ELECTRONIC SUPPLEMENTARY INFORMATION

Dissecting the biosynthesis of the polyketide alkaloid lydicamycin using a complex metabolic network

Heqian Zhang,[a,b,c,d] Xiaojie Li,[a,b,c,d] Shiyu Pan,[a,b,c] Jiaquan Huang[a,b,c*], and Zhiwei Qin[a,b,c*]

[a]Key Laboratory of Cell Proliferation and Regulation Biology, Ministry of Education, Beijing Normal University, Zhuhai, Guangdong, 519087, China.

[b]Center for Biological Science and Technology, Advanced Institute of Natural Sciences, Beijing Normal University, Zhuhai, Guangdong, 519087, China.

[c]Guangdong Zhuhai-Macao Joint Biotech Laboratory, Beijing Normal University, Zhuhai, Guangdong, 519087, China.

[d]These authors contributed equally to this work

E-mail: z.qin@bnu.edu.cn

Table of contents

1. Experimental

1a. Standard microbiology methods. *S. ginsengnesis* G7 was cultivated at 30°C on ISP2 agar and genetically manipulated on MS agar (See ESI Table 2 for recipes). For sgRNA design and cloning, the Benching web tool (https://benchling.com/) was used to design the sequences, and CasOT[44] to detect off-target risk, sgRNAs with higher specificity were selected. In this study, plasmid pSCBE3-HF was used as the template for target fragment amplification, and 20-bp guide sequence was eventually inserted. Briefly, (1) pSCBE3-HF plasmid was used as the template to PCR-amplify two insertion fragments, with 20-bp guide sequence introduced for gene editing; (2) The pSCBE3-HF plasmid was linearized by restriction enzyme *Xba*I and *Nhe*I; (3) The target fragments were ligated to the *Xba*I-*Nhe*I linearized pSCBE3-HF plasmid to make functional sgRNA expression cassettes by Gibson assembly; (4) The newly constructed vector was replicated in *E. coli* DH5α or *E. coli* DH10B, then confirmed by colony PCR and sequencing[1]. For conjugation and mutant screening, the experiments were undertaken to transformed final vector to *Streptomyces ginsengnesis G7* using *E. coli* ET12567/pUZ8002 as described previously [2]. The ex-conjugants were scraped out from the conjugation plates after overlaying with apramycin (25 μg/mL) for 7 days and transferred to a new plate with apramycin (25 μg/mL) for another 7 days. The spores were harvested and spread onto MS agar without apramycin for an extra 3 days growth to loss the CRISPR plasmid. The resulting single colonies were first spread onto the MS plates with and without apramycin (25 μg/mL) for phenotype screening and then checked by Sanger sequencing.

1b. Standard chemistry methods and materials. Unless stated otherwise all chemicals were supplied by Macklin. All solvents were of HPLC grade or equivalent. Unless otherwise stated samples were analysed by LCMS/MS on a Agilent G6500 UHPLC system attached to a quadrupole time-of-flight (Q-ToF) mass spectrometer. The spray chamber conditions were: nebulizer, 5 L/min; drying gas, 200; sheath gas temperature, 350°C, sheath gas flow, 11 L/min; drying gas on, 5 L/min. The instrument was calibrated using API-TOF Reference Mass Solution Kit according to the manufacturer's instructions. The following analytical LCMS method was used throughout this study: Phenomenex Kinetex C_{18} column (100 × 2.1 mm, 100 Å); mobile phase A: water + 0.1% formic acid; mobile phase B: acetonitrile + 0.1% formic acid. Elution gradient: 0–1 min, 20% B; 1–12 min, 20%–100% B; 12–14 min, 100% B; 14–14.1 min, 100%–20% B; 14.1–17 min, 20% B; flow rate 0.3mL/min; injection volume 10 μL.

1c. Production, purification and structure elucidation of lydicamycin. To isolate compound **1-5**, *S. ginsengnesis* G7 was cultivated on ISP2 agar (2 L; approx. 60 plates) at 30°C for seven days. The agar was sliced into small pieces and extracted once with methanol (2 L) using ultrasonication to improve the extraction. 3 g of the crude extract was filtered and concentrated by evaporation then diluted with 50% methanol (3 mL). This sample was then loaded onto an

open column (4.6 cm × 30 cm) preloaded with Sephadex LH-20 in 50% methanol. Elution was achieved using 50% methanol (600 mL each) and a total number of 35 fractions were finally collected with each in approximate 10 mL. The solvent was removed from each fraction, and the residue dissolved in methanol (1 mL) and this was tested for antibacterial activity using a disk diffusion assay against *B. subtilis*. The antibacterial activity was located exclusively in fraction 9 which was further separated by chromatography over a chromatography using a Phenomenex Gemini semi-prep reversed-phase column (C_{18}, 110 Å, 150 × 4.6 mm) using a Agilent 1290 series HPLC system and eluting with the following isocratic method: mobile phase A: water + 0.1% formic acid; mobile phase B: acetonitrile + 0.1% formic acid; 0–20 min 53% B; flow rate 1 mL/min; injection volume 20 µL. Absorbance was monitored at 285 nm. The resulting solids (**1**, 3.31 mg; **2**, 4.95 mg; **3**, 2.34 mg; **4**, 3.33 mg; **5**, 1.59 mg;) were subjected to analysis by HRMS as described in the main text and the structure determined as shown in Fig. 1 and Fig. 3 of the main paper.

1d. Molecular networking of lydicamycin. A molecular network was created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.1 Da and a MS/MS fragment ion tolerance of 0.1 Da. A network was then created where edges were filtered to have a cosine score above 0.1 and more than 1 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

2. ESI Tables

Table S1. Strains, plasmids and primers used in this work.

Strain name	**Description**	**Source or Ref.**
E. coli ET12567	*dam- dcm- hsdS-*	[3]
E. coli DH10B	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ–	Tsingke Biotechnology Co., Ltd., China
E. coli DH5α	F– φ80*lac*ZΔM15Δ(*lacZYA-argF*) *U169 rec*A1 *end*A1 *hsd*R17 (rK– mK+) *pho*A *sup*E44 λ- *thi*–1 *gyr*A96 *rel*A1	Tsingke Biotechnology Co., Ltd., China
Streptomyces ginsengnesis G7	Wild-type strain	This work
Streptomyces ginsengnesis G7_Δlyd65	TPU-0037-A and Lydicamycin (BGC9) reduced	This work
Streptomyces ginsengnesis G7_Δlyd67	Lydicamycin (BGC9) reduced	This work
Plasmids and ePACs	**Description**	**Source or Ref.**
pSCBE3-HF	pSCBE3 derivate with nCas9 replaced by HF-nCas9	[4]
pUZ8002	Non-transmissible RK2 derivative with a mutation in *oriT*	[3]

Primer name	**Sequence**	**Description**
pSCBE3-HF-promoter-fwd	CGGGGACCTGCAGGTCGACTTGTTCACATTCGAACGGTCTC	pSCBE3-HF template left flank 1

pSCBE3-HF-terminator-rev	TATGTCCTGCGGGTAAATAGGCTACAACTCCTGAGGCTACA	pSCBE3-HF template right flank 2
*lyd65*_spacer-scaffold-fwd	GCATCCAGGCGCTCGCCCAGGTTTTAGAGCTAGAAATAGCAAG	pSCBE3-HF template left flank 2 of *lyd65*
*lyd65*_spacer-promoter-rev	CTGGGCGAGCGCCTGGATGCGGCCACGACTTTACAACAC	pSCBE3-HF template right flank 1 of *lyd65*
*lyd67*_spacer-scaffold-fwd	AGCGAGCAATCGCTCCGCTTGTTTTAGAGCTAGAAATAGCAAG	pSCBE3-HF template left flank 2 of *lyd67*
*lyd67*_spacer-promoter-rev	AAGCGGAGCGATTGCTCGCTGGCCACGACTTTACAACAC	pSCBE3-HF template right flank 1 of *lyd67*
seq-sgRNA-fwd	TGATGCCACGATCCTCGCCCT	Test sgRNA in pSCBE3-HF
seq-sgRNA-rev	GCAAGGTCGCGCTGATTGCTGG	Test sgRNA in pSCBE3-HF
seq- *lyd65*-fwd	GCACCGCACACTCCCCTCAGGCAC	Test *lyd65* base editing in genome
seq- *lyd65*-rev	GCCAGGAGGTGAGCTGCCGGCCGG	Test *lyd65* base editing in genome

seq- *lyd67*-fwd	TTTTATAACTGCGGTTGATTGGTGT	Test *lyd67* base editing in genome
seq- *lyd67*-rev	GTTGGGGGCCGCCGCGGTGCAGGCA	Test *lyd67* base editing in genome

Table S2. The media used in this work.

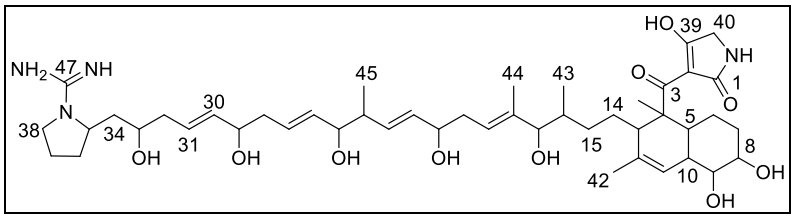
Media	**Recipe (per litre)**	**Water**	**pH**
LB	10 g tryptone 5 g yeast extract 10 g NaCl +/- 20 g agar	Deionised	7.5
MS	20 g soy flour 20 g mannitol 20 g agar	Tap	As made
ISP2	4 g glucose 4 g yeast extract 10 g malt extract 10g agar	Tap	7.2

Table S3. The antibiotics and concentrations used in this work.

Antibiotic	**Final concentration used for selection (μg/ml)**
Apramycin	50/25
Ampicillin	100
Chloramphenicol	25
Nalidixic Acid	25
kanamycin	25

Table S4. The NMR assignment and comparison for **2**, 400 MHz, methanol-d4.

Note. For the reason of the low yield, certain chemical shits were not detected.



Position	δ_C ppm (Ref. 5)	δ_H ppm (Ref. 5)	δ_H ppm (Ref. 6)	δ_C ppm (this study)	δ_H ppm (this study)
1	180.94				
2	102.88				
3	203.95			202.7	
4	54.36				
5	33.53	3.06 (1H, m)	3.12	33.1	3.13 (1H, m)
6	23.44	1.38,1.58 (2H, m)	1.73, 1.42		1.39, 1.60(2H, m)
7	29.57	1.55,1.72 (2H, m)	1.75,1.62		1.62, 1.71(2H, m)
8	71.11	3.43 (1H, m)	3.48		3.48 (1H, m)
9	75.57	3.64 (1H, m)	3.69		3.64 (1H, m)
10	43.75	2.48 (1H, m)	2.54		
11	120.17	4.85 (1H, m)	4.9		
12	141.22				
13	44.97	2.60 (1H, m)	2.66		
14	29.86	1.10, 1.85 (2H, m)	1.89, 1.12		
15	37.86	1.12, 1.86 (2H, m)	1.89,1.19		
16	37.86	1.55 (1H, m)	1.59		
17	83.71	3.57 (1H, d, 8.3)	3.62	83.1	3.57 (1H, d, 4.49)

18	139.66				
19	124.06	5.33 (1H, t, 6.8)	5.37		5.36 (1H, t, 8.29)
20	36.73	2.19, 2.36 (2H, m)	2.38, 2.25	36.8	2.25-2.15, 2.37-2.25(2H, m)
21	73.65	4.03 (1H, m)	4.08	73.7	4.0-4.1 (1H, m)
22	134.32	5.48 (1H, m)	5.53		5.55-5.45 (1H, m)
23	134.89	5.62 (1H, m)	5.64	134.9	5.65-5.55 (1H, m)
24	43.9	2.23 (1H, m)	2.25	44.0	2.25-2.15 (1H, m)
25	77.53	3.84 (1H, t, 6.6)	3.85	77.6	3.84 (1H, m)
26	134.74	5.48 (1H, m)	5.53		5.55-5.45 (1H, m)
27	129.54	5.62 (1H, m)	5.57	129.7	5.65-5.55 (1H, m)
28	42.13	2.21 (1H, m)	2.35	42.1	2.25-2.15 (1H, m)
29	73.11	4.08 (1H, m)	4.06	73.1	4.10-4.01 (1H, m)
30	136.92	5.53 (1H, m)		136.9	5.55-5.45 (1H, m)
31	128.02	5.66 (1H, m)	5.47	127.9	5.65-5.55 (1H, m)
32	41.46	2.20, 2.30 (2H, m)	2.33,2.19	41.4	2.25-2.15, 2.37-2.25(2H, m)
33	69.86	3.66 (1H, m)	3.72	69.9	3.67 (1H, m)
34	41.65	1.63, 1.67 (2H, m)	1.73,1.68	41.7	1.64, 1.67 (2H, m)
35	57.52	4.03 (1H, m)	4.07	57.3	4.10-4.00 (1H, m)
36	32.13	1.89, 2.08 (2H, m)	2.14, 1.95	32.1	1.92, 2.12-2.02 (1H, m)
37	24.09	2.06 (2H, m)	2.09	24.1	2.12-2.02 (1H, m)
38	48.19	3.36, 3.50 (2H, m)	3.53, 3.40	48.2	3.35, 3.48 (2H, m)
39	192.47				
40	50.72	3.52 (2H, m)	3.56	49.7	3.51 (2H, m)

41	17.84	1.40 (3H, s)	1.45	17.7	
42	23.36	1.72 (3H, s)	1.77	22.7	1.76 (3H, m)
43	17.08	0.76 (3H, d, 6.6)	0.81	17.0	0.76 (3H, m)
44	11.94	1.58 (3H, s)	1.63	11.7	1.59 (3H, s)
45	16.42	0.95 (3H, d, 6.8)	0.97	16.6	0.96 (3H, d, 6.7)
46	—	—	1.67	—	—
47	155.77			155.7	

3. ESI Figures

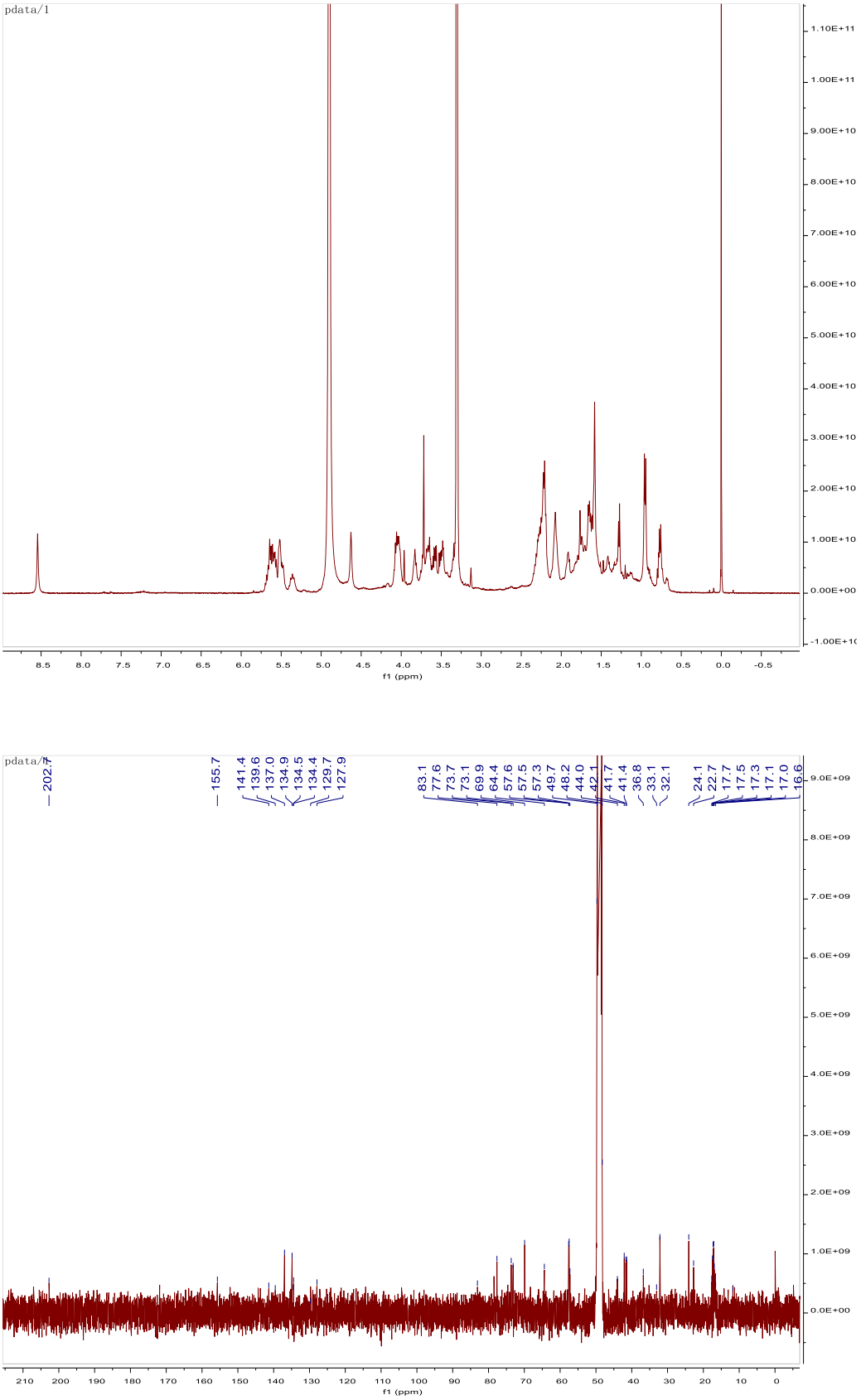


Fig. S1. ^{1}H (up) and ^{13}C NMR (bottom) of **2**, 400 MHz, methanol-d4.

4. ESI References

1. Y. Tong; C. M. Whitford; H. L Robertsen; K Blin; T. S. Jørgensen; A. K. Klitgaard; T. Gren; X Jiang; T. Weber; S. Y. Lee. *Proc Natl Acad Sci U S*, 2019, **116**, 20366.

2. Z. Qin; J. T. Munnoch; R. Devine; N. A. Holmes; R. F. Seipke; K. A. Wilkinson; B. Wilkinson and M. I. Hutchings. *Chem. Sci*, 2017, **8**, 3218.

3. D. J. MacNeil; J. L. Occi; K. M. Gewain; T. MacNeil; P. H. Gibbons; C. L. Ruby; S. J. Danis. *Gene*, 1992, **115**, 119.

4. Z. Zhong; J. Guo, L. Deng; L. Chen; J. Wang; S. Li; W. Xu; Z. Deng; Y. Sun. *bioRxiv*, 2019, https://www.biorxiv.org/content/10.1101/630137v1

5. T. Furumai, k. Eto, T. Sasaki, H. Higuchi, H. Onaka, N. Saitoa, T. Fujitab, H. Naokib, Y. Igarashi. *J Antibiot*, 2002, **55**, 873-880.

6. Y. Hayakawa, N. Kanamaru, N. Morisaki, K. Furihata, H. Seto. *J Antibiot*, 1991, **44**, 288-292.