Identification of the Peptide Epimerase MslH Responsible for D-Amino Acid Introduction at the C-terminus of Ribosomal Peptides

Zhi Feng[†], Yasushi Ogasawara*[,‡], and Tohru Dairi*[,‡]

[†]*Graduate School of Chemical Sciences and Engineering* and [‡]*Graduate School of Engineering, Hokkaido University, Sapporo, Hokkaido 060-8628, Japan*

*Corresponding authors

E-mail: yogasawa@eng.hokudai.ac.jp and dairi@eng.hokudai.ac.jp

Tel: +81-11-706-7815

Supporting Tables and Figures

Table S1. Oligonucleotides used in this study. The introduced restriction sites were underlined. The SKIK and His-tag sequences introduced are shown with italic and bold letters, respectively.

primer names	sequences
SKIK-His-mslA-N(NdeI)	CGCATATG*TCTAAAATAAAA***CACCACCACCACCACCAC**TCCGCTGTTTACGAGCCG
SKIK-His-mslA-C(HindIII)	ACAGAAGCTTTCACCAGAAGCAGACGATCG
mslH-pET-N(NdeI)	ATAGTAACATATGACCCGGCTGACGGTGGCCCTGTC
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

Table S2. 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 (top: for fragment 1, bottom: for fragment 2)	plasmids used for PCR template
pWHM3-msl-MS-271-W21F	MS-271-W21F (2124.5)	CLGVGSCNDFAGCGYAIVCFF	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)	CLGVGSCNDFAGCGYAIVCFY	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)	CLGVGSCNDFAGCGYAIVCFV	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)	CLGVGSCNDFAGCGYAIVCF	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

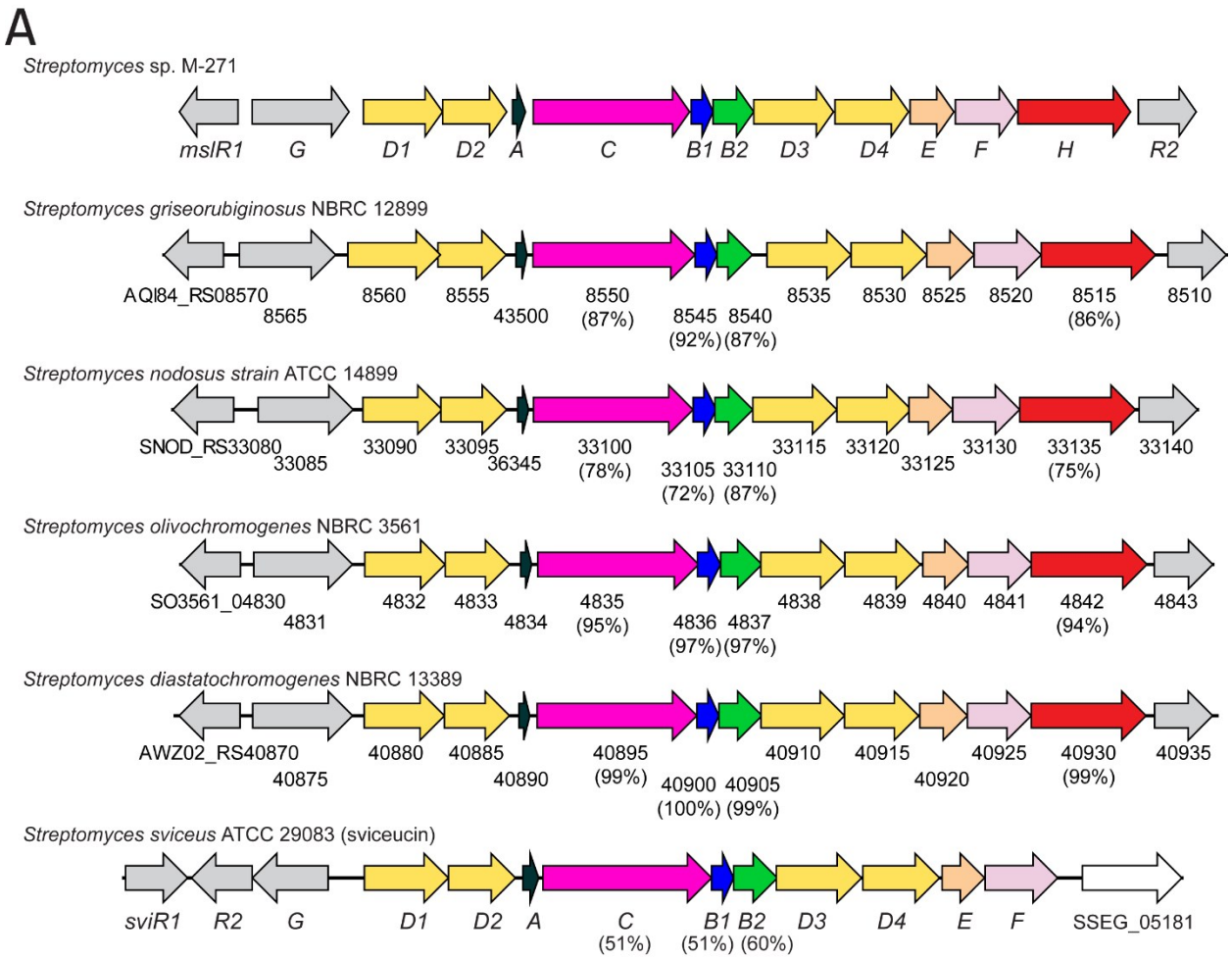


Figure S1. *Msl*-like clusters. (A) The gene organizations of the clusters. Identities to Msl proteins from *Streptomyces* sp. M-271 are shown in parentheses. (B) Sequence alignment of precursor peptides.

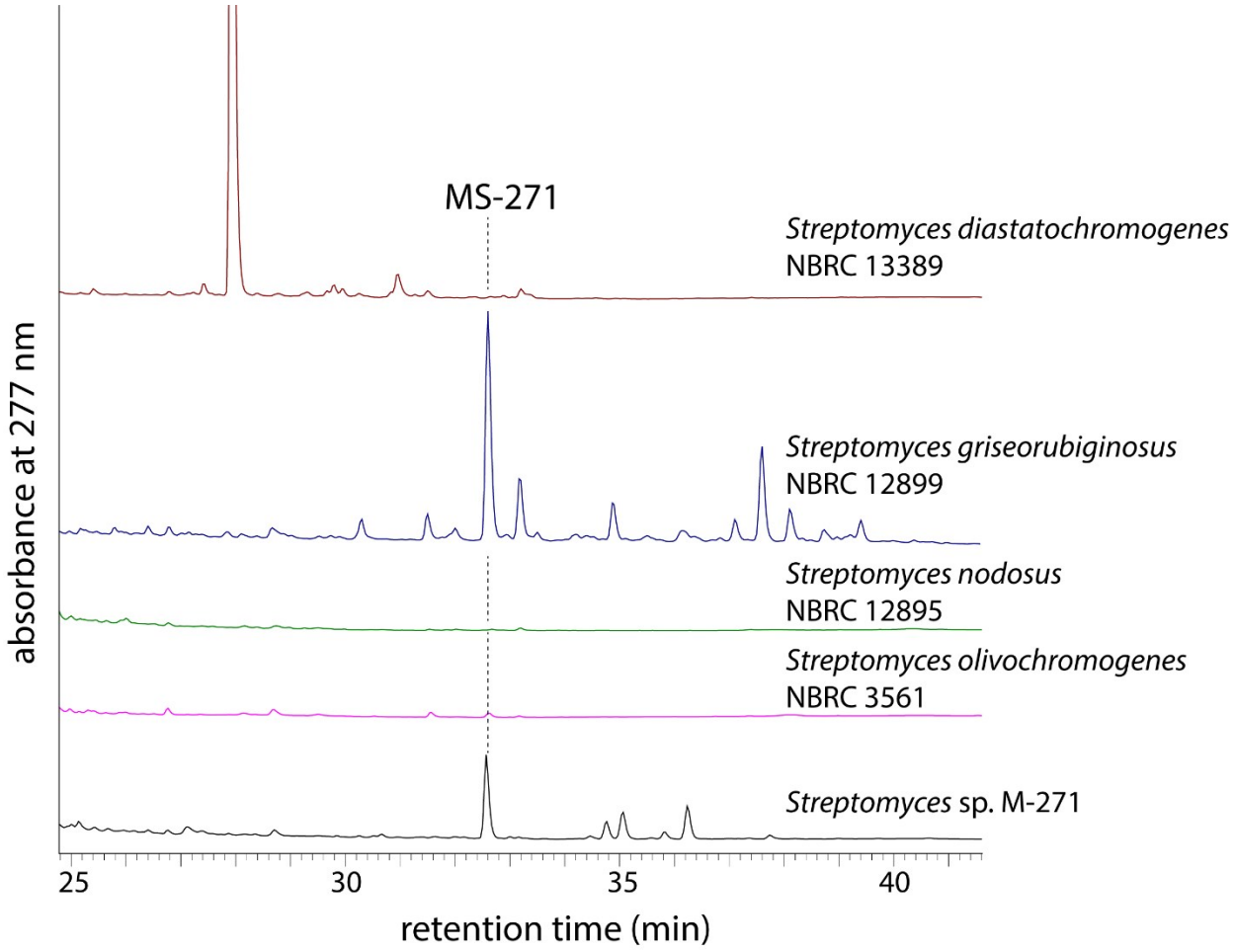


Figure S2. Production of MS-271 by *Streptomyces* strains harboring *msl*-like gene clusters.

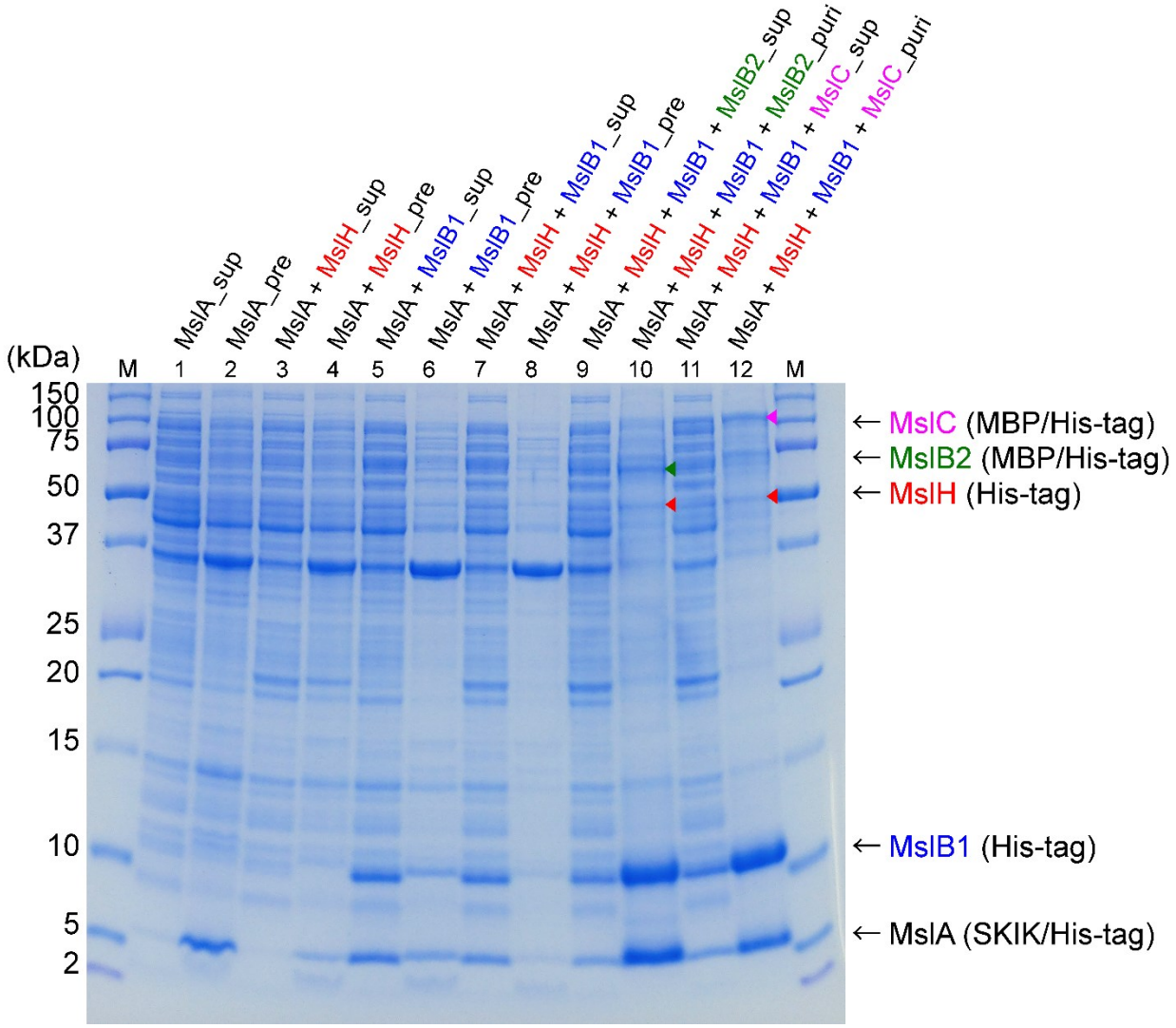


Figure S3. SDS-PAGE analysis of protein expressions of MslA (5.8 kDa), MslB1 (10.4 kDa), MslH (49 kDa), MslB2 (65 kDa), and MslC (110 kDa) for *in vivo* analysis. Samples include supernatant (sup) and precipitate (pre) fractions of the transformants expressing MslA (Lanes 1 and 2), MslA + MslH (Lanes 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, the Ni-NTA purified fractions (puri) from the transformants expressing MslA + MslH + MslB1 + MslB2 (Lane 10) and MslA + MslH + MslB1 + MslC (Lane 12) were also analyzed (MslH, MslB2 and MslC are marked with triangles.).

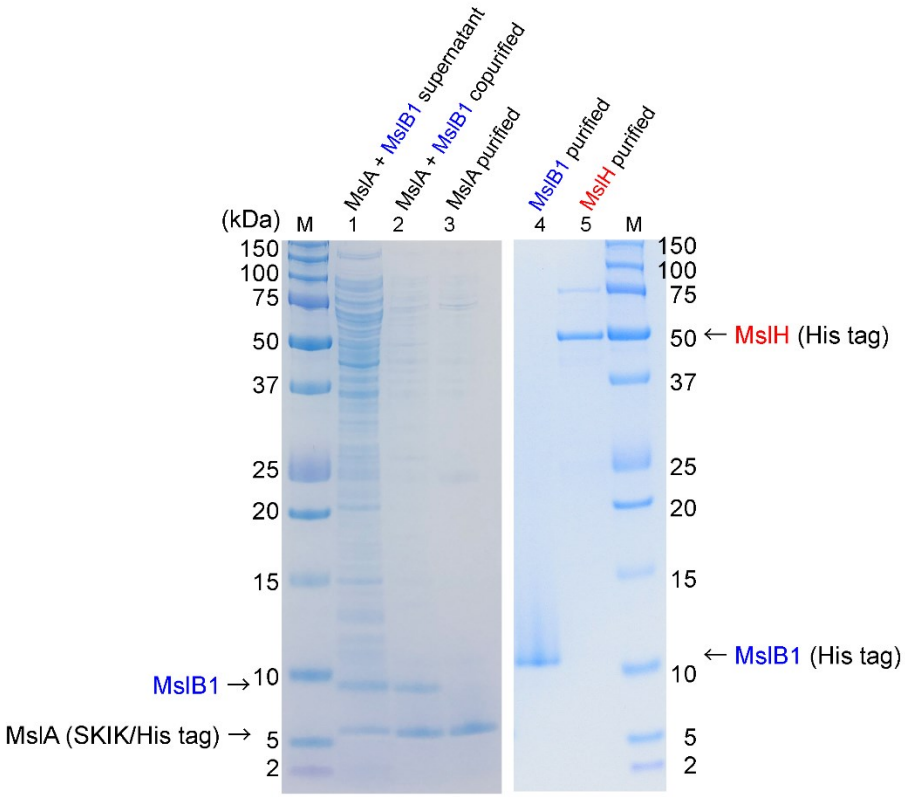


Figure S4. SDS-PAGE analysis of purified proteins for *in vitro* assay. Lane 1: Cell-free extract from the *E. coli* BL21(DE3) transformant harboring pET21-SKIK-His-mslA and pCDF-mslB1. Lane 2: Elution fraction of His-tagged MslA (5.8 kDa) after the Ni-NTA column was washed with high salt buffer (wash 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).

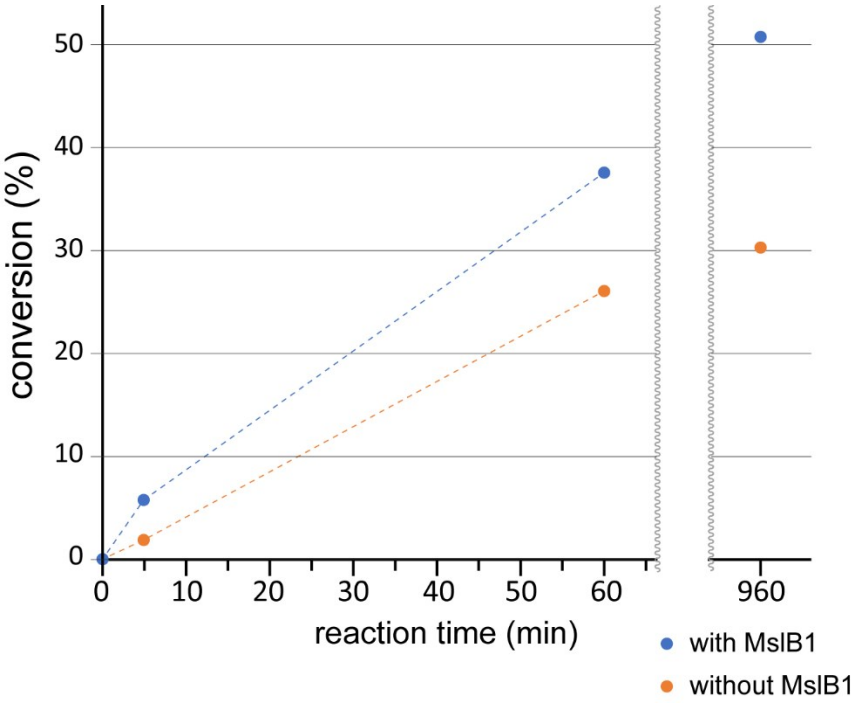


Figure S5. Plot of reaction time (5 min, 60 min, and 16 h) versus conversion from MslA to its epimer by MslH reaction in the presence and absence of MslB1. The conversion (%) was calculated based on the peak area of chromatograms of UV absorbance monitored at 340 nm.

Summary (rank 1-5)

template	3D model	confidence	% identity	template information
d1t70a_ 4 385		100.0	20	Fold:Metallo-dependent phosphatases Superfamily:Metallo-dependent phosphatases Family:DR1281-like
d1t71a_ 1 398		100.0	17	Fold:Metallo-dependent phosphatases Superfamily:Metallo-dependent phosphatases Family:DR1281-like
c4b2oB_ 4 385		100.0	21	PDB header:hydrolase Chain: B: PDB Molecule:ymdb phosphodiesterase; PDBTitle: crystal structure of bacillus subtilis ymdb, a global2 regulator of late adaptive responses.
d2z06a1 4 313		99.9	30	Fold:Metallo-dependent phosphatases Superfamily:Metallo-dependent phosphatases Family:TTHA0625-like
d1usha2 3 420		99.8	16	Fold:Metallo-dependent phosphatases Superfamily:Metallo-dependent phosphatases Family:5'-nucleotidase (syn. UDP-sugar hydrolase), N-terminal domain

Alignment with d1t70a_

Figure S6. Phyre2 analysis of MslH.

Top 10 threading templates used by I-TASSER

5bx9A (PslG from Pseudomonas aeruginosa)
Classification: HYDROLASE
Organism(s): Pseudomonas aeruginosa PAO1

1OI8 (5'-Nucleotidase (E. coli) with an Engineered Disulfide Bridge (P90C, L424C))
Classification: HYDROLASE
Organism(s): Escherichia coli

1HPU (5'-NUCLEOTIDASE (CLOSED FORM), COMPLEX WITH AMPCP)
Classification: HYDROLASE
Organism(s): Escherichia coli

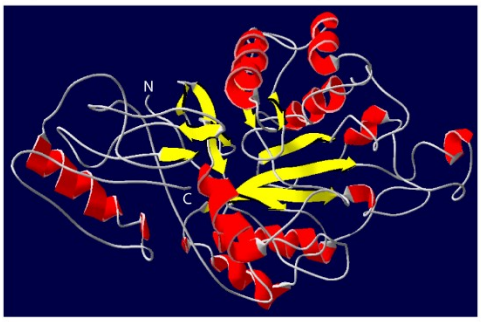
1T70 (phosphatase from Deinococcus radiodurans)
Classification: HYDROLASE
Organism(s): Deinococcus radiodurans R1

3JC8 (type IVa pilus machine in a piliated state)
Classification: MOTOR PROTEIN
Organism(s): Myxococcus xanthus DK 1622

2Z1A (5'-nucleotidase precursor from Thermus thermophilus HB8)
Classification: HYDROLASE
Organism(s): Thermus thermophilus HB8

1OID (5'-Nucleotidase (E. coli) with an Engineered Disulfide Bridge (S228C, P513C))
Classification: HYDROLASE
Organism(s): Escherichia coli

Top final model predicted by I-TASSER



Predicted function using COFACTOR and COACH

Ligand Binding Sites

Rank	C-score	PDB Hit	Lig Name	Bigand Binding Site Residue
1	0.06	5bxaA	MAN	19,20,49,322
2	0.05	1uh3A	GLD	32,35,36
3	0.05	NA	NA	11,12,13,14,42,44,46,49,86,140,144,146,220,259,262
4	0.05	NA	NA	15,18,43,45,67,70,71,83,133,135
5	0.05	4ucfC	POL	76.77

Enzyme Commission (EC) numbers and active sites

Rank	CscoreEC	PDB Hit	TM-score	RMSDa	IDENa	Cov	EC Number	Active Site Residues
1	0.175	1kwgA	0.634	4.09	0.080	0.770	3.2.1.23	NA
2	0.174	1px8A	0.696	4.24	0.078	0.859	3.2.1.37	NA
3	0.174	1uhvB	0.696	4.26	0.076	0.859	3.2.1.37	36
4	0.174	2bfgA	0.696	4.25	0.094	0.859	3.2.1.37	141
5	0.173	2bs9E	0.695	4.27	0.086	0.859	3.2.1.37	NA

Figure S7. I-TASSER analysis of MslH.

Figure S8. InterProScan analysis of MslH. Putative metal binding sites and active sites are labeled.

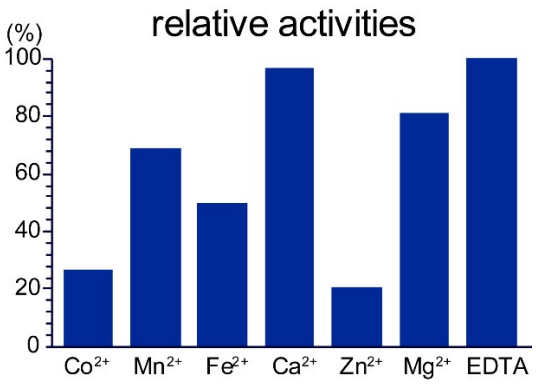


Figure S9. Metal dependency of the MslH reaction.

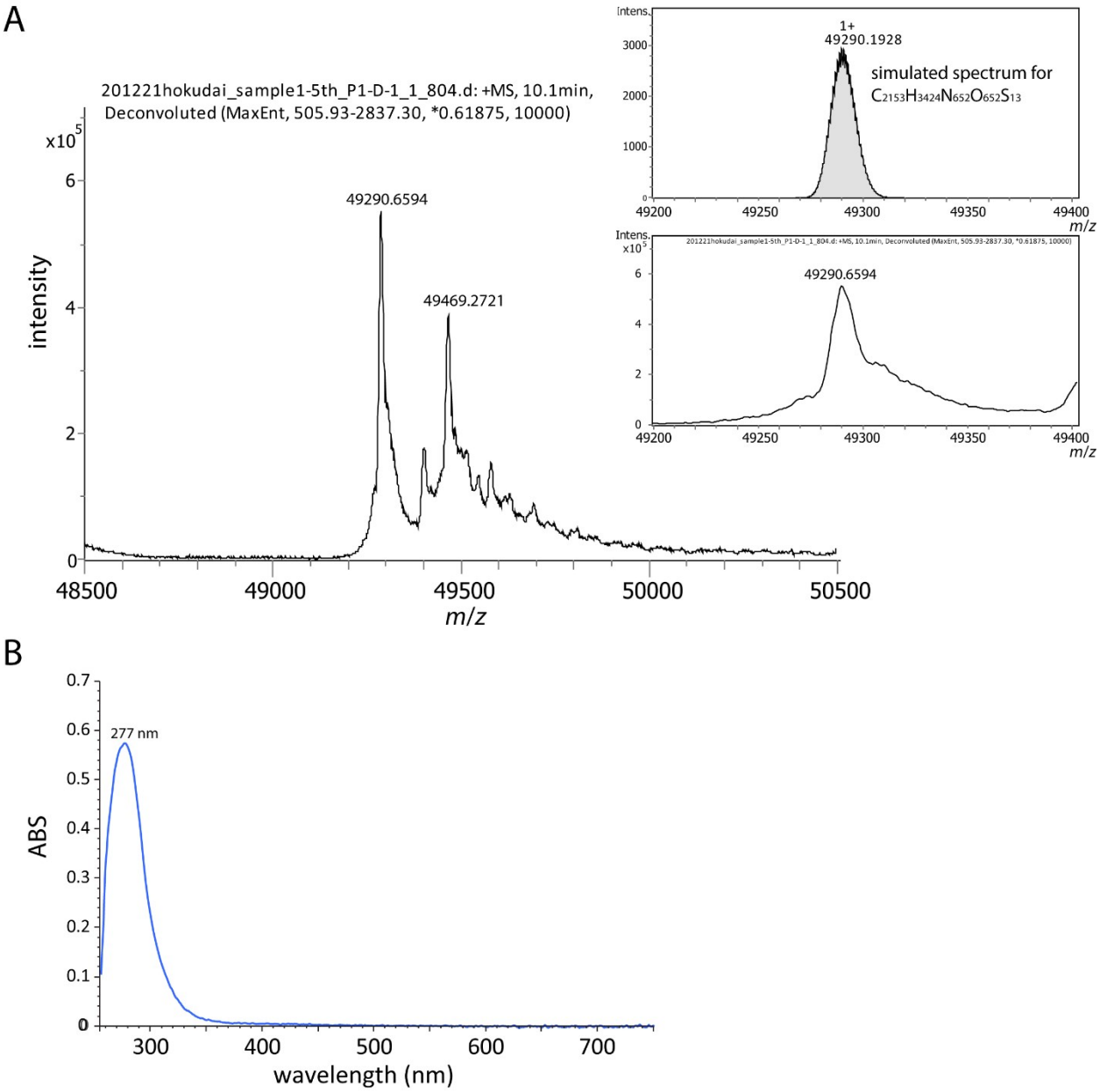


Figure S10. Spectroscopic analysis of MslH. (A) protein MS spectrum of purified MslH. The inlet spectra are enlarged spectrum from *m/z* 49200 to 49400 (bottom) and simulated spectrum for MslH (top). A peak at 49290.6 was consistent with the molecular formula ($C_{2153}H_{3424}N_{652}O_{652}S_{13}$) and its estimated mass (highest signal at 49290.2) calculated based on the amino acid sequence of MslH without N-terminal Met residue. A peak at 49469.3 is due to α-*N*-6 gluconoylation (+178) commonly observed for N-terminal His-tag fused protein (K. F. Geoghegan *et. al.*, *Anal*. *Biochem*. **1999**, *267*, 169). (B) UV-vis spectrum of MslH.

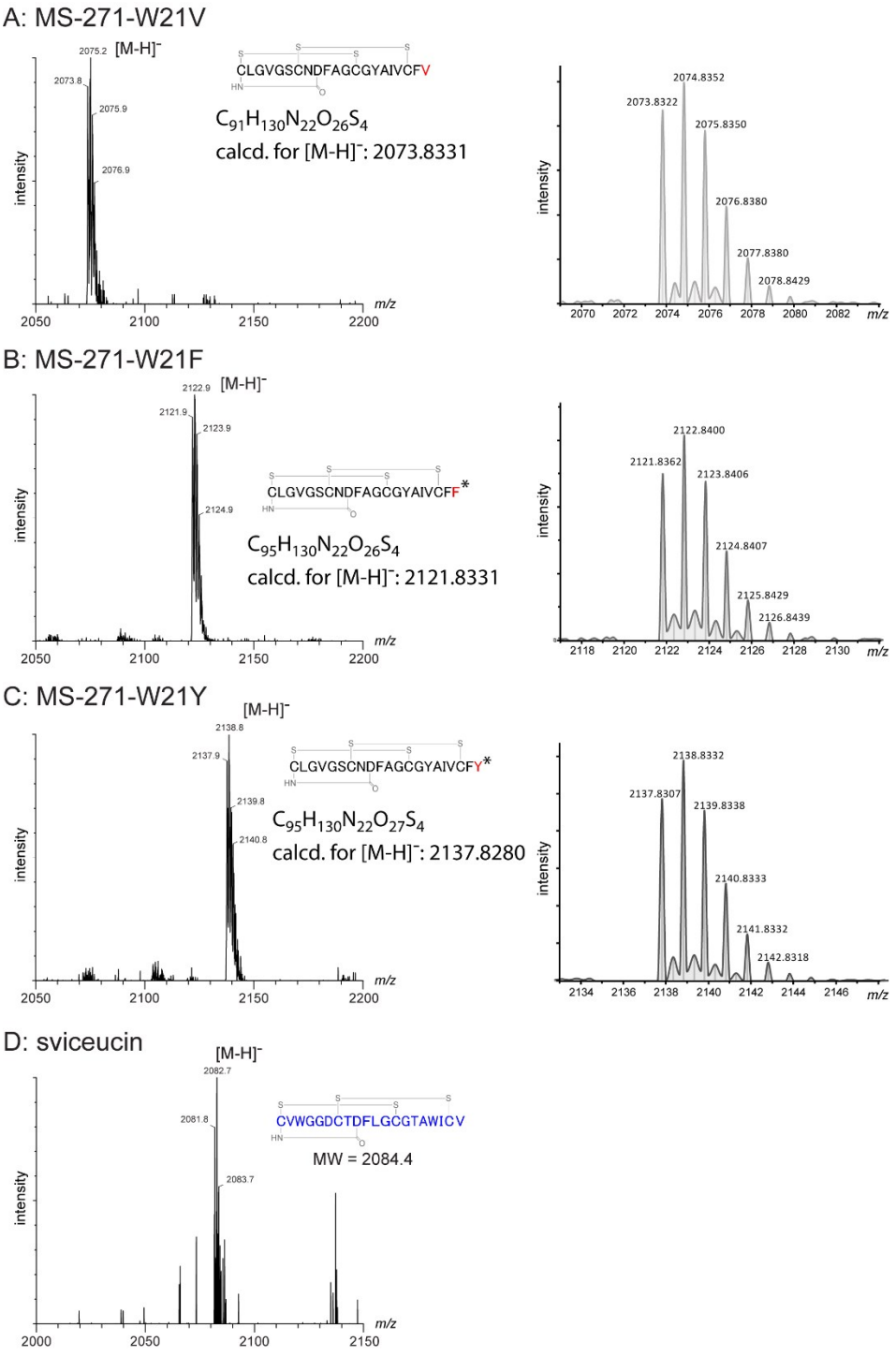


Figure S11. Mass spectra (ESI-negative, low-resolution [left] and high-resolution [right] data) of MS-271 derivatives. (A) MS-271-W21V, (B) MS-271-W21F, (C) MS-271-W21Y, and (D) sviceucin. D-Amino acid-containing products confirmed by chiral analysis are marked with an asterisk.

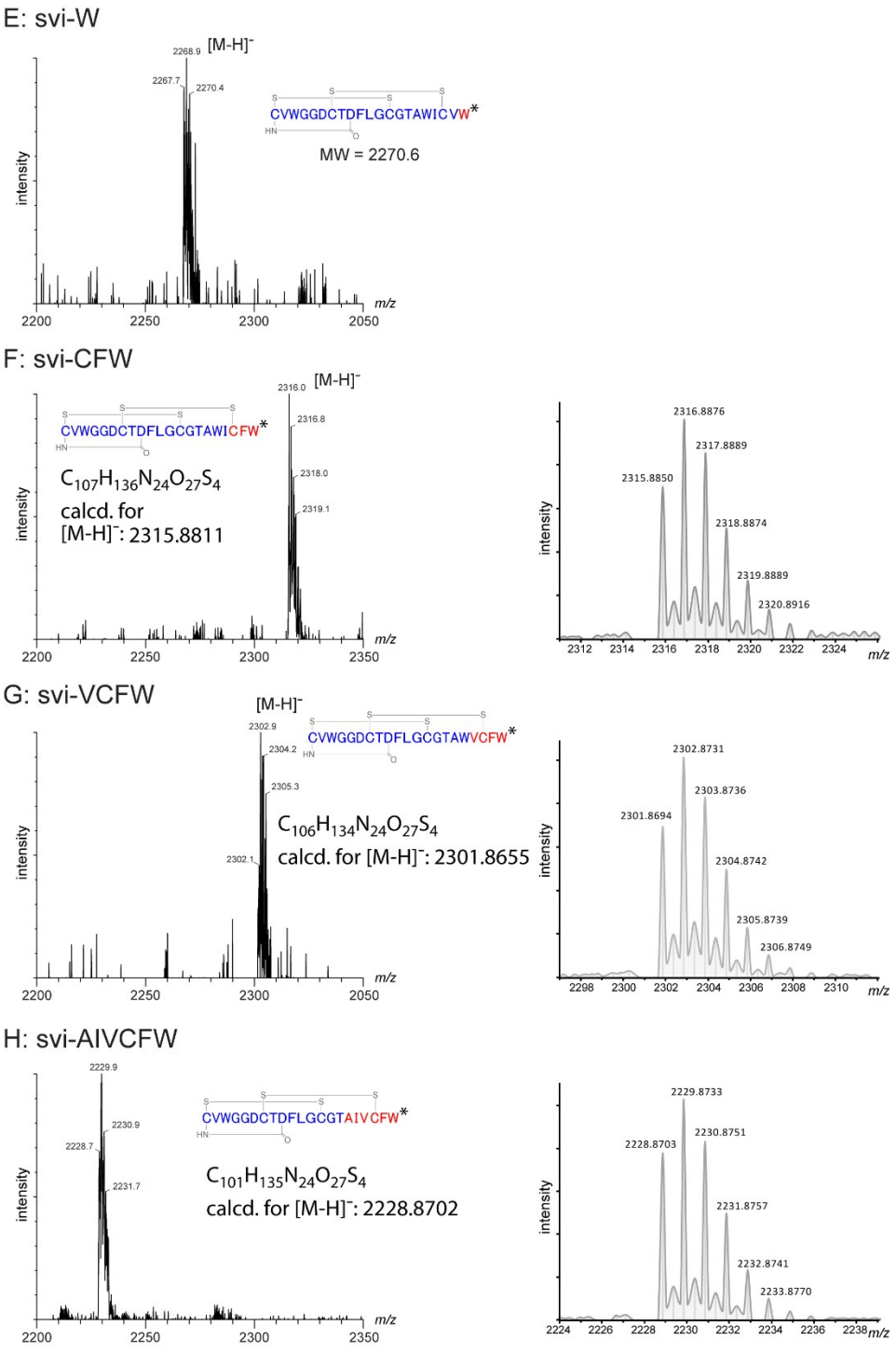


Figure S11 (continued). Mass spectra of MS-271 derivatives. (E) svi-W, (F) svi-CFW, (G) svi-VCFW, and (H) svi-AIVCFW.

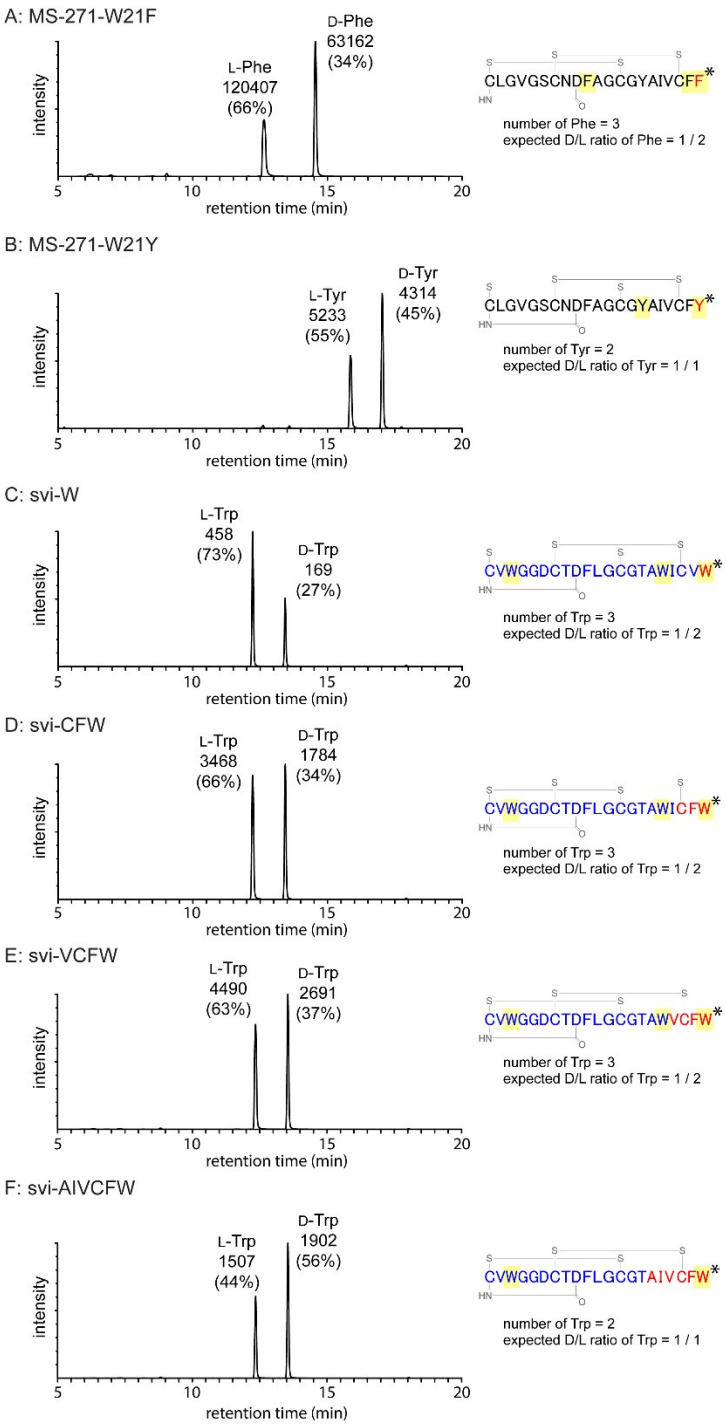


Figure S12. Chiral analysis of MS-271 derivatives. (A) Phe in MS-271-W21F, (B) Tyr in MS-271-W21Y, (C) Trp in svi-W, (D) Trp in svi-CFW, (E) Trp in svi-VCFW, and (F) Trp in svi-AIVCFW. The LC-MS chromatograms of L-FDLA derivatives of Phe, Tyr, and Trp were monitored at *m*/*z* 458, 474, and 497, respectively. 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.

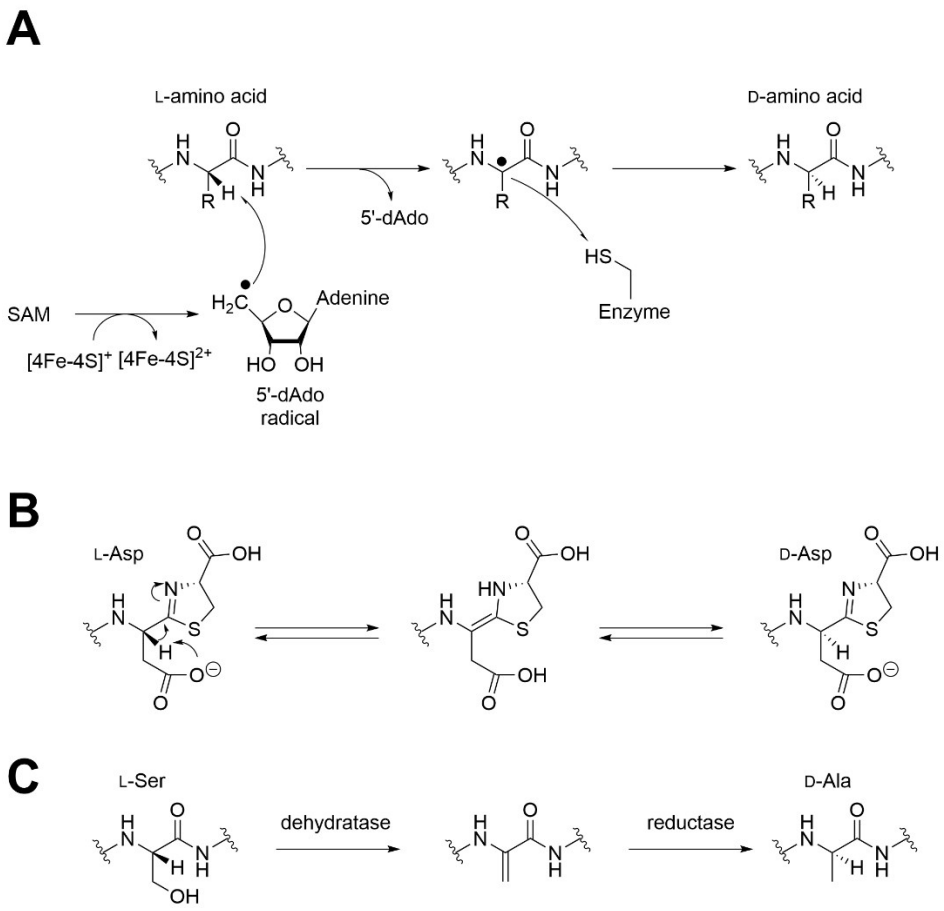


Figure S13. Reported D-amino acid introduction in RiPP biosynthesis. Reaction mechanisms of (A) a radical SAM-dependent epimerase, (B) an α/β-hydrolase family epimerase, and (C) a two-step introduction of D-amino acid residue.