used as the template for PCR. The upstream region (~2000bp) of *mslH* were amplified with two primers, mslH-up_F(PstI) and mslH-up_R(BamHI). Downstream region (~2000bp) of *mslH* were amplified with two primers, mslH-down_F(HindIII) and mslH-down_R(PstI). The amplified fragments of upstream region of *mslH* were digested with *Pst*I and *Bam*HI and inserted into the same sites of pWHM3 to give pWHM3-mslH-up. The amplified fragments of downstream region of *mslH* were digested with *Pst*I and *Hin*dIII and inserted into the same sites of pWHM3-mslH-up to give pWHM3-mslH-ko1. The resulting plasmid was introduced into *Streptomyces* sp. M271 protoplast. The transformants were grown in SKII (50 ml) supplied with thiostrepton (15 μg/ml thiostrepton) and glycine (0.5 %) to make protoplast. The protoplast was diluted and cultivated on R5 plates for six days. Then the colonies were transfer to ATCC5 plates and selected for colony PCR with three pairs of primers (mslH-KO-F-1 and mslH-KO-R-1, mslH-KO-F-2 and mslH-KO-R-2, mslH-KO-F-3 and mslH-KO-R-3) to confirm *mslH* was knocked out successfully. Fermentation of three Δ*mslH* mutants along with the wild type was carried out for 5 days at 30°C on an agar plate containing 25 mL production medium. The metabolites accumulated were analyzed by HPLC and LC-ESI-MS using the same conditions as wild type.

5. Heterologous expression of the *msl* cluster.

A DNA fragment carrying the putative *msl* cluster was amplified by PCR with two primers, msl-cluster_F(XbaI) and msl-cluster_R(HindIII). The amplified fragments were digested with *Hin*dIII and *Xba*I and inserted into the same sites of pWHM3[5] to give pWHM3-msl. The resulting plasmid was introduced into *Streptomyces lividans* TK23, and the transformants were grown in seed medium (10 mL) supplied with thiostrepton (5 μg/ml). Fermentation was carried out for 5 days at 30°C on an agar plate containing 25 mL production medium (15 μg/ml thiostrepton). The metabolites accumulated were analyzed by LC-ESI-MS as described above.

For the overexpression of regulatory genes, the coding regions of *mslR1*, *mslG* and *mslR2* were amplified by PCR using the primer pairs mslR1-N(NdeI)/mslR1-C(XbaI), mslG-N(NdeI)/mslG-C(XbaI), and mslR2-N(NdeI)/mslR2-C(XbaI), respectively. Each PCR product was digested with *Nde*I/*Xba*I and then ligated into the pTYM18ep vector. The recombinant plasmids pTYM18ep-mslR1, pTYM18ep-mslG, and pTYM18ep-mslR2 were introduced into *S. lividans* TK23. After confirmation of correct chromosomal integration of the plasmids by PCR, the resulting heterologous hosts, *S. lividans*:*mslR1*, *:mslG*, and *:mslR2*, were transformed with pWHM3-msl, and the metabolites of the transformants were analyzed as described above except that kanamycin was also added to the seed medium (5 μg/mL) and production medium (15 μg/mL).

6. Chiral analysis of the C-terminal Trp in MS-271.

The chirality of the C-terminal Trp residue in heterologously produced MS-271 was analyzed using Marfey's method. The lasso peptide product obtained by heterologous expression was first hydrolyzed at 115°C for 22 h with 3 M 2-mercaptoethanesulfonic acid (MESA). After neutralization with 1 M aqueous NaOH, tryptophan was purified by HPLC and dissolved in 25 μL of water. To the sample solution was added 10 μL of 1 M sodium bicarbonate and 50 μL of 1% 1-fluoro-2,4-dinitrophenyl-5-L-leucinamide (L-FDLA) in acetone. After incubation for 2 h at 37°C, the reaction was quenched by the addition of 10 μL of 1 M HCl, and then diluted with 300 μL of acetonitrile. The sample containing L-FDLA derivatives was analyzed by LC-MS under the following conditions: Waters ACQUITY UPLC system equipped with a SQ Detector2; InertSustain C18 column (150 × 2.1 mm, GL Science Inc., Tokyo,

Japan); column temperature, 40°C; detection, ESI-negative mode m/z = 497 and PDA; mobile phase: water with 0.05% TFA, acetonitrile with 0.05% TFA. 5% solvent B for 0–10 min and a linear gradient to 95% solvent B for an additional 20 min; flow rate, 0.2 mL/min.

7. Chiral analysis of the MS-271/sviceucin derivatives and peptide product from *in vivo* and *in vitro* experiments

The chirality of the amino acid of interest (Trp, Tyr, or Phe) in peptides was analyzed using Marfey's method. The purified peptide product was first hydrolyzed at 110°C for 10 h with 3 M 2-mercaptoethanesulfonic acid (MESA). The reaction conditions were optimized to inhibit the spontaneous isomerization. After neutralization with 1 M aqueous NaOH, the amino acid of interest was purified by HPLC using the conditions described above and dissolved in 15 μL of water. To the sample solution was added 20 μL of 0.5 M sodium bicarbonate and 50 μL of 1% 1-fluoro-2,4- dinitrophenyl-5-L-leucinamide (L-FDLA) in acetone. After incubation for 2 h at 37°C, the reaction was quenched by the addition of 10 μL of 1 M HCl, and then diluted with 300 μL of acetonitrile. The sample containing L-FDLA derivatives was analyzed by Waters ACQUITY UPLC system equipped with a SQ Detector2 under the following conditions: column: Mightysil RP-18GP Aqua column (150 × 2.0 mm, 3 μm); column temperature, 40°C; detection, ESI-negative mode, single ion monitoring and PDA; mobile phase, A: water with 0.05% TFA, B: acetonitrile with 0.05% TFA, 30% solvent B for 0–2 min and a linear gradient to 95% solvent B for an additional 18 min; flow rate, 0.2 mL/min.

8. Gene deletion of the biosynthetic genes.

Plasmid constructs lacking one of *mslB1*, *-B2*, *-C*, *-E*, *-F*, and *-H* were generated from pWHM3-msl using a RED/ET recombination kit (Gene Bridges, Heidelberg, Germany). To generate a Δ*mslB1* in-frame deletion mutant, a PCR product containing the FRT-PGK-gb2-neo-FRT cassette and terminal homology arms, which had sequence identical to the N- and C-terminal regions of the *mslB1* gene, was prepared using the primer pair mslB1-KO-Fw/mslB1-KO-Rv (Table 1) and FRT-PGK-gb2-neo-FRT as the template. An *Nsi*I site was included in both primers for later removal of the FRT-PGK-gb2-neo-FRT cassette. The PCR product was introduced into *E. coli* DH5α

harboring the plasmids pRedET(tet) and pWHM3-msl, and homologous recombination was induced with L-arabinose. The FRT-PGK-gb2-neo-FRT cassette in the resulting plasmid was removed by digestion with *Nsi*I, and subsequent self-ligation to produce the plasmid pWHM3-msl-ΔmslB1. Likewise, pWHM3-msl-ΔmslB2, -ΔmslC, -ΔmslE, -ΔmslF, and -ΔmslH were constructed using appropriate primers (listed in Table 2). To prepare pWHM3-msl-ΔmslH, the FRT-PGK-gb2-neo-FRT cassette was removed by digestion with *Mfe*I and self-ligation.

9. protein expression in *E. coli* BL21(DE3) using genomic DNA of *Streptomyces* sp. M-271

Plasmid construction

pRSF-mslA-WT: The gene encoding MslA was amplified by PCR from the genomic DNA of *Streptomyces* sp. M-271 with the primers, mslA-WT-N(BamHI) and mslA-WT-C(HindIII). The PCR product was cloned into the *Bam*HI-*Hin*dIII site of the pRSF-Duet-1 vector to construct the plamid, pRSF-mslA-WT.

pET-mslB1-mslB2-WT: A DNA fragment encoding the *mslB1* gene was prepared by PCR from the genomic DNA of *Streptomyces* sp. M-271 with the primers, mslB1-WT-N(NcoI) and mslB1-WT-C(HindIII), and was cloned into the *Nco*I-*Hin*dIII site of the pET-Duet-1 vector to obtain pET-mslB1-WT. Then a DNA fragment encoding the *mslB2* gene was prepared by PCR with the primers, mslB2-WT-N(NdeI) and mslB2-WT-C(XhoI), and was cloned into the *Nde*I-*Xho*I site of the pET-mslB1-WT plasmid to construct pET-mslB1-mslB2-WT.

pCDF-mslE-mslF-WT: A DNA fragment encoding the *mslE* gene was prepared by PCR from the genomic DNA of *Streptomyces* sp. M-271 with the primers, mslE-WT-N(NcoI) and mslE-WT-C(HindIII), and was cloned into the *Nco*I-*Hin*dIII site of the pCDF-Duet-1 vector to obtain pCDF-mslE-WT. Then a DNA fragment encoding the *mslF* gene was prepared by PCR with the primers, mslF-WT-N(NdeI) and mslF-WT-C(XhoI), and was cloned into the *Nde*I-*Xho*I site of the pCDF-mslE-WT plasmid to construct pET-mslE-mslF-WT.

pACYC-mslC-mslH-WT: A DNA fragment encoding the *mslC* gene was prepared by PCR from the genomic DNA of *Streptomyces* sp. M-271 with the primers, mslC-WT-N(NcoI) and mslC-WT-C(HindIII), and was cloned into the *Nco*I-*Hin*dIII site of the pACYC-Duet-1 vector to obtain pACYC-mslC-WT. Then a DNA fragment encoding the *mslH* gene was prepared by PCR with the primers, mslH-WT-N(NdeI) and mslH-WT-C(XhoI), and was cloned into the *Nde*I-*Xho*I site of the pACYC-mslH-WT plasmid to construct pACYC-mslC-mslH-WT.

***in vivo* protein expression**

E. coli BL21(DE3) transformants harboring plasmids were grown at 200 rpm at 37 °C in LB medium (50 mL in a 250 mL Erlenmeyer flask) supplied with appropriate antibiotics and were induced by adding 0.5 mM isopropyl β-D-1- thiogalactopyranoside (IPTG) when the optical density at 600 nm reached about 0.5. The cultivation at 200 rpm was continued for an additional 16 h at 16°C. The cells were resuspended in a 3.5 mL wash buffer I (50 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, pH 8.0) and disrupted by sonication using an ultrasonic disruptor (TOMY, UD-200). The resulting soluble and insoluble fractions were analyzed by SDS-PAGE.

10. *in vivo* and in vitro assay of MslH using genomic DNA of *S. griseorubiginosus* NBRC 12899

Plasmid construction

pET24-mslA: The gene encoding MslA was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, mslA-N(NdeI) and mslA-C(HindIII). The PCR product was cloned into the *Nde*I–*Hin*dIII site of the pET24b vector to construct the plasmid, pET24-mslA.

pET24-SKIK-mslA: The gene encoding MslA was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, SKIK-mslA-N(NdeI) and SKIK-mslA-C(HindIII). The PCR product was cloned into the *Nde*I–*Hin*dIII site of the pET24b vector to construct the plasmid, pET24-SKIK-mslA.

pRSF-SKIK-mslA: The gene encoding MslA was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, SKIK-mslA-pRSF-

N(NdeI) and SKIK-mslA-pRSF-C(XhoI). The PCR product was cloned into the *Nde*I–*Xho*I site of the pRSF-Duet-1 vector to construct the plasmid, pRSF-SKIK-mslA.

pET24-SKIK-His-mslA: The gene encoding MslA was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, SKIK-His-mslA-N(NdeI) and SKIK-His-mslA-C(HindIII). A DNA sequence encoding the N-terminal "MSKIKHHHHHH" peptide was designed in the primer SKIK-His-mslA-N(NdeI). The PCR product was cloned into the *Nde*I–*Hin*dIII site of the pET24b vector to construct the plasmid, pET24-SKIK-His-mslA. The deduced amino acid sequence of the peptide product is "MSKIKHHHHHHSAVYEPPMLQEVGDFDELTKCLGVGSCNDFAGCGYAIVCFW".

pRSF-SKIK-His-mslA: The gene encoding MslA was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, SKIK-His-mslA-N(NdeI)-2 and SKIK-His-mslA-C(XhoI). A DNA sequence encoding the N-terminal "MSKIKHHHHHH" peptide was designed in the primer SKIK-His-mslA-N(NdeI)-2. The PCR product was cloned into the *Nde*I–*Xho*I site of the pRSF-Duet-1 vector to construct the plasmid, pRSF-SKIK-His-mslA. The deduced amino acid sequence of the peptide product is "MSKIKHHHHHHSAVYEPPMLQEVGDFDELTKCLGVGSCNDFAGCGYAIVCFW".

pCDF-His-mslH: The gene encoding MslH was amplified by PCR from the genomic DNA of *S. griseorubiginosus* NBRC 12899 with the primers, His-mslH-N(NdeI) and His-mslH-C(HindIII). The PCR product was cloned into the *Nde*I–*Hin*dIII site of the pET28b vector to construct the plasmid, pET28-NCHis-mslH. The gene encoding MslH was amplified by PCR using the pET28-NCHis-mslH as the template with the primers, His-mslH-N(EcoRI) and His-mslH-C(PstI). The PCR product was cloned into the *Eco*RI–*Pst*I site of the pCDF-Duet-1 vector to construct the plasmid, pCDF-His-mslH.

pACYC-His-mslH: After pCDF-His-mslH was digested with the restriction enzymes *Eco*RI and *Pst*I, the fragment containing *mslH* was ligated to the *Eco*RI–*Pst*I site of the pACYC-Duet-1 vector to construct pACYC-His-mslH.

pET21-SKIK-His-mslA: After pET24-SKIK-His-mslA was digested with the restriction enzymes *Nde*I and *Hin*dIII, the fragment containing *mslA* was ligated to the *Nde*I–*Hin*dIII site of the pET21a vector to construct pET21-SKIK-His-mslA.

pET28-His-mslH: The primer pair mslH-pET-N(NdeI) and mslH-pET-C(EcoRI) was used to amplify the *mslH* gene from the genomic DNA of *S. griseorubiginosus*. The resulting PCR product was cloned into the *Nde*I–*Eco*RI site of the pET28b vector.

pSTV-His-mslH: A DNA fragment containing the *mslH* gene was prepared with PCR using the primers, mslH-pCDFDuet-N(EcoRI) and mslH-pCDFDuet-C(HindIII), and was first cloned into the *Eco*RI–*Hin*dIII site of the pCDF-Duet-1 vector to fuse a 6×His tag sequence at the N-terminus of the *mslH* gene. Second PCR reaction using the resulting plasmid and the primer pair, mslH-pSTV28N-N(NdeI) and mslH-pCDFDuet-C(HindIII), was then carried out and the resulting PCR product was cloned into the pSTV28N vector[6] using the restriction enzymes *Nde*I and *Hin*dIII to construct pSTV-His-mslH.

pCDF-His-mslB1: A DNA fragment containing the *mslB1* gene of *S. griseorubiginosus* was amplified using the primers mslB1-N(BamHI) and mslB1-C(HindIII), and was cloned into the *Bam*HI–*Hin*dIII site of the pCDF-Duet-1 vector to generate pCDF- His-mslB1.

pCDF-mslB1: A DNA fragment containing the *mslB1* gene of *S. griseorubiginosus* was amplified using the primers mslB1-N(NcoI) and mslB1-C(HindIII), and was cloned into the *Nco*I– *Hin*dIII site of the pCDF-Duet-1 vector to generate pCDF-mslB1.

pCola-MBP-His-mslC: A DNA fragment encoding the *mslC* gene was prepared by PCR from the genomic DNA of *S. griseorubiginosus* using the primers mslC-pCola-N(NdeI) and mslC-pCola-C(MfeI), and was cloned into the *Nde*I–*Mfe*I site of the pCola-Duet-1 vector to obtain pCola-His-mslC. A DNA sequence for a 6×His tag was designed in the primer MslC-pCola-N(NdeI). Because expression of soluble MslC

using pCola-His-mslC was unsuccessful, a DNA fragment containing the maltose binding protein (MBP) was inserted into the *Nde*I site of pCola-His-mslC by in-fusion cloning so that the MBP, His-tag, and MslC sequences were expressed in a single open reading frame. The insert DNA fragment was prepared by PCR from the pMAL-c5X vector using the primer pair mal-mslC-F and mal-mslC-R, and pCola-His-msl digested with *Nde*I was used as the vector DNA fragment.

pCola-MBP-His-mslB2: The primer pair mslB2-pCola-N(BamHI) and mslB2-pCola-C(EcoRI) was used to amplify the *mslB2* gene of *S. griseorubiginosus*. The resulting PCR product was cloned into the *Bam*HI–*Eco*RI site of the pCola-Duet-1 vector to obtain pCola-His-mslB2. The DNA fragment containing MBP was then inserted into the *Nco*I site of pCola-His-mslB2 by in-fusion cloning so that the MBP, His-tag, and MslC sequences are expressed in a single open reading frame. The insert DNA fragment was prepared by PCR from the pMAL-c5X vector using the primer pair mal-mslB2-F and mal-mslB2-R, and pCola-His-mslB2 digested with *Nco*I was used as the vector DNA fragment.

pWHM3-msl variants: To generate pWHM3-msl mutant plasmids harboring the desired core peptide sequence, *Spe*I–*Sph*I fragments (3.4 kbp) that included the precursor peptide coding region were prepared by overlap extension PCR and each fragment was replaced with the wild type *Spe*I–*Sph*I (3.4 kbp) fragment in pWHM3-msl. The primer pairs listed in Table 2 were used for the first PCR and the primers msl-SpeI-SphI-F and msl-SpeI-SphI-R were used for the second PCR of the overlap extension PCR.

In vivo peptide epimerization assay

E. coli BL21(DE3) transformants harboring various combinations of plasmids (pET21-SKIK-His-mslA, pSTV-His-mslH, pCDF-His- mslB1, pCola-MBP-His-mslC, and pCola-MBP-His-mslB2) were grown at 200 rpm at 37°C in LB medium (50 mL in a 250 mL Erlenmeyer flask) supplied with appropriate antibiotics and were induced by adding 0.5 mM isopropyl β-D-1- thiogalactopyranoside (IPTG) when the optical density at 600 nm reached about 1. The cultivation at 200 rpm was continued for an additional 16 h at 16°C. The cells were resuspended in a 3.5 mL wash buffer I (50 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, pH 8.0) and disrupted by

sonication using an ultrasonic disruptor (TOMY, UD-200). After centrifugation at 25,000×g for 30 min, the MslA was purified from either the supernatant or the precipitate as follows. When MslA was produced in an insoluble form (MslA alone and MslA+H), the precipitate was resuspended in 900 μL of buffer I (50 mM sodium phosphate, 300 mM NaCl, pH 8) and an aliquot (300 μL) was separated by Tricine-SDS-PAGE (16%T, 3%C polyacrylamide gel)[7]. After reverse staining using EzStain reverse (ATTO) to visualize proteins, a gel slice containing MslA was excised from the gel and MslA was extracted from the gel with an Attoprep filter unit (ATTO). The buffer was exchanged to 10 mM Tris-HCl (pH 8.0) by ultrafiltration (Amicon Ultracel-3, 0.5 mL, Merck). The purified MslA (in ca. 100 μL buffer) was lyophilized and used for chiral analysis. When MslA was produced in a soluble form (MslA+B1, MslA+H+B1, MslA+H+B1+B2, and MslA+H+B1+B2), the supernatant was subjected to Ni-NTA affinity chromatography. The fractions containing MslA were collected and concentrated to 300 μL by Amicon (Ultracel-3, 0.5 mL) and further purified by Tricine-SDS-PAGE as described above.

Preparation of MslA for *in vitro* assay

An *E. coli* BL21(DE3) transformant harboring pET21-SKIK-His- mslA and pCDF-mslB1 was grown in LB medium as described above. After harvest, the cells were resuspended in a 3.5 mL wash buffer I and disrupted by sonication. After centrifugation at 25,000×g for 30 min, the supernatant was loaded onto a column containing Ni-NTA agarose resin. The column was washed with wash buffer I followed by wash buffer II (50 mM sodium phosphate, 8 M urea, 300 mM NaCl, 25 mM imidazole, pH 8.0). MslA was then eluted from the column using elution buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). The buffer was exchanged to buffer II (100 mM Tris-HCl, 300 mM NaCl, pH 8.0) and was reduced by Amicon ultrafiltration units (Ultracel-3, 0.5 mL).

Protein purifications for *in vitro* assay

E. coli BL21(DE3) transformants harboring pET28-His-mslH and pCDF-His-mslB1 were used to produce MslH and MslB1, respectively. After cell lysate preparation as described above, the protein was purified on Ni-NTA resin using wash buffer I and elution buffer. Finally, the protein was prepared in buffer II using Amicon filter units (Ultracel-3 for MslB1 and Ultracel-30 for MslH) for *in vitro* reactions.

In vitro assay of MslH

A reaction mixture (300 μL) containing 220 μM MslA (SKIK- and His-tagged), 220 μM MslB1, and 5 μM MslH in 100 mM Tris-HCl (pH 8.0) was incubated at 30°C for 16 hours. Control reactions omitting MslB1 and/or MslH were also performed. The reactions were terminated by incubating with an equal volume of 2×SDS-PAGE sample buffer (10% 2-mercaptoethanol, 4% SDS, 10 % sucrose, and 0.01 % bromophenol blue in 0.125 M Tris-HCl, pH 6.8) at 100°C for 10 min. MslA was purified by Tricine-SDS-PAGE and subjected to chiral analysis as described above. Adding 300 mM NaCl to the reaction buffer 100 mM Tris-HCl (pH 8.0) had no effect on the results.

In vitro assay of MslH under reduced conditions (DTT added)

A reaction mixture (300 μL) containing 470 μM MslA (SKIK- and His-tagged), 500 μM MslB1, 5 μM MslH, and 10 mM DTT in buffer II was incubated at 30°C. Control reactions omitting MslB1 and/or MslH were also performed. The termination of reactions and the purification and analysis of MslA was mentioned above. Reactions with the synthetic MslA core peptide contained 900 μM MslA-core, 600 μM MslB1, 10 μM MslH, and 10 mM DTT in buffer II (total 150 μL) and were incubated at 30°C for 1 h. Control reactions omitting MslB1 and/or MslH were also performed. The resulting MslA-core peptides were purified by Tricine-SDS-PAGE (16%T, 6%C polyacrylamide gel) and the extracted MslA was concentrated by lyophilization after the removal of small molecules using a Tube-O-Dialyzer Medi 1 k MWCO dialysis unit (Geno Technology).

Metal requirements of MslH

A reaction mixture (230 μL) containing 430 μM MslA, 20 μM MslH, and 5 mM metal chloride ($CoCl_2$, $MnCl_2$, $FeCl_2$, $CaCl_2$, $ZnCl_2$, or $MgCl_2$) or 5 mM EDTA in buffer (100 mM Tris-HCl, pH 8.0) was incubated at 30°C. After 2 h, MslA was purified by Tricine-SDS-PAGE and subjected to chiral analysis as described above.

Table 1. Primers list in this study. The introduced restriction sites were underlined. The SKIK and His-tag sequences introduced are shown with italic and bold letters, respectively.

Name	sequences (5' to 3')
msl-cluster_F(XbaI)	ACATATCTAGAGAGGCCGGAAGTCGGTCCATCCCGAAG
msl-cluster_R(HindIII)	ATCTAAAGCTTGTGGAGTCGACGAGGACCCGGAAGCTCC
mslR1-N(NdeI)	ACTTATCTCATATGCGGATCAATCTCTTCATCCTC
mslR1-C(XbaI)	GGATTCTAGATTCAGGCGTGTTGGGGCAGTACG
mslG-N(NdeI)	CATCACCACATATGGGAGATGTCTCGCGCCGACG
mslG-C(XbaI)	GGATTCTAGATGGCACTAGTACGGCAGTGTCG
mslR2-N(NdeI)	CATCACCACATATGGCCGAGTCGGTCCGGGTAC
mslR2-C(XbaI)	GGATTCTAGACGCGATCACGTCGGCATTCC
mslC-KO-Fw	TGCGGGCGACGCGGCGCGTCGACCACGCGTCGGGGCGGCCCTGGATCGTCATGCATAAATTAACCCTCA CTAAAGGGCGG
mslC-KO-Rv	CGTAGGCGTGTTCCGGGCGGGTGGCGGGGGGTACGTCTCCGGATGCGTAATGCATGTAATACGACTCAC TATAGGGCTCG
mslB1-KO-Fw	TGACCCTCGCCCGCGACGTCACCCTCACCGTCGTCGACTCCGGGGCCGTGATGCATAAATTAACCCTCACT AAAGGGCGG
mslB1-KO-Rv	CACGAGCTTGGCCGCGCTGAGCGCGTCGATGAGGGCCTGGACGTCCTGAATGCATGTAATACGACTCAC TATAGGGCTCG
mslB2-KO-Fw	CCGCCGTGGCCGAACAGGCCCCGCGGCTGCCCTGGTACCGGCAGCTCGCCATGCATAAATTAACCCTCAC TAAAGGGCGG

mslB2-KO-Rv	GGAGGGGCGGAGGGGACGGTCGGAGGTGCTCCG GCGTCCGTCCGGGCGAATGCATGTAATACGACTC A CTATAGGGCTCG
mslE-KO-Fw	GAGGACAGGAGACGGTGGACGCGAGCACCACGG ATGTGACCACGACGGCCATGCATAAATTAACCCT C ACTAAAGGGCGG
mslE-KO-Rv	GACCCAGCCGTCCAGCGCGTATCGGGTGCGCGGA TTCCGGATCAGGAGAATGCATGTAATACGACTCA C TATAGGGCTCG
mslF-KO-Fw	CGAGCACACAGACAACGGACGCCACGGTGCGGG AGATGGTGCACCGACGGATGCATAAATTAACCCT C ACTAAAGGGCGG
mslF-KO-Rv	GACCGCCCGGGTCAGCGCGCCGCCCGGATCGGAG CCGGTGACGTCGACAATGCATGTAATACGACTCA C TATAGGGCTCG
mslH-KO-Fw	CGGCGGTGACCCGGCTGACCGTGGCTCTGTCCGG CGACTGCATGGCGACACAATTGAAATTAACCCTC A CTAAAGGGCGG
mslH-KO-Rv	TCTTCACGCCACGTCCAGAAGCAGCTCCGCCGCG CCGCCGGCCCCCGCCAATTGTGTAATACGACTCA CT ATAGGGCTCG
SKIK-His-mslA-N(NdeI)	CGCATATG*TCTAAAATAAAA***CACCACCACCACCAC CAC**TCCGCTGTTTACGAGCCG
SKIK-His-mslA-C(HindIII)	ACAGAAGCTTTCACCAGAAGCAGACGATCG
mslH-pET-N(NdeI)	ATAGTAACATATGACCCGGCTGACGGTGGCCCTG TC
mslH-pET-C(EcoRI)	AAGAATTCTCACGCCACGTCGAGCGCCAGCTC

mslH-pCDFDuet-N(EcoRI)	TAGTAAGAATTCCATGACCCGGCTGACCGTGG
mslH-pCDFDuet-C(HindIII)	ATGTAAGCTTCACGCCACGTCGAGAGCC
mslH-pSTV28N-N(NdeI)	CATCACCACATATGGGCAGCAGCCATCACC
mslB1-N(BamHI)	TGTAGGATCCTAAGCTGACCCTCGCCCG
mslB1-C(HindIII)	TAGCTAAGCTTTCATGAGGCCACCTCCACG
mslB1-N(NcoI)	CTCAACCATGGATCCTAAGCTGACCCTCGCC
mslC-pCola-N(NdeI)	GCATCTTGACATATG**CATCATCATCATCATCAC**ATGGAATTCGTGGTTCTTCCGGACAGTC
mslC-pCola-C(MfeI)	GCATCTTGACAATTGGCTTCATCCGAGTTCTCCCGAGTGCG
mal-mslC-F	AGAAGGAGATATACATATGAAAATCGAAGAAGGTAAACTGGTAATCTGG
mal-mslC-R	GATGATGATGATGCATAGAAATCCTTCCCTCGATCCCG
mslB2-pCola-N(BamHI)	TGTAGGATCCGATGACCACTCCCGCCGTGGCCGAAC
mslB2-pCola-C(EcoRI)	TGTAGAATTCTCAGCGGCTCCCGTCGATGCG
mal-mslB2-F	AATAAGGAGATATACCATGGAAATCGAAGAAGG
mal-mslB2-R	TGATGGCTGCTGCCCATGCCCATATGTGAAATCCTTCCCTCG

msl-SpeI-SphI-F	CTAGACTAGTGCCACGCCAACCG
msl-SpeI-SphI-R	ACATGCATGCGGGGCGCGAAC
msl-frag1-MS271-W21F-R	GACGTGATCAGAAGAAGCAGACGATCG
msl-frag2-MS271-W21F-F	GATCGTCTGCTTCTTCTGATCACGTC
msl-frag1-MS271-W21Y-R	GACGTGATCAGTAGAAGCAGACGATCG
msl-frag2-MS271-W21Y-F	GATCGTCTGCTTCTACTGATCACGTC
msl-frag1-MS271-W21V-R	GACGTGATCAGACGAAGCAGACGATCG
msl-frag2-MS271-W21V-F	GATCGTCTGCTTCGTCTGATCACGTC
msl-frag1-MS271-dW21-R	GACGTGATCATCAGAAGCAGACGATCG
msl-frag2-MS271-dW21-F	GATCGTCTGCTTCTGATGATCACGTC
msl-frag1-MS271-dI17-R	CAGAAGCAGACCGCGTAGCCGCAG
msl-frag2-MS271-dI17-F	GCTACGCGGTCTGCTTCTGGTG
msl-frag1-sviceucin-R	AGGAAGTCGGTGCAGTCTCCGCCCCACACACACTTCGTGAGCTCCTCGAAGTC

msl-frag2-sviceucin-F	CGGAGACTGCACCGACTTCCTCGGCTGCGGCACCGCCTGGATCTGTGTCTGATCACGTCCGGTGCC
msl-frag1-svi-W-R	GTGATCACCAGACACAGATCCAGGCGG
msl-frag2-svi-W-F	CCTGGATCTGTGTCTGGTGATCACGTC
msl-frag1-svi-CFW-R	GACGTGATCACCAGAAACAGATCCAG
msl-frag2-svi-CFW-F	CTGGATCTGTTTCTGGTGATCACGTC
msl-frag1-svi-VCFW-R	CAGAAACAGACCCAGGCGGTGCC
msl-frag2-svi-VCFW-F	ACCGCCTGGGTCTGTTTCTGGTG
msl-frag1-svi-AIVCFW-R	AAACAGACGATGGCGGTGCCGCAG
msl-frag2-svi-AIVCFW-F	GGCACCGCCATCGTCTGTTTCTGG

mslA-WT-N(BamHI)	ACATGGATCCGATGGGACAACTGCCCCATG
mslA-WT-C(HindIII)	ACAGAAGCTTTCACCAGAAGCAGACGATCG
mslB1-WT-N(NcoI)	CTCAACCATGGCGAAGCTGACCCTCGCCCGCGACGTC
mslB1-WT-C(HindIII)	TAGCTAAGCTTTCACGAGGCCACCTCCACGAGCTTGG
mslB2-WT-N(NdeI)	CATCACCACATATGACCACCCCCGCCGTGGCCGAAC
mslB2-WT-C(XhoI)	TAGTCTCGAGTCATGAGCGGCTCCCTCGGGAGGG
mslE-WT-N(NcoI)	CTCAACCATGGCGAGCCGGGGAGGACAGGAGACGG

mslE-WT-C(HindIII)	TAGCTAAGCTTTCACTCCTTCGGGTCCAGGACCC
mslF-WT-N(NdeI)	CATCACCACATATGTGAGGGGCATGACGAGCACACAGAC
mslF-WT-C(XhoI)	TAGTCTCGAGTCACCGCCGGGCCGCCGTGACC
mslC-WT-N(NcoI)	CTCAACCATGGAATTCGTGGTTCTTCCGGACTGC
mslC-WT-C(HindIII)	TAGCTAAGCTTCATCGGTGTTCTCCCGTAGGCGTG
mslH-WT-N(NdeI)	CATCACCACATATGACCCGGCTGACCGTGGCTC
mslH-WT-C(XhoI)	TAGTCTCGAGTCACGCCACGTCCAGAAGCAGCTCC
mslA-N(NdeI)	AGATATTCATATGTCCGCCATCTACGAGCCC
mslA-C(HindIII)	ACAGAAGCTTTCACCAGAAGCAGACGATCG
SKIK-His-mslA-N(NdeI)-2	AGATATTCATATG*TCTAAAATAAAA***CACCACCAC CACCACCAC**TCCGCTGTTTACGAGCCG
SKIK-His-mslA-C(XhoI)	TAGTCTCGAGTCACCAGAAGCAGACGATCG
His-mslH-N(NdeI)	ATAGTAACATATGACCCGGCTGACCGTGG
His-mslH-C(HindIII)	ATGTAAGCTTCGCCACGTCGAGAGCCAG
His-mslH-N(EcoRI)	TAGTAAGAATTCCATGACCCGGCTGACCGTGG
His-mslH-C(PstI)	TCAACTGCAGAGCCGGATCTCAGTGGTGGTG
SKIK-mslA-N(NdeI)	AGATATTCATATG*TCTAAAATAAAA*TCCGCTGTTTACGAGCCGC
SKIK-mslA-C(HindIII)	ACAGAAGCTTTCACCAGAAGCAGACGATCG
SKIK-mslA-pRSF-N(NdeI)	AGATATTCATATG*TCTAAAATAAAA*TCCGCTGTTTACGAGCCGC
SKIK-mslA-pRSF-C(XhoI)	TAGTCTCGAGTCACCAGAAGCAGACGATCG
mslH-up_F(PstI)	AATTCTGCAGAGTCCGCGATCAGACAGCCGCC
mslH-up_R(BamHI)	AATTGGATCCGCCAGATGCTCGTGGCGCCG
mslH-down_F(HindIII)	ATATAAGCTTTGGTGAATCTTGTGGGTGTCG

mslH-down_R(PstI)	TTAACTGCAGCACGCCTGGCCGACCAGGCCGAG
mslH-KO-F-1	CGTCGACGTGTACCTCGACTACCTCTGC
mslH-KO-R-1	CAATGGCTTCAGTAACGACGGGG
mslH-KO-F-2	ATGACATTTTAGCAAGGCTCTCTC
mslH-KO-R-2	TGTTCCTCGTCGTCCTCGGCAC
mslH-KO-F-3	GCAGCGGACATGGGAAGAAATG
mslH-KO-R-3	GACGGCTCCACGAACTCCTCCAC

Table 2 Plasmid constructions and heterologous expression of MS-271 derivatives in *S. lividans:mslR2*.

plasmid names	expected products (molecular weight)	core peptide sequences	primers used for the first PCR[1] (top: for fragment 1, bottom: for fragment 2)	plasmids used for PCR template
pWHM3-msl-MS-271-W21F	MS-271-W21F (2124.5)	CLGVGSCNDFAGCGYAIV CFF	msl-SpeI-SphI-F and msl-frag1-MS271-W21F-R msl-frag2-MS271-W21F-F and msl-SpeI-SphI-R	pWHM3-msl
pWHM3-msl-MS-271-W21Y	MS-271-W21Y (2140.5)	CLGVGSCNDFAGCGYAIV CFY	msl-SpeI-SphI-F and msl-frag1-MS271-W21Y-R msl-frag2-MS271-W21Y-F and msl-SpeI-SphI-R	pWHM3-msl
pWHM3-msl-MS-271-W21V	MS-271-W21V (2076.5)	CLGVGSCNDFAGCGYAIV CFV	msl-SpeI-SphI-F and msl-frag1-MS271-W21V-R msl-frag2-MS271-W21V-F and msl-SpeI-SphI-R	pWHM3-msl
pWHM3-msl-MS-271-ΔW21	MS-271-ΔW21 (1977.3)	CLGVGSCNDFAGCGYAIV CF	msl-SpeI-SphI-F and msl-frag1-MS271-dW21-R msl-frag2-MS271-dW21-F and msl-SpeI-SphI-R	pWHM3-msl

pWHM3-msl-MS-271-ΔI17	MS-271-ΔI17 (2050.4)	CLGVGSCNDFAGCGYAVCF	msl-SpeI-SphI-F and msl-frag1-MS271-dI17-R msl-frag2-MS271-dI17-F and msl-SpeI-SphI-R	pWHM3-msl
pWHM3-msl-svi	sviceucin (2084.4)	CVWGGDCTDFLGCGTAWICV	msl-SpeI-SphI-F and msl-frag1-sviceucin-R msl-frag2-sviceucin-F and msl-SpeI-SphI-R	pWHM3-msl
pWHM3-msl-svi-W	Svi-W (2270.6)	CVWGGDCTDFLGCGTAWICVW	msl-SpeI-SphI-F and msl-frag1-svi-W-R msl-frag2-svi-W-F and msl-SpeI-SphI-R	pWHM3-msl-svi
pWHM3-msl-svi-CFW	Svi-CFW (2318.7)	CVWGGDCTDFLGCGTAWICFW	msl-SpeI-SphI-F and msl-frag1-svi-CFW-R msl-frag2-svi-CFW-F and msl-SpeI-SphI-R	pWHM3-msl-sviW
pWHM3-msl-svi-VCFW	Svi-VCFW (2304.7)	CVWGGDCTDFLGCGTAWVCFW	msl-SpeI-SphI-F and msl-frag1-svi-VCFW-R msl-frag2-svi-VCFW-F and msl-SpeI-SphI-R	pWHM3-msl-sviCFW
pWHM3-msl-svi-AIVCFW	Svi-AIVCFW (2231.6)	CVWGGDCTDFLGCGTAIVCFW	msl-SpeI-SphI-F and msl-frag1-svi-AIVCFW-R msl-frag2-svi-AIVCFW-F and msl-SpeI-SphI-R	pWHM3-msl-sviVCFW

References

1. a) Kieser. T., Bibb, M. J., Buttner, M. J., Chater, K. F. Hopwood. D. A. *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, England, **2000**; b) Sambrook, J., Russel, W. *Molecular cloning*: *A Laboratory Manual*, 3rd ed., Cold Spring Harbor, NY, **2001**.
2. Onaka, H., Taniguchi, S., Ikeda, H., Igarashi, Y., Furumai, T. *J. Antibiot.* (Tokyo) **2003**, *56*, 950.
3. a) Asamizu, S., Shiro, Y., Igarashi, Y., Nagano, S., Onaka, H, *Biosci. Biotechnol. Biochem.* **2011**, *75*, 2184; b) Du, D., Katsuyama, Y., Onaka, H., Fujie, M., Satoh, N., Shin-Ya, K., Ohnishi, Y. *Chembiochem* **2016**, *17*, 1464.
4. Zerbino, D. R., Binery, E. *Genome research* **2008**, *18*, 821.
5. Vara, J., Lewandowska-Skarbek, M., Wang, Y. G., Donadio, S., Hutchinson, C. R. *J. Bacteriol* **1989**, *171*, 5872.
6. Hayashi, S., Satoh, Y., Ujihara,T., Takata, Y., and Dairi, T. *Scientific Reports*, 2016, **6**, 35441.
7. Schagger, H., *Nat Protoc*, 2006, **1**, 16-22.

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Professor Tohru Dairi for his guidance and support of my master's and doctoral studies as well as the opportunity to work on this project, as well as for the spur and encouragement when my research was not going smoothly. Besides, I realized the importance of having a wide range of knowledge and being self-discipiline as a research through his words and deeds.

Secondly, I also would like to express my great thanks to assistant professors Yasushi Ogasawara and Yasuharu Satoh for their advice on my research and help with daily life in the laboratory. I appreciated the careful revision of my written materials over the past five years by Yasushi Ogasawara, as well as his guidance of my research.

Thirdly, I also thank previous and current lab members. Daiki Naoe, Keisuke Kondo, Tomoyuki Kamide, Mayuko Shigematsu and Mai Naka were in the same year as me who gave me a lot of help in my master's career. Daiki Naoe taught my experimental skills generously and carefully which I thought I couldn't graduate on time without his help. Hayashi Shohei also gave me tremendous help, who encouraged me a lot when my research did not go well. I also learned a lot from his diligence, modesty and persistence.

Finally, I am deeply grateful to my friends and family for the encouragements and supports.

January, 2021

Sapporo