Supporting Information

Escherichia coli allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product

Kenji Watanabe,1 Kinya Hotta,2 Mino Nakaya,3 Alex P Praseuth,4 Clay C C Wang,5 Daiki Inada,3 Kosaku Takahashi,6 Eri Fukushi,7 Hiroki Oguri,3,8 Hideaki Oikawa3**

1. Research Core for Interdisciplinary Sciences, Okayama University, Okayama 700-8530, Japan. 2. Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore. 3. Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan. 4. BioPharmaceuticals Formulation Development, Allergan, Inc., 2525 Dupont Dr. Irvine, California 92612, USA. 5. Department of Pharmacology and Pharmaceutical Sciences, University of Southern California, Los Angeles, California 90033, USA. 6. Division of Applied Bioscience, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan. 7. The Analytical Core facility of NMR and Mass spectrometry, School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan. 8. Creative Research Initiative 'Sousei' (CRIS), Hokkaido University, Sapporo 001-0021, Japan.

Author Email Address: kenji55@cc.okayama-u.ac.jp

Table of Contents

I. SW-163 Biosynthetic Cluster Isolation. *Streptomyces* sp. SNA15896 was provided from DAIICHI SANKYO Co., Ltd. A BAC library was prepared with pECBAC1[1] linearized by *Bam*H I restriction endonuclease using total DNA of *Streptomyces* sp. SNA15896. A polymerase chain reaction (PCR) was performed with the *Streptomyces* sp. SNA15896 total DNA, obtained as described elsewhere, as a template and a pair of NRPS A domain-specific degenerate primers NRPS-degA3FW and NRPS-degA7RV[2]. The PCR products generated with a set of primers sw_probe forward and reverse (Supporting Information Table 1) were used as templates for preparing NRPS A domain-specific probe[3] using the PCR DIG probe synthesis kit (Roche Diagnostics). The colony blot hybridization screening[4] with the probe using the CDP-Star detection system (Roche Diagnostics) yielded BAC 1-N8 with an open reading frame (ORF) whose sequence is highly homologous to the previously identified SAM (*S*-adenosyl-L-methionine)-dependent methyltransferase Ecm18[5]. Ecm18 catalyzes a unique biotransformation of the disulfide bridge of **2** into a thioacetal bridge in **1**. Since the same thioacetal bridge exists in **4**, we predicted an *ecm18*-like gene would be involved in the SW-163 biosynthesis as well. Further analysis revealed the presence of two NRPS genes, each coding for two modules. Because the four modules were equipped with domains[6] necessary to generate a peptide with the sequence and modifications expected of the monomeric tetrapeptide SW-163 core structure, we were able to predict accurately the involvement of the gene cluster in SW-163 biosynthesis. The BAC covered more than 160 kilobases of contiguous *Streptomyces* sp. SNA15896 DNA. DNA sequence analysis for identifying putative ORFs was conducted with the program FramePlot[7], and the predicted functions of the ORFs were determined by comparison to known proteins using BLAST peptide sequence database search program[8]. The SW-163 biosynthetic gene cluster was found to span over 38 kb of the BAC and contain seven genes predicted to be responsible for the biosynthesis of **8**[9] (*swb1, swb2, swb10, swb11, swb13, swb14* and *swb18*), two genes for the biosynthesis of **9** (*swb6* and *swb7*), six genes for the SW-163 peptide backbone formation and modifications (*swb8, swb9, swb12, swb16, swb17* and *swb20*), two in transcriptional regulation (*swb3* and *19*), a resistance gene (*swb15*) and two

ABC transporter genes (*swb4* and *swb5*) (Figure 2b). The *Nde* I–*Eco*R I fragment of *swb15* was prepared by PCR using the primer pair *swb15* forward and reverse (Supporting Information Table 1) and subsequently cloned into pKW409.

II. General Method for Assembling Multi-Monocistronic Constructs. Initially, ORFs were individually cloned from BAC harboring the biosynthetic gene cluster of interest, and inserted into a pET28b (Novagen)-derived pKW409 as either a PCR-generated *Nde* I–*Eco*R I or *Nde* I–*Xho* I fragment. In pKW409, the unique *Xba* I recognition site was moved to the 5' side of the T7 promoter, and a *Spe* I recognition site was created at the 3' side of the T7 terminator. This assembly method exploits the compatibility of the cohesive ends generated by *Xba* I and *Spe* I digestion. Upon ligation between an *Xba* I and *Spe* I cohesive ends, the recognition site becomes uncleavable by either *Xba* I or *Spe* I restriction endonucleases. The cassette arrangement not only facilitated evaluation of the expression level of each gene individually, but also was necessary for rapid construction of the multi-monocistronic gene assemblies.

III. Construction and Site-Directed Mutagenesis of the Expression Vectors. The common plasmid pKW532 necessary for the *E. coli* production of **10**, the chromophore in **1**, **2** and **14**, was prepared previously[2]. The second plasmid pKW539 (Supporting Information Figure 3b) necessary for the *E. coli* production of **14** was also prepared previously[2]. An alternative second plasmid pKW756 (Supporting Information Figure 3b) used for the *E. coli* production of **14** is functionally identical to pKW539 except pKW756 carries *swb15*, the presumed self-resistance gene from SW-163 biosynthetic gene cluster, in place of *ecm16*, the self-resistance gene from echinomycin biosynthetic gene cluster. pKW756 was constructed to determine the ability of Swb15 to confer resistance to *E. coli* against echinomycin, SW-163s and their hybrid molecule **14** during their productions. Thus, pKW756 carries the genes for the SW-163 self-resistance-conferring factor *swb15* (orange), peptide-forming proteins

ecm1 (green), acyl carrier protein *fabC* (green), peptide-modifying protein *ecm17* (pink) and *Bacillus subtilis* 4'-phosphopantetheinyl transferase *sfp*[10] (orange). Prior to the incorporation of *swb15* into the plasmid, the expression level of *swb15* in *E. coli* was checked (Supporting Information Figure 2b).

The third plasmid pKW755 (Supporting Information Figure 3a) necessary for the *E. coli* production of **15**, which carries *ecm6* and *swb17*, was prepared as follows. First, plasmid pKW677 was prepared for the expression of *swb17* by cloning the *swb17* fragment PCR-amplified from BAC 1-N8 using the primers (Supporting Information Table 1) into pET21c (Novagen)-derived pKW423. In pKW423, the unique *Xba* I recognition site of pET21c was moved to the 3' region of the T7 promoter, and a *Spe* I recognition site was newly created at the 5' region of the T7 terminator. Thus, *swb17* was integrated into an *Xba* I–*Spe* I cassette equipped with its own T7 promoter, ribosome-binding site and T7 terminator in pKW423. The expression level of *swb17* in *E. coli* was checked using pKW423 (Supporting Information Figure 2a). The *Xba* I/*Spe* I double-digested fragment from pKW677 containing *swb17* was cloned into pKW470 to yield pKW684, a pET28-based plasmid carrying *swb17* and *ecm6*. Finally, the *Xba* I/*Spe* I double-digested fragment from pKW684 containing *swb17* and *ecm6* was transferred to pKW407 to create pKW755, a *swb17–ecm6* expression vector containing the RSF origin of replication.

IV. Swb17 Protein Production in *E. coli*. Overexpression of *swb17* were performed as follows: BL21 (DE3) harboring pKW677, a pET28b-based plasmid carrying the *swb17* gene, was grown overnight in 10 ml of 2xYT medium with 50 μg ml^{-1} kanamycin at 37 °C. Each liter of fresh 2xYT medium with 50 μg ml^{-1} kanamycin was inoculated with 5 ml of the overnight culture and incubated at 37 °C until the optical density at 600 nm reached 0.7. Then, expression of each gene was induced with 100 μM IPTG at 15 °C. Incubation was continued for another 24 h, after which cells were harvested by centrifugation at 2,500 x *g* at 4 °C. All subsequent procedures were performed at 4 °C or on ice. Harvested cells were resuspended in lysis buffer [0.1 M Tris-HCl (pH 7.2), 0.3 M NaCl, 10 mM

imidazole, 10 mM $MgCl_2$, 1 mg ml^{-1} lysozyme, 10 μg ml^{-1} DNase I and 10% (v v^{-1}) glycerol]. Cells were disrupted by sonication, and the lysate was clarified by centrifugation at 40,000 x *g*. The supernatant were recovered as the soluble fraction, and was loaded onto lane 2 of a Tris-HCl 4–15% linear gradient gel (Bio-Rad) for SDS–PAGE analysis. The gel was stained with CBB (Supporting Information Figure 2a).

V. Swb15 Self-Resistance Protein Production in *E. coli*. Overexpression of *swb15* were performed as follows: BL21 (DE3) harboring pKW647, a pCold1-based plasmid carrying the *swb15* gene, was grown overnight in 10 ml of 2xYT medium with 100 μg ml^{-1} carbenicillin at 37 °C. Each liter of fresh 2xYT medium with 100 μg ml^{-1} carbenicillin was inoculated with 5 ml of the overnight culture and incubated at 37 °C until the optical density at 600 nm reached 0.7. Then, expression of each gene was induced with 100 μM IPTG at 15 °C. Incubation was continued for another 24 h, after which cells were harvested by centrifugation at 2,500 x *g* at 4 °C. All subsequent procedures were performed at 4 °C or on ice. Harvested cells were resuspended in lysis buffer [0.1 M Tris-HCl (pH 7.2), 0.3 M NaCl, 10 mM imidazole, 10 mM $MgCl_2$, 1 mg ml^{-1} lysozyme, 10 μg ml^{-1} DNase I and 10% (v v^{-1}) glycerol]. Cells were disrupted by sonication, and the lysate was clarified by centrifugation at 40,000 x *g*. The supernatant were recovered as the soluble fraction. Then the soluble fraction containing Swb15 protein was applied to a HisTrap HP column (5 ml; Amersham Biosciences) which was previously equilibrated with the binding buffer [0.1 M Tris-HCl (pH 7.2), 0.3 M NaCl and 10% (v v^{-1}) glycerol] and supplemented with 10 mM imidazole at a flow rate of 1 ml/min. The column was washed with the binding buffer supplemented with10 mM imidazole. Proteins were then eluted using a gradient of 10 to 500 mM imidazole over 100 ml of binding buffer. Fractions containing protein with target molecular weight were pooled and dialyzed against 0.1 M Tris-HCl (pH 7.2), 1 mM dithiothreitol, 1 mM ethylenediaminetetraacetic acid, and 25% (v v^{-1}) glycerol. Protein concentration was estimated with the Bio-Rad protein assay kit using bovine immunoglobulin G as a standard. The purified Swb15 protein

was loaded onto lane 2 of a 4–15% linear gradient polyacrylamide Tris-HCl gel (Bio-Rad) for SDS–PAGE analysis. The gel was stained with CBB (Supporting Information Figure 2b).

Supporting Information Table 1. DNA sequences of the primers used for generating the oligonucleotide probes for isolating the SW-163 biosynthetic cluster and cloning the *swb15* gene.

Primer name	Sequence
sw_probe forward	5'-CCGCACTAGTGATTTACACCAGCGGGAGCACCGGTAAGCCCAAGGGCG-3'
sw_probe reverse	5'-CGCCGGTCCAGTAGAGGCGGTCGCCGTCCCGG-3'
swb15 forward	5'-AGC**CATATG**ACCGTGAGGACGACTGCCGACACTCAGCAGACCG-3'
swb15 reverse	5'-GCTC**GAATTC**CAGCTGCCCAGGTAATCCGCCAGATGCTCGCCCG-3'

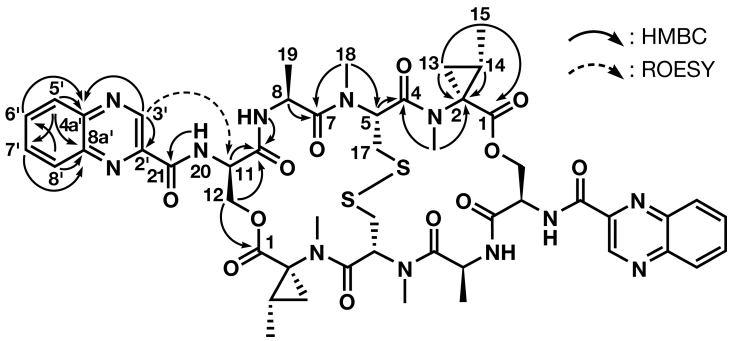
Restriction endonuclease sites are in bold.

Supporting Information Table 2. DNA sequences of the primers used for cloning the *swb17* gene.

Primer name	Sequence
*swb17*_1 forward	5'-GGTGAACGAG**CATATG**ATTCCCCTGTCGTACGCGCAGCGTCGGCTCTGG-3'
*swb17*_1 reverse	5'-GCGCTGCTGGGCCTCGTCGGCGA**TGTACA**GCGAGCCGCCGCAGCC-3'
*swb17*_2 forward	5'-GGCGTTGCCGGGGAGC**TGTACA**TCGGGGGCGAAGGCGTGGCGCGC-3'
*swb17*_2 reverse	5'-CCCTTCTTGTCGGTCG**GAATTC**CACGGGCGTCGCAGCTCCCGAAGCATCCGG-3'

Restriction endonuclease sites are in bold.

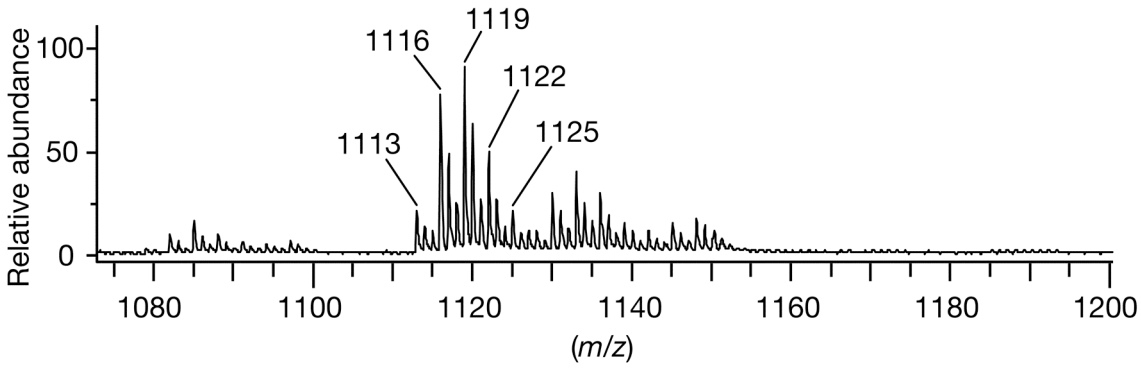
Supporting Information Table 3. ^{1}H and ^{13}C NMR Data of Ecolimycin C 14 in DMSO-d_6. Chemical shifts were recorded as δ (ppm) using DMSO as an internal standard at 2.49 (^{1}H) and 39.7 (^{13}C) ppm. The molecular formula, $C_{50}H_{58}O_{12}N_{12}S_2$, of **14** was established by FAB–HR–MS data [*m*/*z* 1083.3849 $(M+H)^+$, error +3.2 mmu] by positive ionization mode. $[\alpha]_D^{25}$ –58.2° (*c* 0.52, $CHCl_3$)



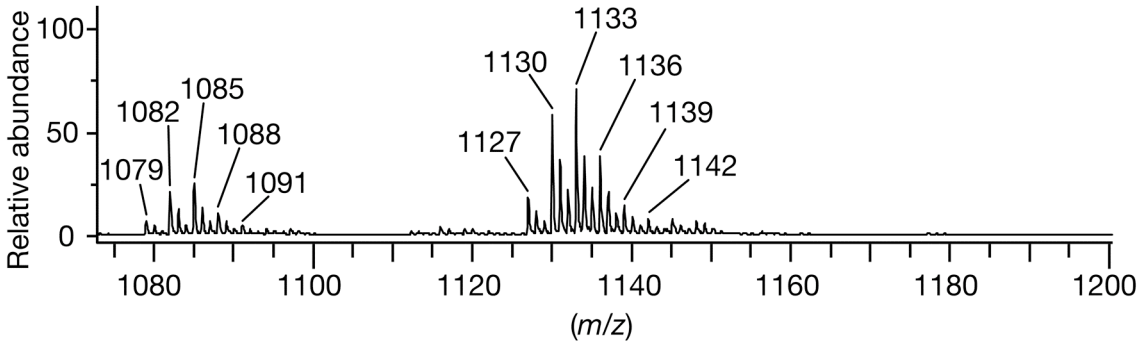
Number	δ H			δ C	
1				170.1	(s)
2				46.4	(s)
4				170.7	(s)
5	6.18	(1H, dd,	J = 2.7, 11.3 Hz)	57.2	(d)
7				171.2	(s)
8	4.47	(1H, dq,	J = 7.0, 7.0 Hz)	46.9	(d)
9	8.21	(1H, s)			
10				168.9	(s)
11	4.85	(1H, dd,	J = 4.1, 9.6 Hz)	50.9	(d)
12	4.29	(1H, d,	J = 10.6 Hz)	64.3	(t)
	4.79	(1H, dd,	J = 4.1, 10.6 Hz)		
13	1.18	(1H, dd,	J = 5.6, 9.5 Hz)	25.1	(t)
	1.51	(1H, dd,	J = 5.6, 7.8 Hz)		
14	2.01	(1H, m)		23.3	(d)
15	1.05	(3H, d,	J = 6.4 Hz)	11.9	(q)
16	3.34	(3H, s)		35.9	(q)
17	2.27	(1H, dd,	J = 2.7, 13.4 Hz)	43.7	(t)
	3.50	(1H, dd,	J = 11.3, 13.4 Hz)		
18	2.71	(3H, s)		30.0	(q)
19	1.26	(3H, d,	J = 7.0 Hz)	15.2	(q)
20	8.37	(1H, d,	J = 9.6 Hz)		
21				162.9	(s)
2'				143.64	(s)
3'	9.50	(1H, s)		143.66	(d)
4a'				143.2	(s)
5'	8.12	(1H, m)		129.0	(d)
6'	8.01	(1H, m)		131.8	(d)
7'	8.02	(1H, m)		132.3	(d)
8'	8.22	(1H, m)		129.3	(d)
8a'				139.4	(s)

Supporting Information Figure 1. HPLC–MS spectra of **(a)** deuterated **3**, **(b)** deuterated **4**, **(c)** deuterated **5**, **(d)** deuterated **6** and **(e)** deuterated **7** produced by *Streptomyces* sp. SNA15896 upon feeding of [*methyl*-D_3]-L-methionine. *Streptomyces* sp. SNA15896 was incubated at 30 °C for 3 days in 50 ml of R2YE medium[3]. Subsequently, 10 ml of the culture was used to inoculate 100 ml of the same medium containing 1 mg ml^{-1} of [*methyl*-D_3]-L-methionine (Cambridge Isotope Laboratories). The culture was grown at 30 °C for another 8 days. The culture was centrifuged to remove the cells, and 100 ml of the supernatant was extracted with ethyl acetate (2 **x** 100 ml). The ethyl acetate extracts were combined and concentrated *in vacuo* to give a dry material. The strain produced the expected compounds as confirmed by HPLC–MS analysis.

Supporting Information Figure 1a: HPLC–MS spectrum of deuterated **3**.

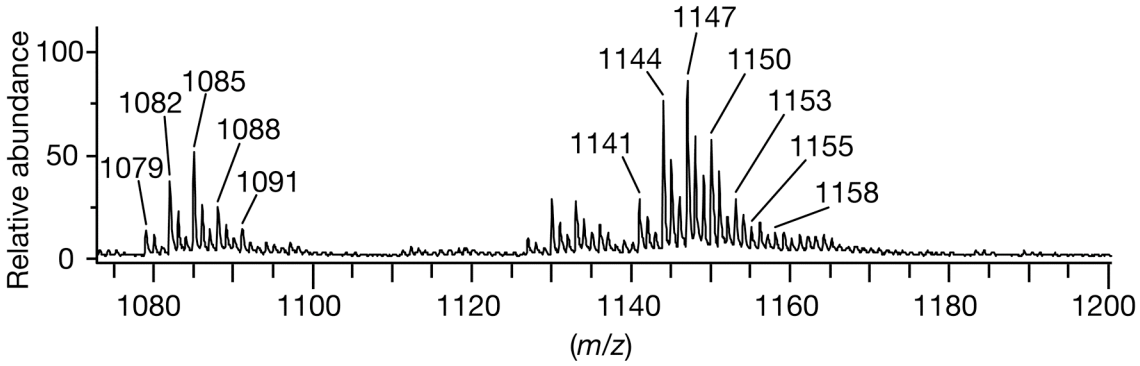


Supporting Information Figure 1b: HPLC–MS spectrum of deuterated **4**.

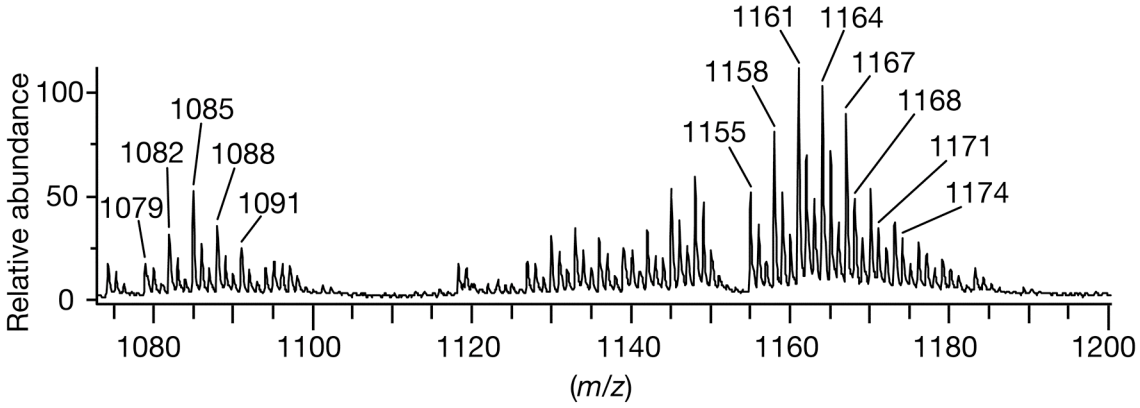


Supporting Information Figure 1 (continued)

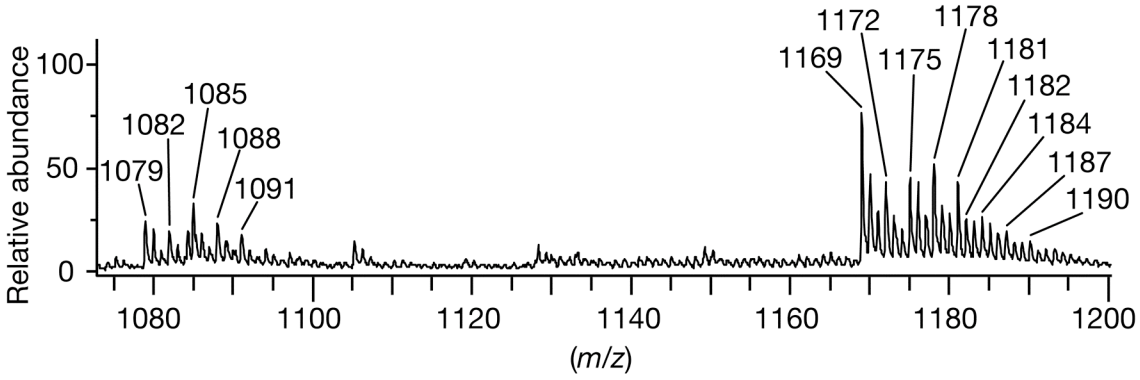
Supporting Information Figure 1c: HPLC–MS spectrum of deuterated **5**.



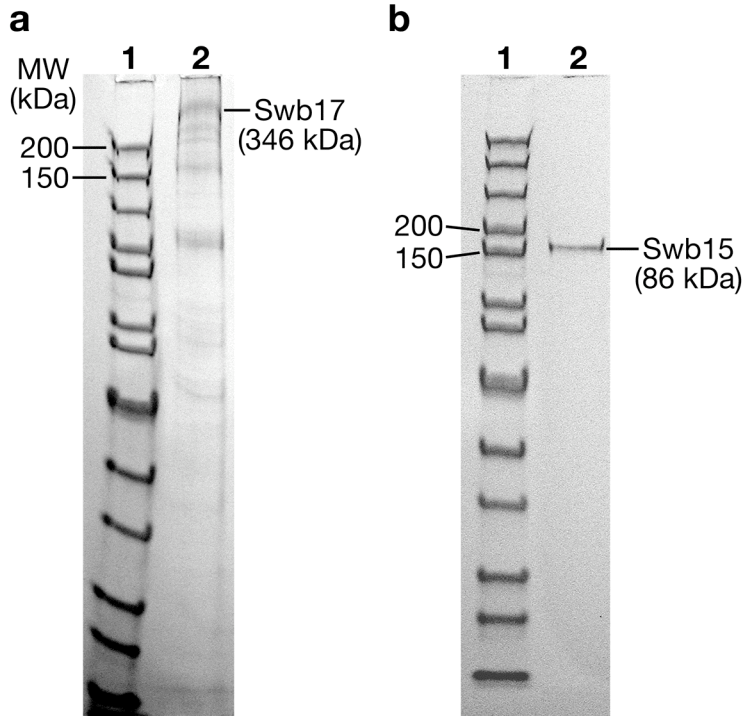
Supporting Information Figure 1d: HPLC–MS spectrum of deuterated **6**.



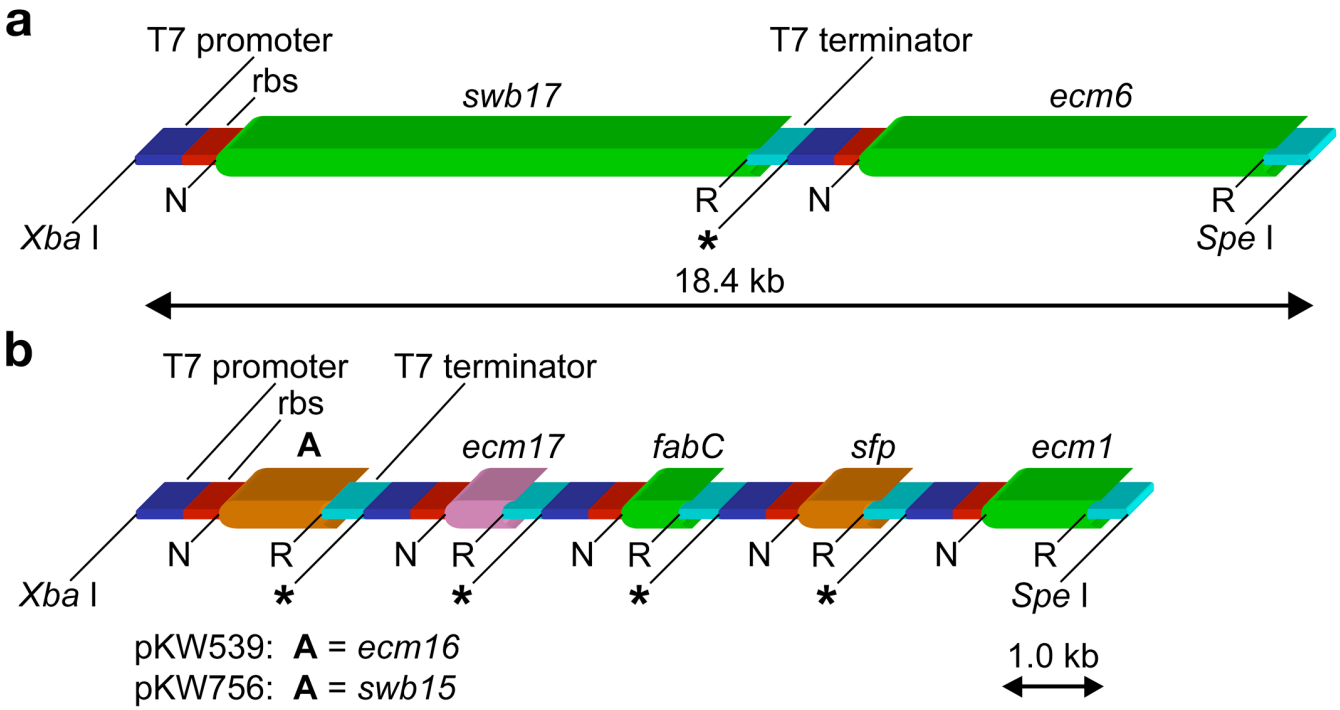
Supporting Information Figure 1e: HPLC–MS spectrum of deuterated **7**.



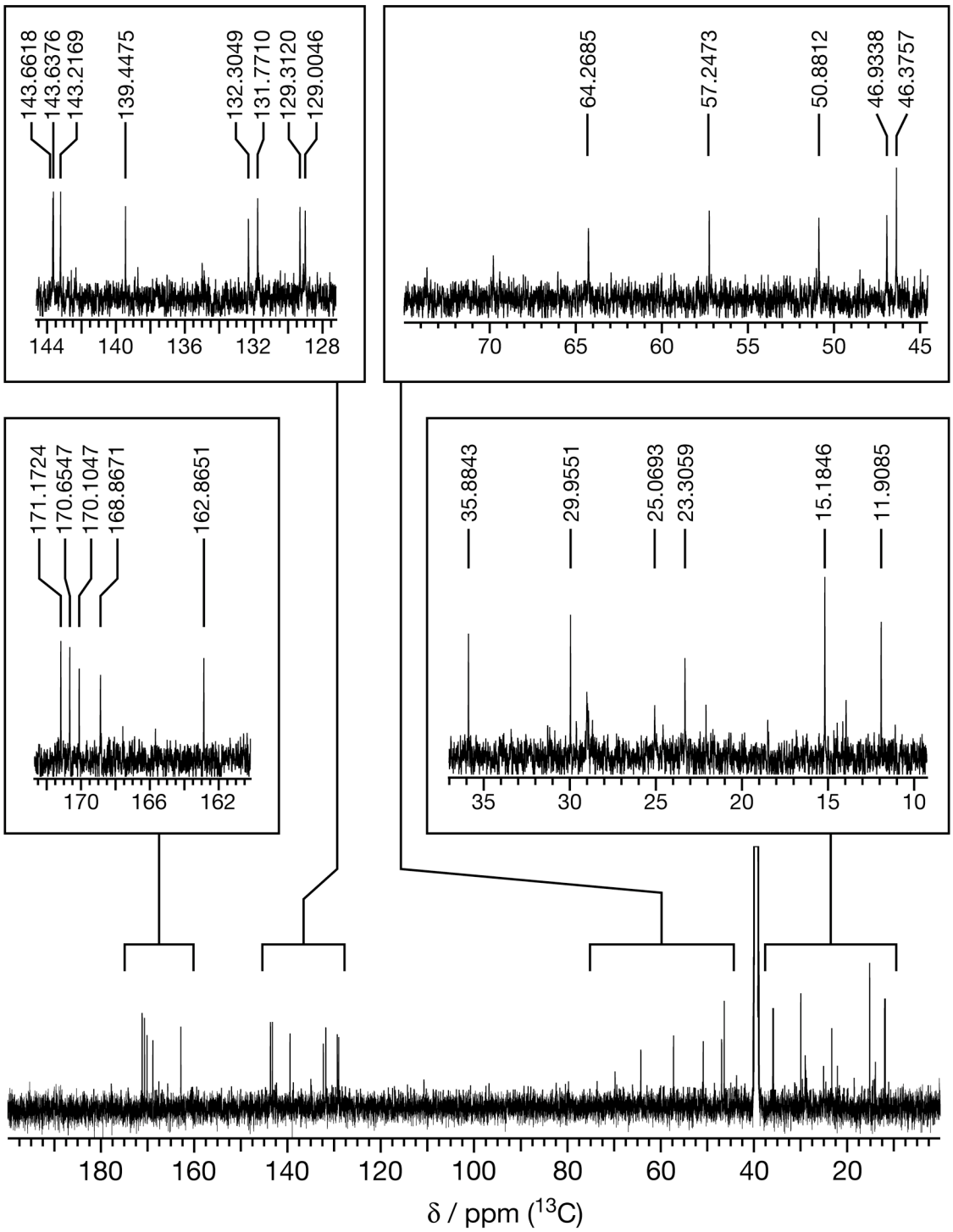
Supporting Information Figure 2. SDS–PAGE analyses of the expression levels of the genes of interest. **(a)** SDS-PAGE analysis of the *swb17* expression level. **Lanes 1:** Molecular weight marker. **Lane 2:** Soluble fraction prepared from the lysate of *E. coli* harboring pKW677 that carries the gene for Swb17 (346 kDa). **(b)** SDS-PAGE analysis of the *swb15* expression level. **Lanes 1:** Molecular weight marker. **Lane 2:** Nickel affinity chromatography-purified Swb15 (86 kDa) produced in *E. coli* harboring pKW647 that carries the *swb15* gene.



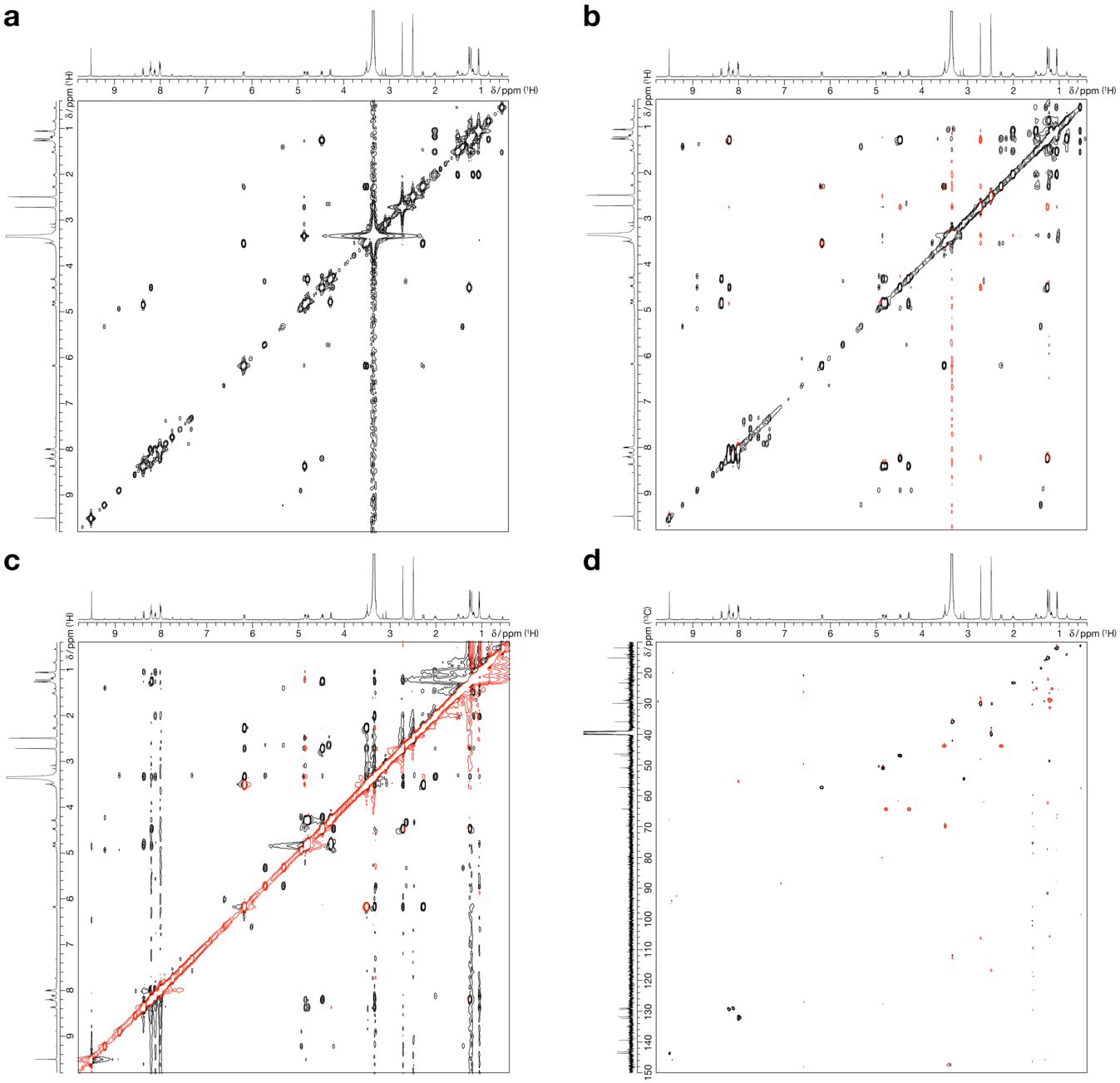
Supporting Information Figure 3. Maps of plasmids employed for the *E. coli* production of ecolimycin C **14**. T7 promoters, T7 terminators and ribosome binding sites are shown in blue, light blue and red, respectively. Abbreviation: rbs - ribosome binding site; N - *Nde* I restriction endonuclease recognition site; E - *Eco*R I restriction endonuclease recognition site; ✱ - product ligation between cohesive *Xba* I and *Spe* I restriction endonuclease recognition sites. **(a)** Map of pKW755 carrying mixed NRPS genes for the biosynthesis of **14**. pKW755 is a pET28b-based plasmid carrying *ecm6* (green), *swb17* (green), a kanamycin resistance gene (not shown), the *lac* repressor gene *lacI* (not shown) and a RSF origin of replication (not shown). pKW755 is used for the production of the NRPS proteins for the biosynthesis of **14**, a hybrid NRP molecule. **(b)** Map of plasmids pKW539 and pKW756, carrying auxiliary genes for the hybrid NRP **14** biosynthesis and the self-resistance against **14**. pKW539 and pKW756 are pCDFDuet-1-based plasmid carrying streptomycin resistance gene (not shown), the *lac* repressor gene *lacI* (not shown) and a CDF origin of replication (not shown). pKW756 is functionally identical to pKW539 except pKW539 carries *ecm16*, the self-resistance gene from echinomycin biosynthetic gene cluster, while pKW756 carries *swb15*, the presumed self-resistance gene from SW-163 biosynthetic gene cluster.



Supporting Information Figure 4. ^{13}C NMR spectrum of **14** produced by the engineered *E. coli* strain.



Supporting Information Figure 5. 2D NMR spectra of **14**. **(a)** ^{1}H-^{1}H COSY, **(b)** TOCSY, **(c)** ROESY and **(d)** HSQC spectra of **14** produced by the engineered *E. coli* strain. The mixing time of all of the 2D spectra was set to 80 ms.



VI. References

(1) Frijters, A. C. J.; Zhang, Z.; van Damme, M.; Wang, G.-L.; Ronald, P. C.; Michelmore, R. W. *Theor. Appl. Genet.* **1997**, *94*, 390-399.

(2) Watanabe, K.; Hotta, K.; Praseuth, A. P.; Koketsu, K.; Migita, A.; Boddy, C. N.; Wang, C. C.; Oguri, H.; Oikawa, H. *Nat. Chem. Biol.* **2006**, *2*, 423-8.

(3) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; The John Innes Foundation: Norwitch, UK, 2000.

(4) Sambrook, J.; Russel, D. W. *Molecular Cloning: A Laboratory Manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2001.

(5) Kinashi, H.; Shimaji, M.; Sakai, A. *Nature* **1987**, *328*, 454-6.

(6) Chiu, H. T.; Hubbard, B. K.; Shah, A. N.; Eide, J.; Fredenburg, R. A.; Walsh, C. T.; Khosla, C. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8548-53.

(7) Ishikawa, J.; Hotta, K. *FEMS Microbiol. Lett.* **1999**, *174*, 251-3.

(8) McGinnis, S.; Madden, T. L. *Nucleic Acids Res.* **2004**, *32*, W20-5.

(9) Reid, D. G.; Doddrell, D. M.; Williams, D. H.; Fox, K. R. *Biochim. Biophys. Acta* **1984**, *798*, 111-4.

(10) Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585-95.