Supporting Information

Functional and Structural Analyses of the Split-DH Domain in the Biosynthesis of Macrolactam Polyketide Cremimycin

Daisuke Kawasaki, Akimasa Miyanaga, Taichi Chisuga, Fumitaka Kudo and Tadashi Eguchi*

Department of Chemistry, Tokyo Institute of Technology, 2-12-1 O-okayama, Meguro-ku, Tokyo 152-8551, Japan

List of Supporting Information

Experimental

General procedure

All commercial reagents derived from TCI, Kanto Chemical and Sigma Aldrich were used as provided unless otherwise indicated. ^{1}H- and ^{13}C-NMR spectra were recorded in $CDCl_3$ with a JEOL ECS-400 or a JEOL ECZ-500 spectrometer. HRFABMS was performed on a JEOL JMS-700. Optical rotations were determined on a JASCO P-2200 polarimeter using the sodium D line (λ = 589 nm).

Synthesis of *racemic*-3-hydroxynonanoyl-SNAC

Boron trifluoride-diethyl ether complex (2.81 g, 19.8 mmol) was gently added over 15 min to a stirred suspension of heptanal (1.13 g, 9.9 mmol), ethyl bromoacetate (5.0 g, 29.7 mmol) and Zn dust (3.24 g, 49.5 mmol) in THF (30 mL) according to literature method.[1] After stirring for 2 h at room temperature, the mixture was filtrated, treated with 5% HCl and extracted with ethyl acetate. The organic layer was washed with brine and dried with Na_2SO_4. The solvent was removed and the residue was purified with silica gel column chromatography (0–25% AcOEt in hexane) to yield ethyl 3-hydroxynonanoate (1.48 g, 7.3 mmol, 74%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 4.18 (q, 2H, J = 7.2 Hz), 4.00 (m, 1H), 2.94 (br, 1H), 2.51 (dd, 1H, J = 16.8 Hz, 3.2 Hz), 2.40 (dd, 1H, J = 16.4 Hz, 8.8 Hz), 1.59–1.28 (m, 13H), 0.88 (t, 3H, J = 6.8 Hz); ^{13}C NMR (125 MHz, $CDCl_3$): δ 173.0, 67.9, 60.5, 41.3, 36.4, 31.7, 29.1, 23.3, 22.5, 14.1, 14.0

Ethyl 3-hydroxynonanoate (77.6 mg, 0.38 mmol) was dissolved in EtOH (1.5 mL). Sodium hydroxide (39.3 mg, 0.98 mmol) in water (0.5 mL) was added to the solution and the resulting mixture was stirred for 3 h at room temperature. The reaction was quenched with 1 M HCl and the mixture was extracted with ethyl acetate. The organic layer was washed with brine and dried with Na_2SO_4. The solvent was removed to yield 3-hydroxynonanoic acid (64.6 mg, 0.37 mmol, 97%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 4.04 (m, 1H), 2.58 (dd, 1H, J = 16.4 Hz, 3.2 Hz), 2.48 (dd, 1H, J = 16.4 Hz, 9.2 Hz), 1.59–1.29 (m, 10H), 0.89 (t, 3H, J = 6.8 Hz); ^{13}C NMR (125 MHz, $CDCl_3$): δ 177.8, 68.1, 41.1, 36.4, 31.7, 29.1, 25.4, 22.6, 14.0

A solution of 3-hydroxynonanoic acid (64.6 mg, 0.37 mmol) in dehydrated CH_2Cl_2 (1 mL) was mixed with a solution of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide-HCl (108.6 mg, 0.57 mmol) and *N,N*-dimethylaminopyridine (18.1 mg, 0.15 mmol) in dehydrated CH_2Cl_2 (1 mL). *N*-acetylcysteamine (56 mg, 0.47 mmol) was then added to the mixture. After stirring for 15 h at room temperature, the reaction was quenched with saturated NH_4Cl aqueous solution and the mixture was extracted with CH_2Cl_2. The organic layer was dried with Na_2SO_4 and evaporated. The residue was purified with silica gel column chromatography (2% methanol in chloroform) to yield *racemic*-3-hydroxynonanoyl-SNAC (64.8 mg, 0.24 mmol, 62%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.80 (br, 1H), 4.06 (m, 1H), 3.46

(m, 2H), 3.05 (m, 2H), 2.76 (dd, 1H, J = 15.4 Hz, 3.4 Hz), 2.69 (dd, 1H, J = 15.4 Hz, 8.6 Hz), 1.97 (s, 3H), 1.56–1.28 (m, 10H), 0.88 (t, 3H, J = 6.6 Hz); ^{13}C NMR (100 MHz, $CDCl_3$): δ 199.5, 170.4, 68.8, 51.0, 39.3, 36.7, 31.7, 29.1, 28.8, 25.4, 23.2, 22.5, 14.0

Preparation of (*S*)- and (*R*)-3-hydroxynonanoyl-SNAC

Racemic 3-hydroxynonanoyl-SNAC was separated into its enantiomers by preparative HPLC with a Hitachi instrument (Chromaster 5110 Pump and UV Detector L-7405; Hitachi) equipped with a CHIRAL ART Amylose-SA column (5 μm, 250×10 mm; YMC). The (*S*)- and (*R*)-enantiomers were detected at 230 nm and eluted with an isocratic solvent system (eluent: 7% 2-propanol in hexane, flow rate: 4.7 mL/min) in 23 and 26 min, respectively. (*S*)-3-hydroxynonanoyl-SNAC; $[\alpha]_{589}^{27}$ = 10.27 ± 0.46 (c = 0.21, $CHCl_3$); HRFABMS (positive mode): $[M + H]^+$ ion at *m/z* 276.1655 (calcd $[M + H]^+$ ion for $C_{13}H_{26}NO_3S$ at *m/z* 276.1628); the NMR data was identical to that of *racemic*-3-hydroxynonanoyl-SNAC. (*R*)-3-hydroxynonanoyl-SNAC; $[\alpha]_{589}^{27}$ = −14.55 ± 1.00 (c = 0.25, $CHCl_3$); HRFABMS (positive mode): $[M + H]^+$ ion at *m/z* 276.1649 (calcd $[M + H]^+$ ion for $C_{13}H_{26}NO_3S$ at *m/z* 276.1628). The NMR data was identical to that of *racemic*-3-hydroxynonanoyl-SNAC.

Synthesis of (*R*)-3-hydroxynonanoyl-SNAC MTPA esters

(*R*)-3-Hydroxynonanoyl-SNAC (2 mg, 7 μmol) and *N,N*-dimethylaminopyridine (1.5 mg, 10 μmol) was dissolved in super dehydrated CH_2Cl_2 (0.10 mL). Then, distilled trimethylamine (2 μL, 15 μmol) and (*S*)- or (*R*)-α-methoxy-α-(trifluoromethyl)phenylacetyl chloride (MTPA-Cl) (6 μL, 29 μmol) was added to the solution, and the resulting mixture was stirred at room temperature for 12 h. The reaction was quenched with 0.1 mL of water and stirred for additional 30 min. The mixture was extracted with ethyl acetate (5 mL) and the organic layer was washed with 0.3 M HCl aq., sat. $NaHCO_3$ aq. and brine, and dried with Na_2SO_4. The solvent was removed *in vacuo* and the resultant residue was purified with preparative TLC (83% ethyl acetate in hexane) to give (*R*)-MTPA ester (2.1 mg, 4 μmol, 57%) and (*S*)-MTPA ester (3.0 mg, 6 μmol, 86%). (*R*)-MTPA-(*R*)-3-hydroxynonanoyl-SNAC ester; ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.52 (m, 2H), 7.40 (m, 3H), 5.85 (br, 1H), 5.49 (m, 1H), 3.55 (s, 3H), 3.41 (m, 2H), 3.04 (t, 2H, J = 6.2 Hz), 2.95 (dd, 1H, J = 15.6 Hz, 8.4 Hz), 2.82 (dd, 1H, J = 15.6 Hz, 4.0 Hz), 1.95 (s, 3H), 1.26–1.18 (m, 10H), 0.86 (t, 3H, J = 7.0 Hz); ^{13}C NMR (125 MHz, $CDCl_3$): δ 196.1, 170.4, 166.2, 132.0, 129.7, 128.4, 127.3, 73.5, 55.6, 47.9, 39.0, 33.8, 31.5, 29.7, 29.0, 28.8, 24.6, 23.1, 22.4, 14.0; HRFABMS (positive mode): $[M + H]^+$ ion at *m/z* 492.2005 (calcd $[M + H]^+$ ion for $C_{23}H_{33}F_3NO_5S$ at *m/z* 492.2026), (*S*)-MTPA-(*R*)-3-hydroxynonanoyl-SNAC ester; ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.52 (m, 2H), 7.41 (m, 3H), 5.82 (br, 1H), 5.49 (m, 1H), 3.51 (s, 3H), 3.38 (m, 2H), 2.99 (t, 2H, J = 6.0 Hz), 2.91 (dd, 1H, J = 15.6 Hz, 8.0 Hz), 2.78 (dd, 1H, J = 15.6 Hz, 4.4 Hz), 1.95 (s, 3H),

1.31–1.24 (m, 10H), 0.88 (t, 3H, J = 7.0 Hz); ^{13}C NMR (100 MHz, $CDCl_3$): δ 195.9, 170.3, 166.0, 131.8, 129.7, 128.4, 127.5, 73.5, 55.4, 47.6, 39.0, 33.9. 31.5, 29.7, 28.9, 28.8, 24.9, 23.1, 22.5, 14.0; HRFABMS (positive mode): $[M + H]^+$ ion at *m/z* 492.2042 (calcd $[M + H]^+$ ion for $C_{23}H_{33}F_3NO_5S$ at *m/z* 492.2026)

Construction of an expression system for fusion-DH protein

The *cmiP3-cmiP2* DNA fragment connected by artificial linker, which was optimized for overexpression in *E. coli*, was synthesized by Eurofins Genetics (Figure S3). The *fusion-DH* (DH_N-*linker*-DH_C) gene region was amplified by PCR with primer 1 and 2 (Table S2) and cloned into the pColdI vector by using the *Nde*I and *Xho*I restriction sites to obtain pColdI-*fusion-DH*. The plasmid was introduced into *E. coli* BL21(DE3) to construct the expression system for the fusion-DH protein.

Construction of an expression system for CmiP2-KR protein

The *cmiP2* gene was amplified by PCR with primer 3 and 4 (Table S2) using the cosmid vector pOJ446 that contains a part of cremimycin biosynthetic gene cluster derived from *Streptomyces* sp. MJ635-86F5 as a template DNA.[2] The amplified fragment was cloned into the pMD19 T-vector to obtain pMD19-*cmiP2*. The *cmiP2-KR* fragment was amplified by PCR using primers 5 and 6 (Table S2) from pMD19-*cmiP2*. The amplified fragment was cloned into the *Nde*I and *Hin*dIII sites of pColdI to construct pColdI-*cmiP2-KR*. The plasmid was introduced into *E. coli* Rosetta 2 (DE3) to construct the expression system for the CmiP2-KR protein.

Expression and purification of the fusion-DH and CmiP2-KR proteins

Each transformant was cultivated at 37 °C in LB medium containing 50 μg/ml ampicillin (plus 50 μg/mL chloramphenicol for CmiP2-KR production) with 200 rpm agitation. After the optical density at 600 nm reached 0.6, protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (final conc. 0.2 mM), and the cells were cultured for an additional 20 h at 15 °C with 200 rpm agitation. The cells were harvested by centrifugation and washed with lysis buffer (buffer A for the fusion-DH: 50 mM HEPES−NaOH (pH 7.0), 10% glycerol, buffer B for CmiP2-KR: 50 mM HEPES−NaOH (pH 7.2), 10% glycerol). The wet cells were suspended in lysis buffer and the suspension of cells was disrupted by sonication (QSONICA) for 5 sec with a 5 sec interval (total 2 min) on ice. Cell debris was removed by centrifugation. The supernatant was loaded onto a TALON resin (Clontech) column (15 × 15 mm) after equilibrated with lysis buffer. The column was washed with lysis buffer containing 20 mM imidazole to remove unbound proteins. The protein was then

eluted with lysis buffer containing 200 mM imidazole. The protein solution was desalted with a PD-10 desalting column (GE Healthcare Life Science) equilibrated with lysis buffer, and concentrated by using an Amicon Ultra 10K centrifugal filter (Merck Millipore) at 5000 *g*. The fusion-DH protein was further purified by a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Science) with lysis buffer A containing 150 mM NaCl. The purified protein was then desalted and concentrated with a PD-10 column and an Amicon Ultra 10K centrifugal filter, respectively. The concentrations of proteins were determined by using NanoDrop (Thermo Fisher Scientific) with extinction coefficients (fusion-DH: ε_{280nm} = 33570 M^{-1} cm^{-1}, CmiP2-KR: ε_{280nm} = 52730 M^{-1} cm^{-1}).

Preparation of split-DH protein

The fusion-DH protein (final conc. 50 μM) in cleavage buffer (20 mM Tris-HCl (pH 8.4), 150 mM NaCl, 2.5 mM $CaCl_2$) was treated with 2 U/mL of biotinylated thrombin protease (Novagen). After incubation at 4 °C for 20 h, the thrombin was captured with agarose and filtrated. The protein solution was desalted with a PD-10 column and concentrated with an Amicon Ultra 10K centrifugal filter to give the split-DH protein.

***In vitro* assay of DH reaction**

The reaction mixture (100 μL) contained 5 μM DH, 1 mM 3-hydroxynonanoyl-SNAC ((*R*)- or (*S*)-enantiomer) in buffer A with 1% DMSO. After incubation at 28 °C for 15 min, the solution was extracted with ethyl acetate and the organic layer was evaporated. The dried residue was dissolved in 100 μL methanol and applied to HPLC analysis with a Hitachi instrument (ELITE La Chrom L-2130 Pump and UV Detector L-2455; Hitachi) equipped with a PEGASIL ODS SP100 column (5 μm, 250×4.6 mm; Senshu). The reaction product was detected at 260 nm and eluted with an isocratic solvent system (eluent: 80% MeOH in H_2O, flow rate: 1.0 mL/min) in 8.5 min.

***In vitro* assay of DH reverse reaction**

The reaction mixture (100 μL) contained 5 μM DH, 1 mM (*E*)-non-2-enoyl-SNAC[2] in buffer A with 1% DMSO. After incubation at 28 °C for 1 h, the solution was extracted with ethyl acetate and the organic layer was evaporated. The dried residue was dissolved in 100 μL methanol and applied to HPLC analysis with a Hitachi instrument (Chromaster 5110 Pump and UV Detector L-7405; Hitachi) equipped with a CHIRAL ART Amylose-SA column (5 μm, 250×4.6 mm; YMC). The reaction product was detected at 230 nm and eluted with an isocratic solvent system (eluent: 7% 2-propanol in hexane, flow rate: 1.0 mL/min) in 26 min.

***In vitro* assay of CmiP2-KR reaction**

The reaction mixture (100 μL) contained 10 μM CmiP2-KR, 1 mM 3-oxononanoyl-SNAC[2] and 2 mM NADPH in buffer B with 1% DMSO. After incubation at 28 °C for 14 h, the solution was extracted with ethyl acetate and the organic layer was evaporated. The dried residue was dissolved in 100 μL methanol and analyzed with HPLC by the same way of DH reverse reaction.

Isolation of DH product

Racemic-3-hydroxynonanoyl-SNAC (5.0 mg, 2 mM) was dissolved in buffer A (total 9 mL) with 2% DMSO. The fusion-DH (final 5 μM) was added to the solution and the mixture was stirred at room temperature for 10 h. The solution was extracted with ethyl acetate and the organic layer was evaporated. The residue was purified with preparative TLC (75% ethyl acetate in hexane) to yield (*E*)-non-2-enoyl-SNAC (1.0 mg). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.94 (dt, 1H, J = 15.6 Hz, 6.8 Hz), 6.14 (dt, 1H, J = 15.6 Hz, 1.6 Hz), 5.87 (br, 1H), 3.47 (q, 2H, J = 6.1 Hz), 3.09 (t, 2H, J = 6.4 Hz), 2.21 (dq, 2H, J = 7.3 Hz, 1.2 Hz), 1.97 (s, 3H), 1.51–1.22 (m, 8H), 0.89 (t, 3H, J = 6.8 Hz); ^{13}C NMR (125 MHz, $CDCl_3$): δ 190.5, 170.3, 146.9, 128.2, 39.9, 32.3, 31.5, 28.8, 28.2, 27.9, 23.2, 22.5, 14.0

Crystallization and structural determination of the fusion-DH protein

Crystals of the fusion-DH protein were grown using sitting-drop vapor diffusion by mixing fusion-DH protein solution [10 mg/mL in 5 mM HEPES-Na (pH 8.0), 10% glycerol] with an equal volume of reservoir solution [0.1 M MES-Na (pH 6.5), 30% polyethylene glycol monomethyl ether 5000 and 0.2 M ammonium sulfate] at 20 °C. Crystals were flash-frozen in a liquid nitrogen stream prior to X-ray data collection. Diffraction data were collected at the beamline BL-5A at the Photon Factory (Tsukuba, Japan) and subsequently processed with XDS software[3]. The initial phase was determined by molecular replacement using the Molrep program[4] with the structure of the DH domain from module 10 of the rifamycin PKS (PDB entry 4LN9) as a search model. Protein model building of fusion-DH was performed automatically with the ARP/wARP program[5] and subsequently inspected by Coot[6]. Refmac[7] was used to refine the structure. The structural representations were prepared with PyMOL (https://pymol.org). The geometries of the final structure were evaluated using the program MolProbity[8]. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB entry 6K97).

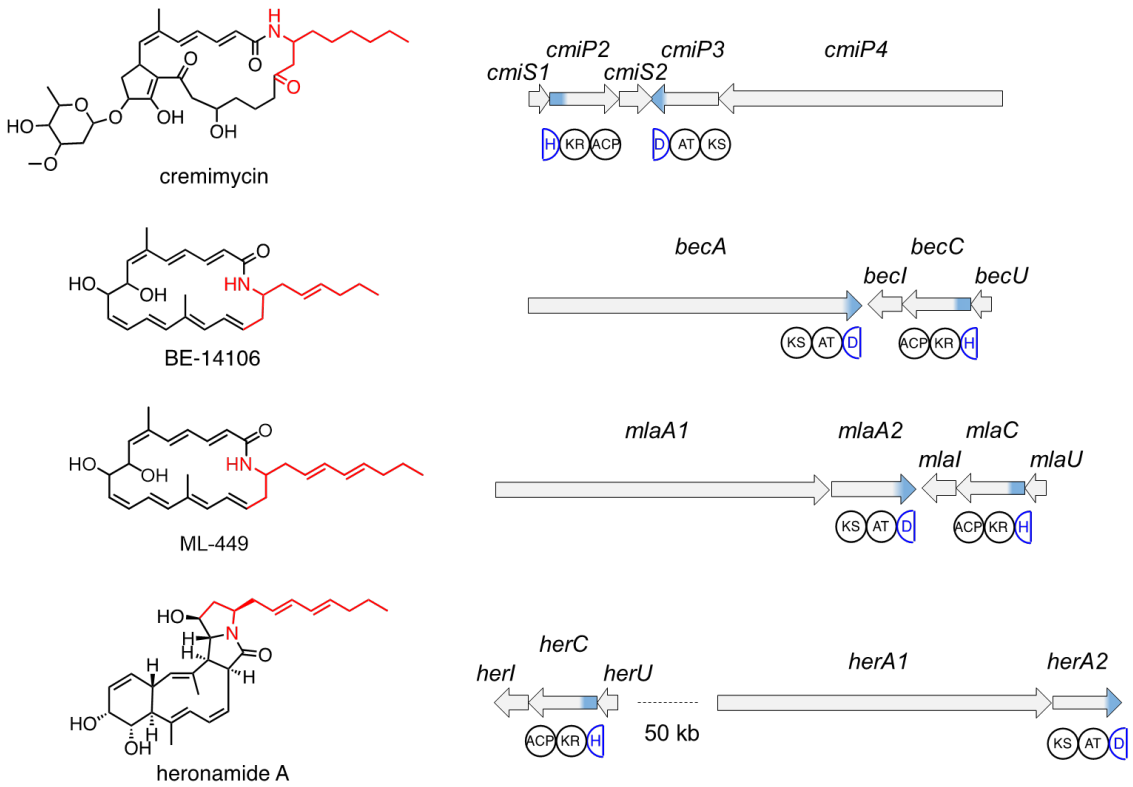

Figure S1. Biosynthetic genes for 3-amino fatty acid units in cremimycin related compounds. 3-Amino fatty acid units are shown in red. DH_N and DH_C subdomains are shown in blue.

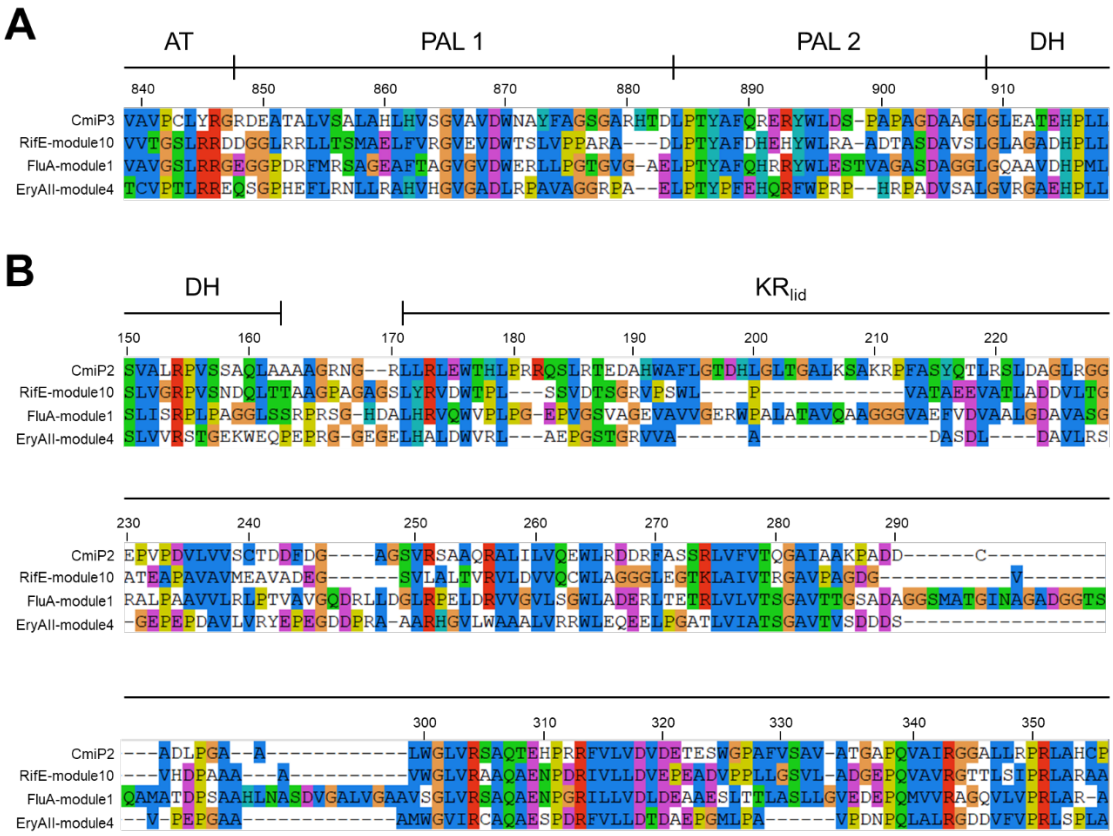

Figure S2. Multiple sequence alignment in pre or post DH regions. A) Comparison of N-terminal region of CmiP3 DH_N with other canonical *cis*-AT PKSs. B) Comparison of C-terminal region of CmiP2 DH_C with other canonical *cis*-AT PKSs. The amino acid sequences were aligned by ClustalX.[9] The sequences of the following enzymes were used for analysis: RifE in rifamycin biosynthesis (AAC01714), FluA in fluvirucin biosynthesis (AFS18277), EryAII in erythromycin biosynthesis (CAM00064).

5’CATATGGCGGTACCATGCCTGTATCGCGGACGTGACGAAGCGACCGCGCTAGTTTCGGC
GCTGGCGCATTTGCACGTTAGCGGGGTCGCAGTTGACTGGAACGCCTATTTCGCGGGTTC
TGGTGCGCGCCATACCGATTTACCGACCTATGCCTTTCAGCGCGAGCGCTATTGGCTGGA
TTCTCCGGCACCAGCTGGCGATGCGGCTGGTCTGGGACTGGAAGCGACAGAACATCCGT
TACTGGCCACTGCCACTGAGCTTCCGGATGGTGGTTACCTGTTCACTGGCCGTCTGGCGT
TGAGGGAGCATCCATGGTTAGCTGACCACACCATTGCGGGTACCACGATTGTACCGGGTA
CCGCCTTTGTCGAGCTTGCGCTACATGCAGCAGACATTGCCGGCTGTGACGAAATCACGG
AACTCGTCCTGCATACCCCCTTAGTCCTGAGCACCCAGAGCAGCTCACTCCTGCAAGTGG
CTGTTGGCCCGGCAGATCCCTCAGGTGCGCGATCCCTGACGATACGGAGTCATGGGGAA
GACGTGCGGTTGTGGGTGGAACACGCCGATGGAAGCATTGGCCCGGGTGAAACAGCCC
CGGAACTGCCTGAACCGGTGCGCGTAGGTCTGGTGCCGAGAGGTAGCGGGATGACGGG
CTCCGAAGAGGCCTTTGCCTCTGCAGATGGCACGGATGAAGCCTGGCCGCCACCTGGAG
GCGACGCATGGGATACGGCCGGCTTGTACGCTCGTCTTGCAGATCGTGGCTTTCAGTATG
GGGAGACATTTCGCGGCTTACGTGCTGCGTGGTCAAGTGGCGAGGACATCTACGCAGAT
GTGGAAGTGGGTGCGCCTGCGTCGAGTCCGAAACCGGAAGCGTTCCACGTGCATCCGGC
CTTGCTGGACGCTGCTTTGCATGCAGCTCTCGGGCCACTTCTGGATGGCGAAGAAGGCC
TGTTCCTGCCCTTTGCGCTGCGTCGCGTCCGTGTACACCACAGTGGTGCGAAATCACTCC
GTGTGCACATCACCCCCGATGGCGATAAGTCCGTGAGCCTGTCGGCGGTTGATGCTGCCG
GGAATGCGGTGGTTAGCGTTGGCTCCGTTGCATTACGCCCTGTCTCGTCTGCCCAACTGG
CAGCCGCAGCGGGTCGCAACGGCTAACTCGAG3’

Figure S3. The codon-optimized *cmiP3* and *cmiP2* fragments connected by artificial linker. The *fusion-DH* (*DH_N-linker-DH_C*) gene region is shown in red. The inserted linker region is colored yellow. The restriction sites are underlined.

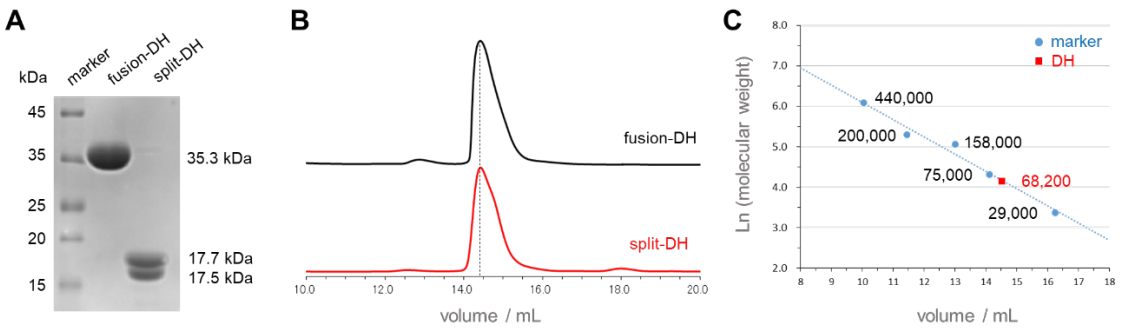

Figure S4. SDS-PAGE and gel filtration analyses of the fusion-DH and the split-DH. A) SDS-PAGE analysis of the fusion-DH and the split-DH. B) The charts of gel filtration chromatography. Black and red lines indicate the chromatograms of the fusion-DH and the split-DH, respectively. C) The logarithmic plot of molecular weight versus elution volume. The standard proteins (440 kDa of ferritin, 200 kDa of β-amylase, 158 kDa of aldolase, 75 kDa of albumin, and 29 kDa of carbonic anhydrolase) are represented by blue circles. The elution volume of fusion-DH (red square) was fitted to the standard curve.

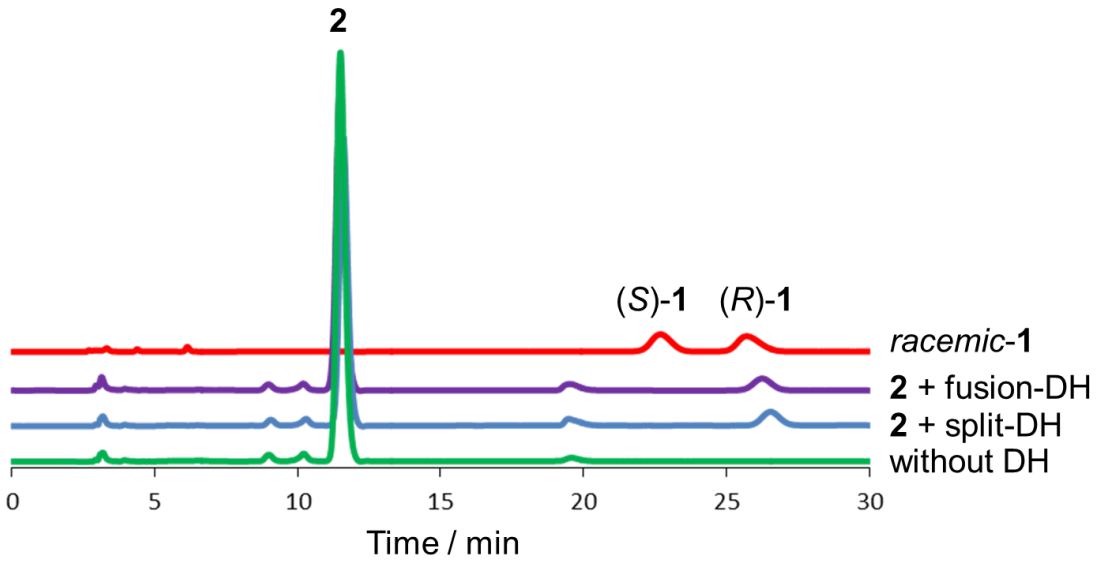

Figure S5. HPLC analysis of hydration reaction by the fusion-DH and the split-DH. Red line: the synthesized 3-hydroxynonanyl-SNAC (**1**) (racemic), purple line: the reaction of the fusion-DH with (*E*)-non-2-enoyl-SNAC (**2**), blue line: the reaction of the split-DH with **2**, green line: the reaction without DH.

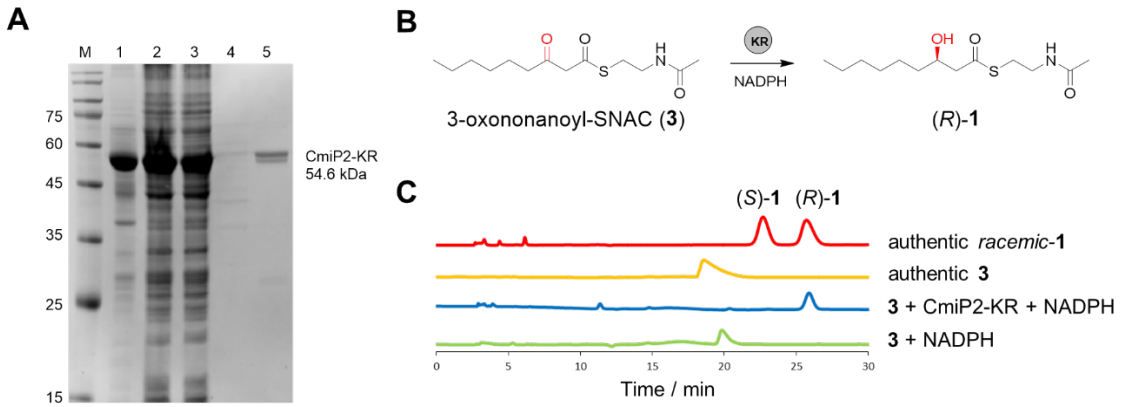

Figure S6. *In vitro* analysis of CmiP2-KR. A) SDS-PAGE analysis of the CmiP2-KR protein. Each lane 1: insoluble fraction of cell-free extract, lane 2: soluble fraction of cell-free extract, lane 3: flow-through, lane 4: elution fraction with 5 mL of lysis buffer B containing 20 mM imidazole, lane 5: elution fraction with 5 mL of lysis buffer B containing 200 mM imidazole. B) A scheme for the CmiP2-KR reaction. C) HPLC analysis of enzymatic reaction. Red line: the synthesized **1** (racemic), orange line: the synthesized 3-oxononanoyl-SNAC (**3**), blue line: the reaction of CmiP2-KR with **3**, green line: the reaction without CmiP2-KR.

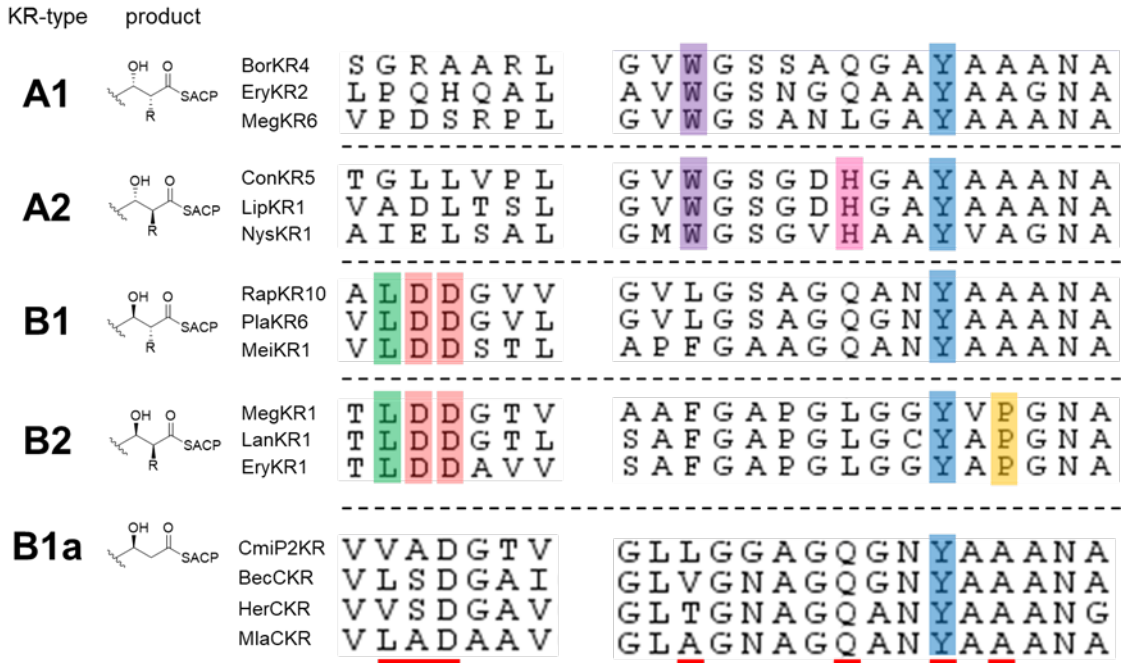

Figure S7. Sequence alignment of the four different stereochemical types of KR domains and CmiP2KR. A catalytic Tyr residue is conserved in all enzymes. A-type KRs for the formation of (*S*)-3-hydroxyacyl thioester product have Trp residue and are classified into A1 ((*R*)-α-substituent) or A2 ((*S*)-α-substituent) depending on whether they have also His residue. B-type KRs have LDD motif and are classified into B1 ((*R*)-α-substituent) or B2 ((*S*)-α-substituent) depending on whether they have also Pro residue. B1a-type KRs for the formation of (*R*)-3-hydroxyacyl thioester product have V or L/A or S/D motif instead of LDD motif in B1 type. B1a-type KRs are a newly proposed family in the present study. BorKR4: KR domain from module 4 of borrelidin PKS (CAE45670), EryKR1, KR2: KR domain from module 1 and 2 of erythromycin PKS (AAV51820), MegKR1, KR6: KR domain from module 1 and 6 of megalomicin PKS (AAG13917, AAG13919), ConKR5: KR domain from module 5 of concanamycin A PKS (AAZ94388), LipKR1: KR domain from module 1 of lipomycin PKS (ABB05102), NysKR1: KR domain from module 1 of nystatin PKS (AAF71766), RapKR10: KR domain from module 10 of rapamycin PKS (CAA60459), PlaKR6: KR domain from module 6 of pladienolide PKS (BAH02269), MeiKR1: KR domain from module 1 of meilingmycin PKS (ADC45586), LanKR1: KR domain from module 1 of lankamycin PKS (BAC76493), BecCKR: KR domain from BecC of BE-14106 PKS (ACO94458), HerCKR: KR domain from HerC of heronamide PKS (AKD43751), MlaCKR: KR domain from MlaC of ML-449 PKS (ACO94486).

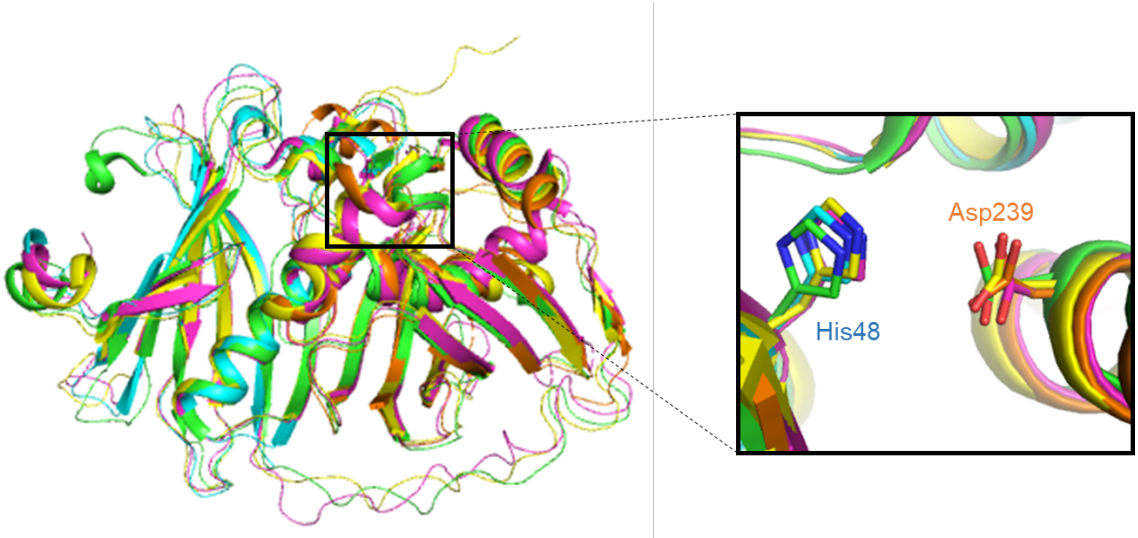

Figure S8. Structural comparison of the fusion-DH and other canonical *cis*-AT PKS DHs. CmiP3-DH_N, CmiP2-DH_C, EryDH4, the DH domain from module 10 of the rifamycin PKS, (RifDH10) and the DH domain from module 1 of the fluvirucin B1 PKS (FruDH1) are shown in cyan, orange, green, magenta and yellow, respectively. Catalytic residues are shown as sticks.

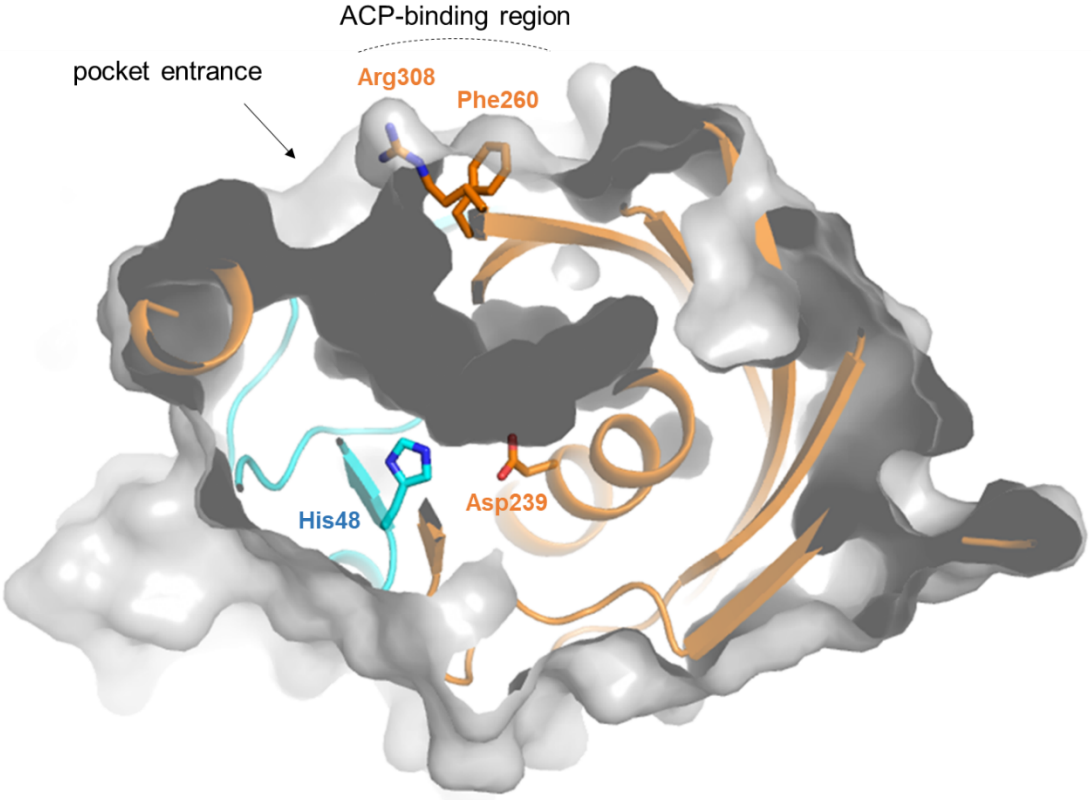

Figure S9. The active-site pocket of the fusion-DH (DH_N: cyan, DH_C: orange). Catalytic residues and proposed ACP-binding residues, which are based on the previous study[10], are shown as sticks.

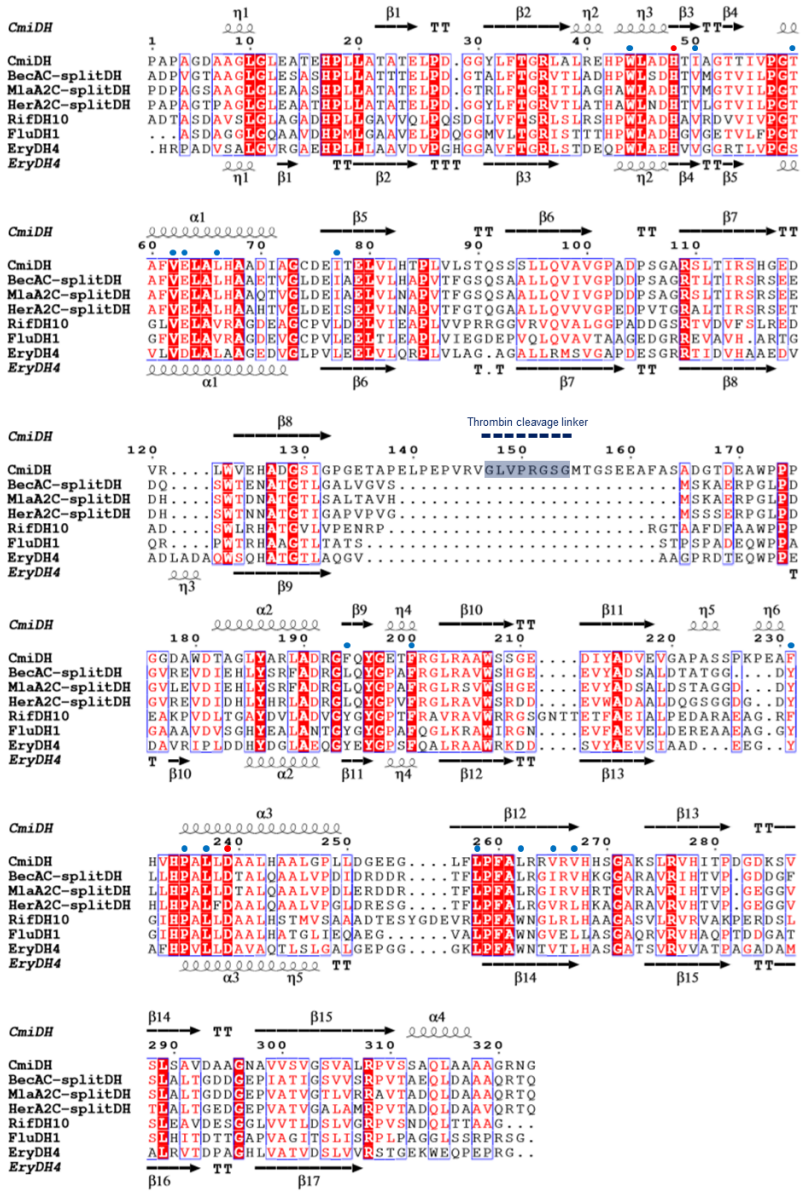

Figure S10. Multiple sequence alignment among split-DH homologs and other canonical *cis*-AT PKS DHs. The sequences of the following enzymes were used for analysis: the split-DH domain from CmiP3 and CmiP2 (CmiDH), the split-DH domain from BecA (ACO94456) and BecC (ACO94458) in BE-14106 biosynthesis (BecAC-splitDH), the split-DH domain from MlaA2 (ACO94484) and MlaC (ACO94486) in ML-449 biosynthesis (MlaA2C-splitDH), the split-DH domain from HerA2 (AKD43769) and HerC (AKD43751) in heronamide biosynthesis (HerA2C-splitDH), RifDH10 (AAC01714) from the rifamycin PKS, FluDH1 (AFS18277) from the fluvirucin B1 PKS and EryDH4 (CAM00064) from the erythromycin PKS. The amino acid sequences were aligned by ClustalX[9] and presented by ESPript[11]. The secondary structural elements of CmiDH and EryDH4 are indicated above and below the sequences, respectively. The artificial linker region of fused CmiDH protein is highlighted in gray box. Catalytic residues are shown as red circles. Residues in the DH_N–DH_C interface are shown as blue circles.

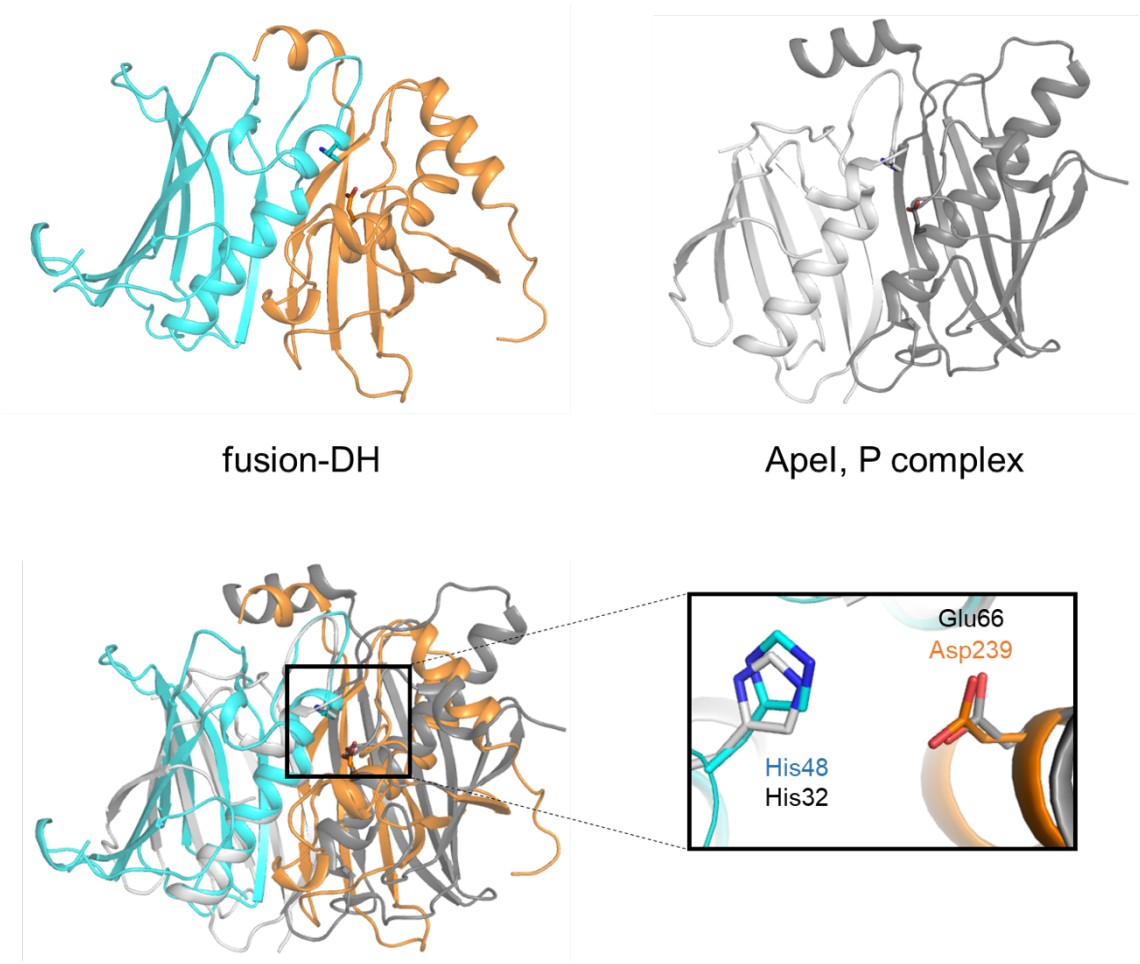

Figure S11. Structural comparison of the fusion-DH (DH_N: cyan, DH_C: orange) with the heterodimer of ApeI (gray) and ApeP (dark gray).

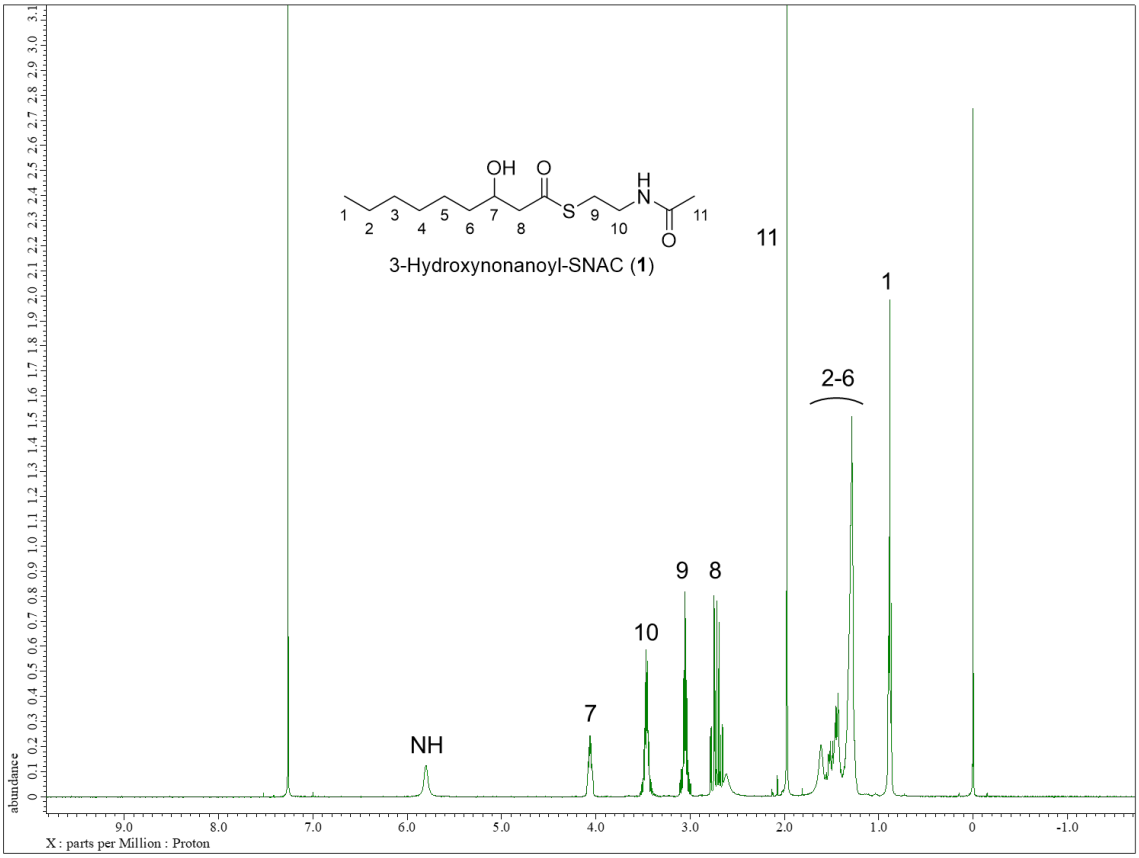

Figure S12. ^{1}H-NMR spectrum of 3-hydroxynonanoyl-SNAC (**1**).

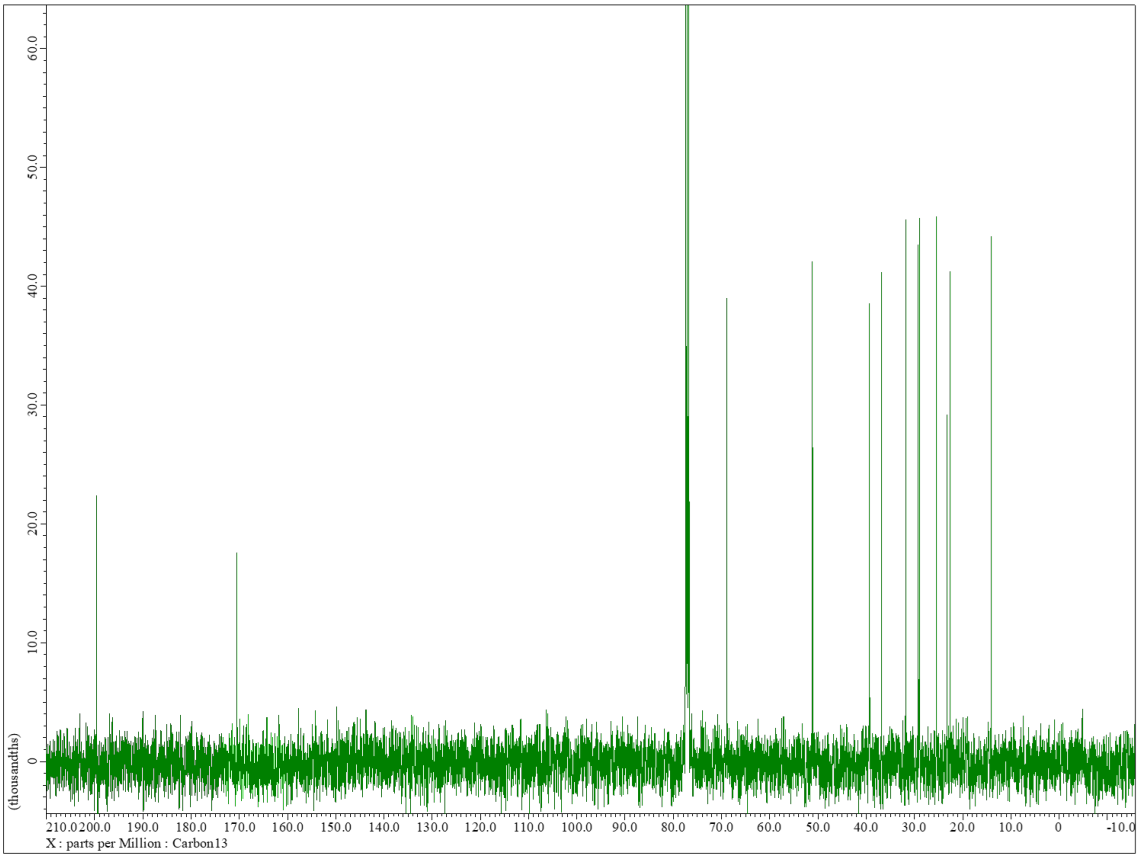

Figure S13. ^{13}C-NMR spectrum of **1**.

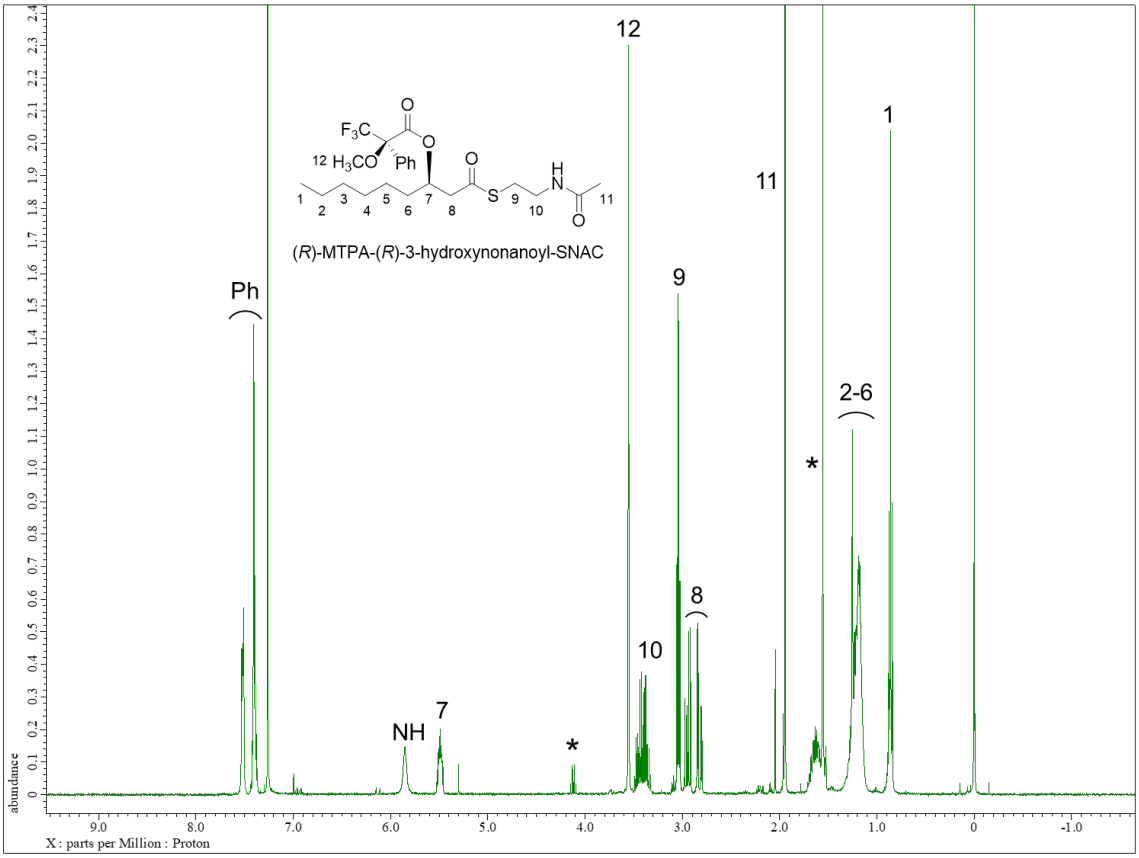

Figure S14. ^{1}H-NMR spectrum of (*R*)-MTPA-(*R*)-**1**. The asterisks indicate impurities.

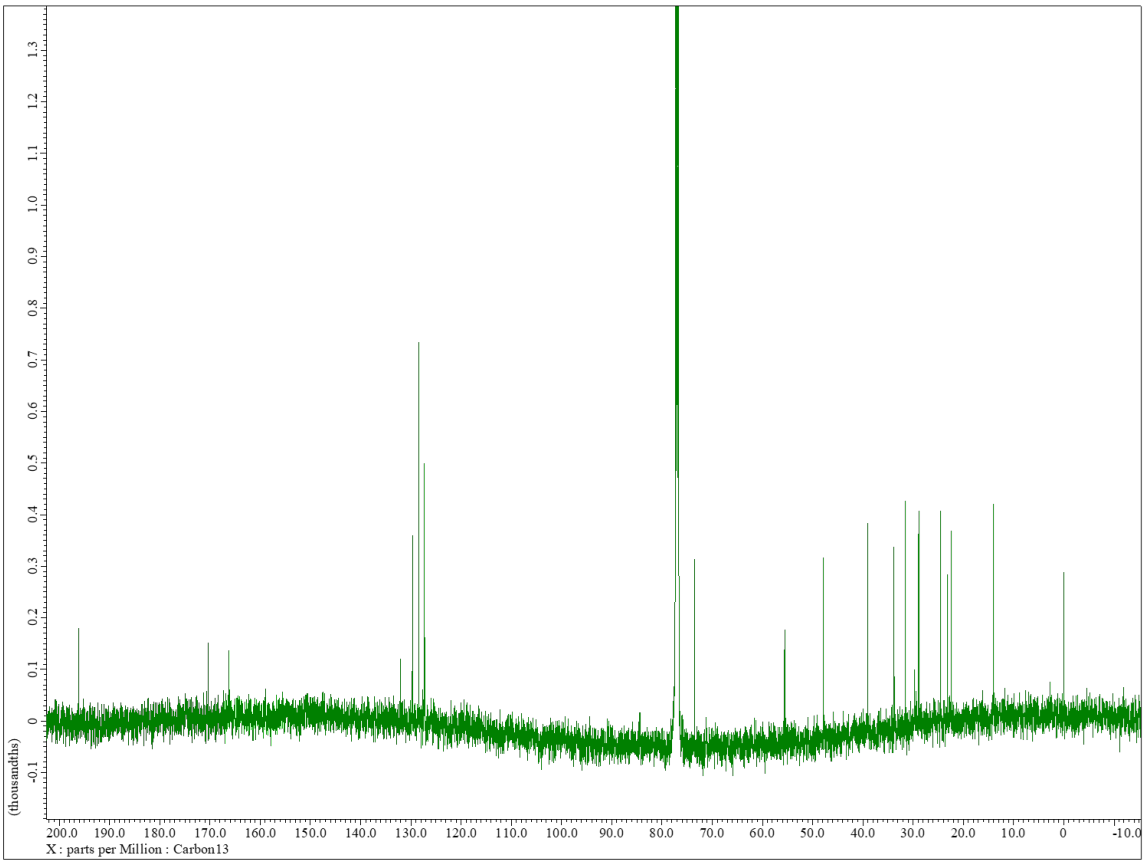

Figure S15. ^{13}C-NMR spectrum of (*R*)-MTPA-(*R*)-**1**.

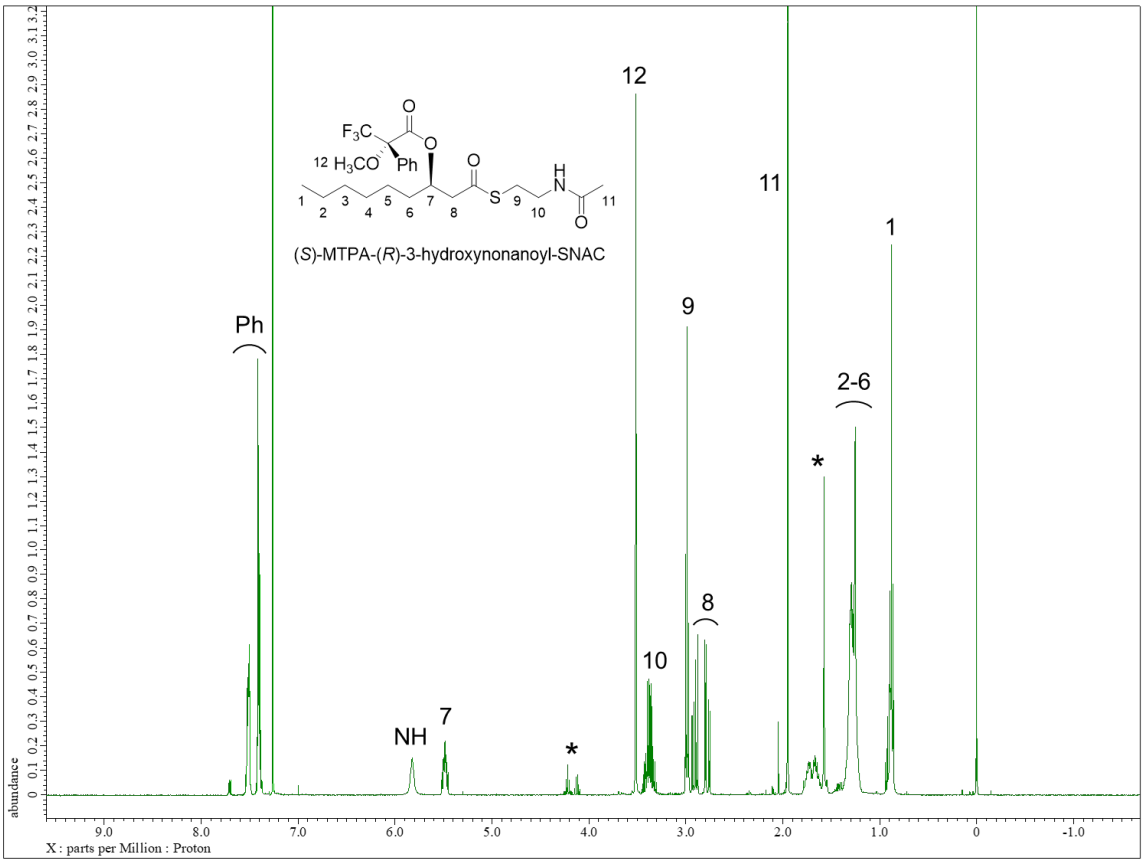

Figure S16. ^{1}H-NMR spectrum of (*S*)-MTPA-(*R*)-**1**. The asterisks indicate impurities.

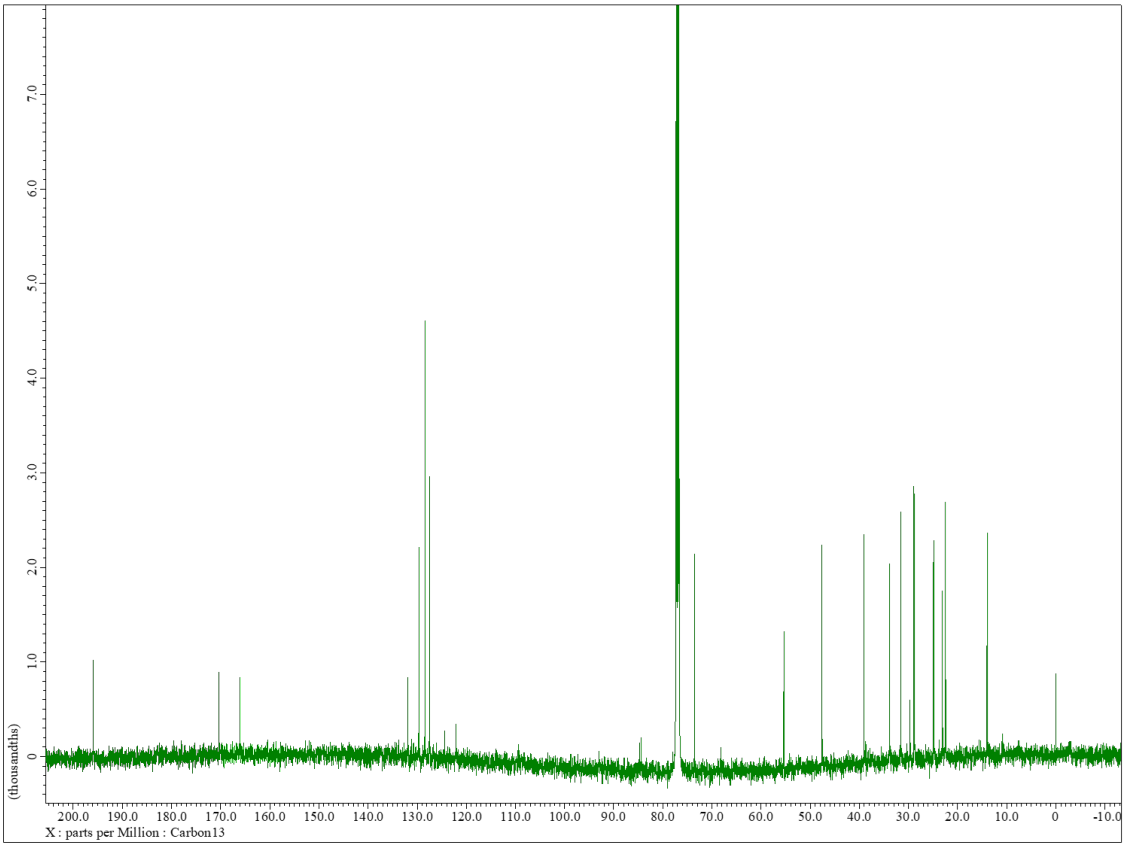

Figure S17. ^{13}C-NMR spectrum of (*S*)-MTPA-(*R*)-**1**.

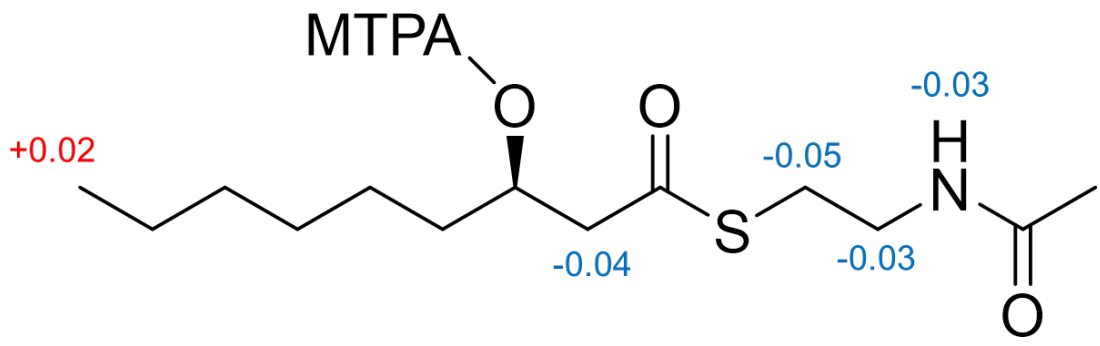

Figure S18. $\Delta\delta_{S\text{-}R}$ values (ppm) of MTPA ester for (*R*)-**1**.

Table S1. Structural data collection and refinement statistics

Dataset	fusion-DH
Data collection statistics	
Beamline	PF BL-5A
Wavelength (Å)	1.00000
Space group	*C*2
Unit-cell parameters	
a (Å)	154.6
b (Å)	66.6
c (Å)	63.0
β (º)	91.0
Resolution (Å)	50.00–2.50
(outer shell)	(2.60–2.50)
Unique reflections	22,153 (2,463)
Redundancy	3.4 (3.5)
Completeness (%)	99.3 (98.6)
R_{merge} (%)	4.9 (60.0)
Mean $\langle I/\sigma(I)\rangle$	14.6 (2.2)
Refinement statistics	
R_{work} (%)	19.5
R_{free} (%)	23.5
Number of polypeptides in the asymmetric unit	2
No. of non-hydrogen atoms	
Protein	3,941
Solvent	33
Average B-factors (Å^2)	
Protein	55.2
Solvent	54.6
r.m.s.d. from ideality	
Bond length (Å)	0.014
Bond angles (º)	1.676
Ramachandran plot	
Favored (%)	96.8
Allowed (%)	3.2
Disallowed (%)	0.0
Molprobity clashscore	2.56

Table S2. List of primers

No	Primer	Sequence (5’ to 3’)
1	fusion-DH F	GCTATTGGCATATGTCTCCGG
2	pCold R	GGCAGGGATCTTAGATTCTG
3	cmiP2 F	CGCATATGACCGGATCCGAGGAGGC
4	cmiP2 R	ACCTGGCTCGAGTCCGGTGCGTGTC
5	cmiP2-KR F	GCAACTTCATATGGCAGCAGGCCGGAACGG
6	cmiP2-KR R	GTCAAGCTTTCAGTGGCCGAGCGCGGAAG

References

1. Chattopadhyay, A., and Salaskar, A. (2000) A practical method for the Reformatsky reaction of aldehydes. *Synthesis 4*, 561-564.

2. Amagai, K., Takaku, R., Kudo, F., and Eguchi, T. (2013) A unique amino transfer mechanism for constructing the β-amino fatty acid starter unit in the biosynthesis of the macrolactam antibiotic cremimycin. *ChemBioChem 14*, 1998–2006.

3. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* (2010), 66, 125–132.

4. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* 66, 22–25.

5. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* 3, 1171–1179.

6. Emsley, P., and Cowtan, K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.

7. Murshudow, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.

8. Chen, V. B., Arendall, W. B. 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Rhichardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21.

9. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.

10. Keatinge-Clay, A. T. (2008) Crystal structure of erythromycin polyketide synthase dehydratase. *J. Mol. Biol. 384*, 941-953.

11. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42, W320-W324.