2014

Iterative Assembly of Two Separate Polyketide Chains by the Same Single-module Bacterial Polyketide Synthase in the Biosynthesis of HSAF

Yaoyao Li
Shandong University

Haotong Chen
University of Nebraska-Lincoln

Yanjiao Ding
Shandong University

Yunxuan Xie
University of Nebraska-Lincoln

Haoxin Wang
Shandong University

See next page for additional authors

Published in final edited form as:
Angew Chem Int Ed Engl. 2014 July 14; 53(29): 7524–7530. doi:10.1002/anie.201403500.

Iterative Assembly of Two Separate Polyketide Chains by the Same Single-module Bacterial Polyketide Synthase in the Biosynthesis of HSAF**

Dr. Yaoyao Li[#],
Key Laboratory of Chemical Biology, School of Pharmaceutical Sciences, Shandong University, Jinan 250100, China

Haotong Chen[#],
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Yanjiao Ding,
Key Laboratory of Chemical Biology, School of Pharmaceutical Sciences, Shandong University, Jinan 250100, China

Yunxuan Xie,
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Dr. Haoxin Wang,
Key Laboratory of Chemical Biology, School of Pharmaceutical Sciences, Shandong University, Jinan 250100, China; State Key Laboratory of Microbial Technology, School of Life Science, Shandong University, Jinan 250100, China

Dr. Ronald L. Cerny [Prof.],
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Dr. Yuemao Shen[*] **[Prof.]**, and
Key Laboratory of Chemical Biology, School of Pharmaceutical Sciences, Shandong University, Jinan 250100, China; State Key Laboratory of Microbial Technology, School of Life Science, Shandong University, Jinan 250100, China

Dr. Liangcheng Du[*] **[Prof.]**
Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

[#] These authors contributed equally to this work.

Keywords

HSAF; *Lysobacter enzymogenes*; natural products; nonribosomal peptide synthetase; polyketide synthase

[**]We thank Dr. Zhongjun Qin for providing *S. coelicolor* strain ZM12. This work was supported in part by NSFC (31329005 and 3120032), the NIH (R01AI097260), Nebraska Research Initiatives, and Program for Changjiang Scholars and Innovative Research
[*]yshen@sdu.edu.cnldu3@unl.edu.

HSAF (**1**) was isolated from the biocontrol agent *Lysobacter enzymogenes* (Figure 1).[1-4] This bacterial metabolite belongs to polycyclic tetramate macrolactams (PTM) that are emerging as a new class of natural products with distinct structural features.[5, 6] HSAF exhibits a potent antifungal activity and shows a novel mode of action.[1-4] The HSAF biosynthetic gene cluster contains only a single-module hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS), although the PTM scaffold is apparently derived from two separate hexaketide chains and an ornithine residue.[1-4] This suggests that the same PKS module would act not only iteratively, but also separately, in order to link the two hexaketide chains with the NRPS-activated ornithine to form the characteristic PTM scaffold. Recently, the Gulder group reported heterologous expression of the ikarugamycin (**4**) biosynthetic gene cluster in *E. coli*,[7] and the Zhang group reported the enzymatic mechanism for formation of the inner 5-memebered ring and demonstrated the polyketide origin of the ikarugamycin skeleton.[8] Ikarugamycin is a *Streptomyces*-derived PTM which has a 5,6,5-tricyclic system (Figure 1). Both the Gulder and Zhang groups showed that a three-gene cluster is sufficient for ikarugamycin biosynthesis. Despite the progress, this iterative polyketide biosynthetic mechanism had not been demonstrated using purified PKS and NRPS. In addition, HSAF has a 5,5,6-tricyclic system, and its gene cluster contains at least six genes.[3] Finally, unlike most PTM compounds, HSAF is produced by a Gram-negative bacterium, *L. enzymogenes*. Here, we report the heterologous production of HSAF analogs in Gram-positive *Streptomyces* hosts, in which the native PKS have been deleted. We also obtained evidence for the formation of the polyene tetramate intermediate in *Streptomyces* when only the single-module hybrid PKS-NRPS gene was expressed. Finally, we showed the *in vitro* production of the polyene tetramate using the individually purified PKS and NRPS. The results provide direct evidence for this iterative polyketide biosynthetic mechanism that is likely general for the PTM-type hybrid polyketide-peptides.

First, we isolated a cosmid clone, Cos4'-1, from the genomic library of *L. enzymogenes* C3, which contains the entire HSAF biosynthetic gene cluster.[2] The gene cluster was then transferred into vectors for expression in *Streptomyces* sp. However, the transformants failed to produce any detectable HSAF or analogs. Subsequently, we made two modifications in the experiments. One was to replace the putative promoter at the 5′-nontranslated region of the PKS-NRPS gene with the *ermE** promoter.[9-11], generating pSETHSAF3 (Figure S1A). The other modification was to use an engineered host, strain SR107 derived from *Streptomyces* sp. LZ35 through deleting its four native PKS gene clusters.[12, 13] This host is expected to provide a relatively "clean" background for the heterologous production of HSAF. We introduced pSETHSAF3 into strain SR107 to generate strain SR107HSAF1 and analyzed the metabolites in the transformant using HPLC. Strain SR107HSAF1 produced approximately seven eminent peaks that were absent in the control strain SR107 (Figure 1). We first focused our attention on the main peak **2** at 18.8 min because it falls in the region that HSAF and analogs would appear.

Compound **2** was isolated (~1 mg/L titer) as yellow powder. HR-ESI-MS gave a quasi molecular ion at *m/z* 495.2837 for $[M+H]^+$ (calculated 495.2853 for $C_{29}H_{38}N_2O_5$). The structure assignments were carried out by the analysis of 1D and 2D NMR data (HSQC, HMBC and ^1H-^1H COSY) (Table S1, Figures S14 – S20 in Supporting Information). The

NMR comparison of **2** with alteramide A (**5**) indicated that the compounds are structurally similar,[14] except for the absence of the hydroxyl group at C3 and the *Z*-geometry for C10,C11 double bond in **2** (Figure 1 and Table S1). The relative configuration of **2** was established by proton couplings and NOE correlations. The large coupling constants (~ 15.0 Hz) between olefinic protons (H8/H9, H21/H22, and H23/H24) led to assignment of the *E*-configuration for the three double bonds, whereas the *Z*-configuration of C10,C11 double bond was deduced from the small coupling constant (11.2 Hz) between H10 and H11. The relative stereochemistry of the bicycle unit was determined from the NOESY experiment, which is identical to alteramide A (**5**).[14] Interestingly, the configuration at C12 in **2** and **5** is opposite to that in **1**. Compounds **2** and **5** may result from an alternative stereospecific cycloaddition that leads to a "wrong" configuration at C12. This consequently may prevent the formation of the 6-membered ring in **1**, as *L. enzymogenes* PKS-NRPS mutant was not able to convert **5** to **1**.

The rest of the eminent peaks (indicated by asterisks in Figure 1) detected at 380 nm appeared unstable, and we were not able to obtain the NMR data. To see if any other isolable HSAF analog was produced in the transformant, we checked the metabolites under other wavelengths. At 318 nm, two peaks were detected at the HSAF region, one at 18.8 min (compound **2**) and the other at 18.6 min (compound **3**) (Figure 1). These two compounds were not produced by the control strain SR107 (at this wavelength, a main peak at 18.6 min was also detected in the control, but showed a different UV-Vis spectrum than compound **3**). Compound **3** was then isolated (~0.4 mg/L titer) for structural determination. It appeared as white powder, with a quasi molecular ion at *m/z* 497.3021 for $[M+H]^+$ (calculated 497.3015 for $C_{29}H_{40}N_2O_5$) as determined by HR-ESI-MS. Comparison of the ^{1}H-NMR spectrum of **3** to that of the previously reported 3-deOH-HSAF readily established the structure of **3** as 3-deOH-HSAF (Figure 1 and S21-23).[15]

The production of compounds **2** and **3** in a *Streptomyces* strain supports the notion that the single-module hybrid PKS-NRPS in pSETHSAF3 is likely sufficient for the assembly of the PTM scaffold. To further prove this point, we transferred pSETHSAF3 into a second *Streptomyces* host that has a completely "clean" background. Strain ZM12 was derived from *S. coelicolor*, probably the best-studied model *Streptomyces*, through deleting all ten native PKS and NRPS gene clusters present in its genome.[16] HPLC showed that this strain produced very few metabolites (Figure S2A). However, upon introducing pSETHSAF3, both compounds **2** and **3** were produced as the main metabolites in strain ZM12, in addition to a number of minor peaks. Since no other PKS-NRPS is present in strain ZM12, our data clearly demonstrated that the biosynthesis of HSAF only requires a single-module hybrid PKS-NRPS. The results also imply that the five domains (KS-AT-DH-KR-ACP) of this PKS module act two separate times to assemble two separate hexaketide chains.

Compounds **2** and **3** lack the 3-hydroxyl group of HSAF (**1**), suggesting that the SD gene (see Figure S1 for the cluster) was not functional in the expression construct pSETHSAF3, where *ermE** promoter was placed in front of the PKS-NRPS gene (Figure S1A). The SD gene encodes the 3-hydroxylase converting 3-deOH-HSAF (**3**) to **1**.[15] We subsequently generated a second expression construct pSETHSAF4, where *ermE** promoter was placed in front of the SD gene (Figure S1B). The construct was introduced into *Streptomyces* sp.

SR107 to generate the transformant strain SR107HSAF2. However, a careful search of the metabolites in this strain did not find any HSAF-like compound (Figure S2B). The reason for this is unclear at this moment; one possibility is that the insertion of the *ermE** cassette in front of SD gene might not lead to the transcription of the PKS-NRPS gene and downstream tailoring genes because SD gene and the rest genes in the cluster do not appear to share the same promoter (Figure S1).[15]

To further demonstrate that the single-module hybrid PKSNRPS is able to assemble two separate polyketide chains and then link them to ornithine, we generated the third *Streptomyces* expression construct pSETHSAF5 that contains only the PKS-NRPS gene under the control of *ermE** promoter (Figure S1C). The construct was introduced into *Streptomyces* sp. SR107 to generate strain SR107PKS/NRPS. HPLC analysis showed that the strain produced three new peaks that were absent in the control strain. The peaks showed absorption λmax around 350-450 nm, suggesting the presence of conjugation systems in all compounds (Figure S3). The peaks were individually collected and analysed by LC-MS/MS. Among them, the peak (compound **6**) with a retention time of 20.8 min gave a quasi molecular ion at *m/z* 475.26 (calculated 475.26 for $[M+H]^+$ for the polyene tetramate **6**) (Figure 2). MS/MS analysis showed fragments of 173.10 and 147.08 that are consistent with the polyene structure. The same polyene was also observed in the recent heterologous expression of ikarugamycin PKS/NRPS.[8] Together, the data support that single module hybrid PKS-NRPS is able to assemble two separate polyketide chains and then link them with ornithine residue to generate the polyene tetramate (**6**).

To obtain direct evidence for this iterative single-module PKS, we expressed the PKS in *E. coli* and purified the 199.8 kDa protein (Figure S4-6). To test its activity, we converted the PKS to its holo form by incubating it with CoA and Svp, a promiscuous 4′-phosphopantetheinyl transferase (PPTase).[17] Due to the huge size of this protein, we treated the PKS with trypsin after the reaction and followed the mass change of the tryptic fragment within the ACP domain of the PKS, to which the PPT moiety and biosynthetic intermediates are covalently linked.[18, 19] Specifically, the trypsin digestion[20] is predicted to release a 26-residue fragment, VKPEQIDADASLNALGLD**S**LLAMELR (the active site serine residue underlined), within the ACP domain.

First, we wanted to confirm that this 26-residue fragment was indeed released from the PKS that was heterologously expressed in *E. coli*. Q-TOF-MS showed that the tryptic fragment of the apo-PKS had *m/z* 928.1374 for $[M+3H]^{3+}$ (calculated 928.1604) (Figure 3), and tandem MS-MS showed that this tryptic fragment had the predicted amino acid sequence (Figure S7). After the correct tryptic PKS fragment was identified, we analyzed the holo-PKS and detected a tryptic fragment of *m/z* 1041.5487 for $[M+3H]^{3+}$ (calculated 1041.5221) and *m/z* 781.4270 for $[M+4H]^{4+}$ (calculated 781.3936) (Figure 3, S8; Table S2). The holo-PKS was also confirmed using phosphopantetheine ejection assay,[21] which detected *m/z* 357.2021 (calculated 357.0891) for the predicted PPT-ejected product (Figure S8). The data showed that the *E. coli* produced PKS was as expected and the enzyme was active.

Next, we sought for potential biosynthetic intermediates that were covalently attached to the ACP domain of the PKS. Upon incubating the holo-PKS with malonyl-CoA, acetyl-CoA

and NADPH, the tryptic fragment released from the PKS showed *m*/*z* 1113.4847 for $[M+3H]^{3+}$ (calculated 1113.5533) and *m*/*z* 835.4001 for $[M+4H]^{4+}$ (calculated 835.4170) (Figure 3, S9; Table S2). This mass change of the ACP fragment is coincident with a hexaketide polyene intermediate attached to the PPT of the PKS (Figure 3). We also varied reaction conditions and searched for other potential biosynthetic intermediates with a varied carbon chain. However, the hexaketide polyene was the only one detected. The results are consistent with that obtained from the heterologous production of HSAF analogs/intermediate in *Streptomyces*. Interestingly, the *in vitro* data suggest that the PKS used malonyl-CoA as both the starter and the extender in the polyketide chain synthesis because a carboxylate appeared present in the intermediate (Figure 3 and S9). To verify this, we generated another PKS expression construct, in which the active site cysteine in the KS domain of this PKS was mutated to alanine (C176A, protein accession number ABL86391) (Figure S10). The point-mutated PKS (mPKS) gene was heterologously expressed in *E. coli*, and the enzyme was purified and confirmed to contain the C176A mutation by Q-TOF-MS detection of the tryptic fragment containing the KS active site, GPSLSIDTA**A**SSSLVAVHLACHSLRR (underlined residue to indicate the C176A mutation), with *m*/*z* 883.16 for $[M+3H]^{3+}$ (calculated 883.45) (Figure S11). Because all the domains, except the KS domain, are still active in this PKS, the synthase is expected to transfer an acyl-CoA (malonyl- or acetyl-CoA) to the ACP but unable to elongate the polyketide chain. So the acyl group from the starter acyl-CoA will be “stuck” on the ACP domain. Indeed the ACP domain of the mPKS was 4′-phosphopantetheinylated by Svp, as evident by the tryptic fragment of *m*/*z* 1041.21 for $[M+3H]^{3+}$ (calculated 1041.52) (Figure S11). Upon incubating the holo-mPKS with both acetyl-CoA and malonyl-CoA in the presence of NADPH, we analyzed the acylated ACP fragment. The tryptic fragment showed *m*/*z* 1069.89 for $[M+3H]^{3+}$ (calculated 1070.19 for malonyl-*S*-ACP, 1055.52 for acetyl-*S*-ACP) and *m*/*z* 802.69 for $[M+4H]^{4+}$ (calculated 802.87 for malonyl-*S*-ACP, 791.86 for acetyl-*S*-ACP) (Figure S12). The results showed that the C176A mutated PKS was active in transferring the malonyl group from malonyl-CoA, but not the acetyl group from acetyl-CoA, to the ACP domain (Figure S12). The result demonstrated that the PKS prefers malonyl-CoA over acetyl-CoA as the starter. It also explains why a carboxylate was observed in the ACP-bound polyketide intermediate (Figure 3).

Finally, we attempted to reconstitute the activity of HSAF PKSNRPS *in vitro* using the individually purified enzymes. The NRPS module (C-A-PCP-TE, 148.6 kDa) was separately expressed and purified.[3] Prior to reconstituting the activity, the purified NRPS was converted to the holo form by incubating with CoA and Svp, followed by incubating with L-ornithine and ATP to form L-ornithine-*S*-NRPS.[3] Since a thioesterase (TE) domain is present in this module, we expect the product(s) be released into the reaction medium.[3, 4, 22] We incubated the polyene-*S*-PKS with L-ornithine-*S*-NRPS and used LC-MS to search for released product from the reconstitution reaction (Figure 4). Total ion chromatography of selected mass range detected a distinct peak (15.48 min under this LC-MS condition, see Supporting Information) that was absent in the control reaction. In HR-ESI-MS analysis, this peak gave *m*/*z* 475.2572 that is coincident with the expected $[M+H]^{+}$ for the polyene tetramate **6** (Figure S13). The result is also in agreement with the *in vivo* data where the production of compound **6** was observed in *Streptomyces* containing the

PKS-NRPS gene alone (Figure 2). The product of the reconstituted PKS and NRPS did not contain the carboxylate from the starter malonyl-CoA, suggesting that a decarboxylation must have taken place after the PKS-linked hexaketide polyene was transferred to the NRPS. Together, the *in vivo* heterologous production data (with whole gene cluster and PKS-NRPS only) and the *in vitro* enzyme assays (with PKS alone and PKS-NRPS reconstitution) support that the PKS module acts iteratively and, together the NPRS module, are sufficient for the synthesis of the framework of PTM.

Based on the results, we proposed a mechanism for the formation of PTM scaffold (Figure 5). First, the five-domain PKS module iteratively catalyzes the formation of a polyene hexaketide. The geometry of the double bonds is unknown, but most likely to be *trans*. This hexaketide is transferred to the four-domain NRPS module, which activates L-orinithine and catalyzes the first amide bond formation, probably between the delta-amino group of L-ornithine and thioester carbonyl group of the hexaketide. This leads to an NRPS-bound polyene-ornithine intermediate. Meantime, the PKS module continues to assemble the second hexaketide, which is subsequently transferred to the NRPS through forming the second amide bond, probably between the alpha-amino group of L-ornithine and the second hexaketide chain. The timing of this transfer is right after the fifth cycle of polyketide chain elongation, in which the newly formed beta-keto group of the second hexaketide chain has not been processed by the KR domain and DH domain. All PTM natural products contain this keto group in the final structure (C25 in HSAF structure, Figure 1).[3, 5, 6, 23] The determining factor for this timing is not known, but probably related to the redox enzymes that are proposed to cooperate with the PKS-NRPS in forming the PTM scaffold.[7, 8, 15, 23] The transfer of the second hexaketide chain leads to a NRPS-bound polyene-ornithine-polyene intermediate. Finally, the intermediate is released from the NRPS by formation of the tetramate moiety, which is through the attack of the nucleophilic alpha carbon of the second hexaketide at the thioester carbonyl carbon of L-ornithine. The presence of the beta-keto group on the second hexaketide makes the alpha-carbon a good nucleophile, which promotes the tetramate formation and product release. This perhaps contributes to the timing of the transfer of this hexaketide to NRPS. In case the beta-keto was fully processed, the second polyene chain would not undergo tetramate formation, which results in an “unproductive process”. In vivo, the polyene tetramate **6** is ultimately converted to the PTM scaffold by tailoring enzymes.[8, 23] Our previous study showed that this NRPS prefers a 12-carbon chain when forming tetramate acyl-ornithine-acyl products in vitro.[3] Thus, the specificity of the domain (most likely the C domain) responsible for transferring the polyketide chain from the PKS to the NRPS is the key determinant for the macrolactam ring size observed in the characteristic PTM framework.

HSAF is the main antifungal factor that the biocontrol agent *Lysobacter enzymogenes* uses to fight against fungal diseases. *Lysobacter* are ubiquitous in the environment, but largely remain untapped for bioactive natural products.[23] HSAF has a potent activity against a broad spectrum of fungi, using a new mode of action. Its chemical structure is distinct from any existing fungicide or antifungal drug. Most interestingly, its biosynthesis involves an iterative PKS mechanism. The hybrid PKS-NRPS has a typical modular organization, including KS-AT-DH-KR-ACP for the PKS module and C-A-PCP-TE for the NRPS

module. There is no obvious remnants of an inactive enoylreductase (ER°) domain or a methyltransferase (CMeT) domain, as seen in several iterative fungal PKS-NRPS with the similar organization, such as PKS for tenellin, lovastain, and compactin.[24-27]. Although this type of iterative PKS-NRPS is commonly seen in fungi, it is not common in bacteria until recent years. Furthermore, PTM-type gene clusters have been identified from numerous bacterial genomes,[5, 6] and the biochemical investigation of this system has only started recently.[3, 7, 8, 28] Here, our studies of HSAF provide direct evidence for this iterative biosynthetic mechanism for bacterial polyketide-peptide natural products. The evidence comes from both the *in vivo* approach — heterologous production of HSAF analogs and the polyene tetramate intermediate in the Gram-positive bacterial hosts with a clean background, and the *in vitro* approach — expression of the ~200 kDa PKS and the ~150 kDa NRPS separately in *E. coli*, purification of the giant enzymes, and reconstitution of the biosynthetic activity. In light of the huge number of uninvestigated PTM-type gene clusters in databases, our studies present here will facilitate the future exploitation of new PTM products.

Experimental Section

Details of experimental procedures, construction of expression vectors, production of HSAF in heterologous hosts, PKS and NRPS gene expression in *E. coli*, protein purification, enzyme reactions and activity assay, and spectroscopic data are included in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Li S, Du L, Yuen G, Harris SD. Mol Biol Cell. 2006; 17:1218–1227. [PubMed: 16394102]

[2]. Yu F, Zaleta-Rivera K, Zhu X, Huffman J, Millet JC, Harris SD, Yuen G, Li XC, Du L. Antimicrob Agents Chemother. 2007; 51:64–72. [PubMed: 17074795]

[3]. Lou L, Qian G, Xie Y, Hang J, Chen H, Zaleta-Rivera K, Li Y, Shen Y, Dussault PH, Liu F, Du L. J Am Chem Soc. 2011; 133:643–645. [PubMed: 21171605]

[4]. Lou L, Chen H, Cerny RL, Li Y, Shen Y, Du L. Biochemistry. 2012; 51:4–6. [PubMed: 22182183]

[5]. Blodgett JA, Oh DC, Cao S, Currie CR, Kolter R, Clardy J. Proc Natl Acad Sci U S A. 2010; 107:11692–11697. [PubMed: 20547882]

[6]. Cao S, Blodgett JA, Clardy J. Organic Letters. 2010; 12:4652–4654. [PubMed: 20843016]

[7]. Antosch J, Schaefers F, Gulder TAM. Angew Chem Int Ed Engl. 2014; 53:3011–3014. [PubMed: 24519911]

[8]. Zhang G, Zhang W, Zhang Q, Shi T, Ma L, Zhu Y, Li S, Zhang H, Zhao Y, Shi R, Zhang C. Angew Chem Int Ed Engl. 2014; 53 in press.

[9]. Bibb MJ, Janssen GR, Ward JM. Gene. 1985; 38:215–216. [PubMed: 2998943]

[10]. Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Bohm I, Mironenko T, Deacon M, Wheatcroft M, Wirtz G, Staunton J, Leadlay PF. J Mol Microbiol Biotechnol. 2002; 4:417–426. [PubMed: 12125822]

[11]. Datsenko KA, Wanner BL. Proc Natl Acad Sci U S A. 2000; 97:6640–6645. [PubMed: 10829079]

[12]. Jiang Y, Wang H, Lu C, Ding Y, Li Y, Shen Y. ChemBiochem. 2013; 14:1468–1475. [PubMed: 23824670]

[13]. Zhao GS, Li SR, Wang YY, Hao HL, Shen YM, Lu CH. Drug Discov Ther. 2013; 7:185–188. [PubMed: 24270382]

[14]. Shigemori H, Bae MA, Yazawa K, Sasaki T, Kobayashi J. Journal of Organic Chemistry. 1992; 57:4317–4320.

[15]. Li Y, Huffman J, Li Y, Du L, Shen Y. MedChemComm. 2012; 9:982–986.

[16]. Zhou M, Jing X, Xie P, Chen W, Wang T, Xia H, Qin Z. FEMS Microbiology Letters. 2012; 333:169–179. [PubMed: 22670631]

[17]. Sanchez C, Du L, Edwards DJ, Toney MD, Shen B. Chemistry & Biology. 2001; 8:725–738. [PubMed: 11451672]

[18]. Hitchman TS, Crosby J, Byrom KJ, Cox RJ, Simpson TJ. Chem Biol. 1998; 5:35–47. [PubMed: 9479478]

[19]. Crosby J, Byrom KJ, Hitchman TS, Cox RJ, Crump MP, Findlow IS, Bibb MJ, Simpson TJ. FEBS Lett. 1998; 433:132–138. [PubMed: 9738947]

[20]. Kayser JP, Vallet JL, Cerny RL. J Biomol Tech. 2004; 15:285–295. [PubMed: 15585825]

[21]. Dorrestein PC, Bumpus SB, Calderone CT, Garneau-Tsodikova S, Aron ZD, Straight PD, Kolter R, Walsh CT, Kelleher NL. Biochem. 2006; 45:12756–12766. [PubMed: 17042494]

[22]. Du L, Lou L. Nat Prod Rep. 2010; 27:255–278. [PubMed: 20111804]

[23]. Xie Y, Wright S, Shen Y, Du L. Nat Prod Rep. 2012; 19:1277–1287. [PubMed: 22898908]

[24]. Yakasai AA, Davison J, Wasil Z, Halo LM, Butts CP, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. Journal of the American Chemical Society. 2011; 133:10990–10998. [PubMed: 21675761]

[25]. Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR. Science. 1999; 284:1368–1372. [PubMed: 10334994]

[26]. Ma SM, Li JW, Choi JW, Zhou H, Lee KK, Moorthie VA, Xie X, Kealey JT, Da Silva NA, Vederas JC, Tang Y. Science. 2009; 326:589–592. [PubMed: 19900898]

[27]. Abe Y, Suzuki T, Ono C, Iwamoto K, Hosobuchi M, Yoshikawa H. Mol Genet Genomics. 2002; 267:636–646. [PubMed: 12172803]

[28]. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. Nat Commun. 2014; 4 in press.

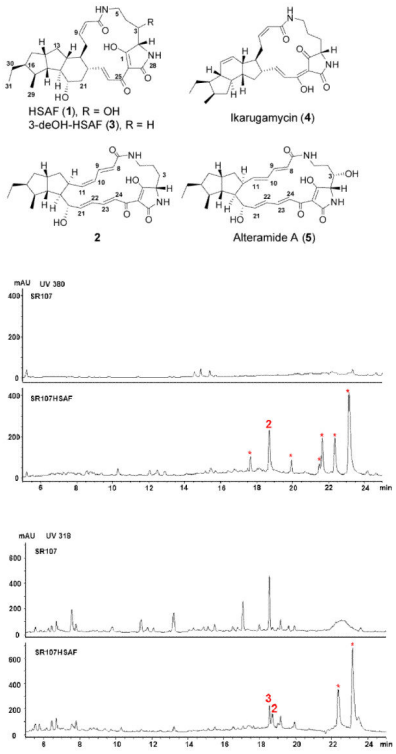


Figure 1.

Chemical structures of HSAF and other PTM analogs and HPLC analysis of the HSAF analogs produced in *Streptomyces*. SR107, metabolites from non-transformed *Streptomyces* sp. LZ35 strain SR107, in which four native PKS gene clusters had been deleted; SR107HSAF, metabolites from strain SR107HSAF1 transformed with the PKS-NRPS biosynthetic gene cluster under the control of *ermE** promoter. The first two HPLC profiles show the metabolites detected at 380 nm, and the bottom two show the metabolites detected at 318 nm. The asterisks indicate peaks that were absent in the control.

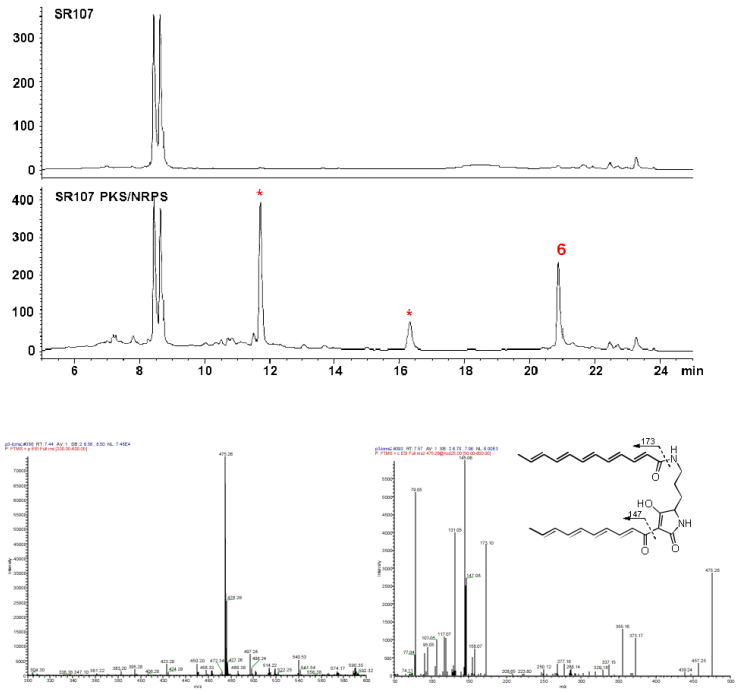


Figure 2.

Production of the polyene tetramate (**6**) in *Streptomyces* transformed with the PKS-NRPS only. Top figure, HPLC analysis; bottom figure, MS analysis of **6**, with the full ESI-MS at the left and MS/MS analysis at the right. SR107, metabolites from non-transformed *Streptomyces* sp. LZ35 strain SR107; SR107 PKS/NRPS, metabolites from the strain transformed with only the PKS-NRPS gene under the control of *ermE** promoter. The metabolites were detected at 380 nm. The asterisks indicate peaks that were absent in the control.

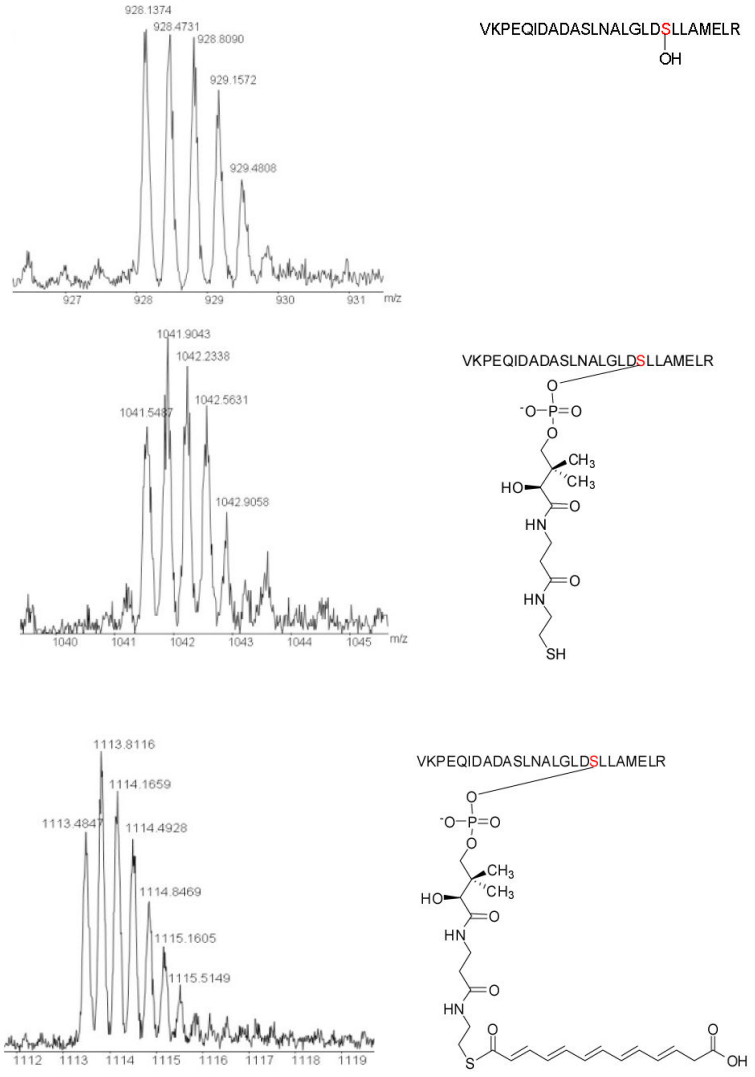


Figure 3.

Q-TOF MS detection of the 26-residue tryptic PKS fragment that harbors the serine active site of the ACP domain. Top panel, the tryptic fragment from the PKS produced in *E. coli*, with *m*/*z* 928.1374 for $[M+3H]^{3+}$ (calculated 928.1604); mid panel, the tryptic fragment from the 4′-phosphopantetheinylated PKS, with *m*/*z* 1041.5487 for $[M+3H]^{3+}$ (calculated 1041.5221); bottom panel, the tryptic fragment from acylated PKS, with *m*/*z* 1113.4847 for $[M+3H]^{3+}$ (calculated 1113.5533), showing the hexaketide polyene synthesized by the single-module PKS.

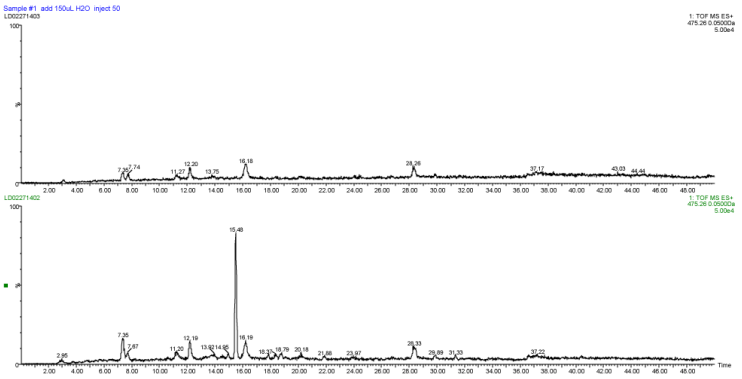


Figure 4.

Analysis of the in vitro reconstituted PKS-NRPS. Total ion chromatography of the mass (475.2597) for compound **6** is shown for the control reaction (top panel) and the PKS-NRPS reaction (bottom). The peak at 15.48 min in the PKS-NRPS reaction gave *m*/*z* 475.2572 in HR-ESI-MS.

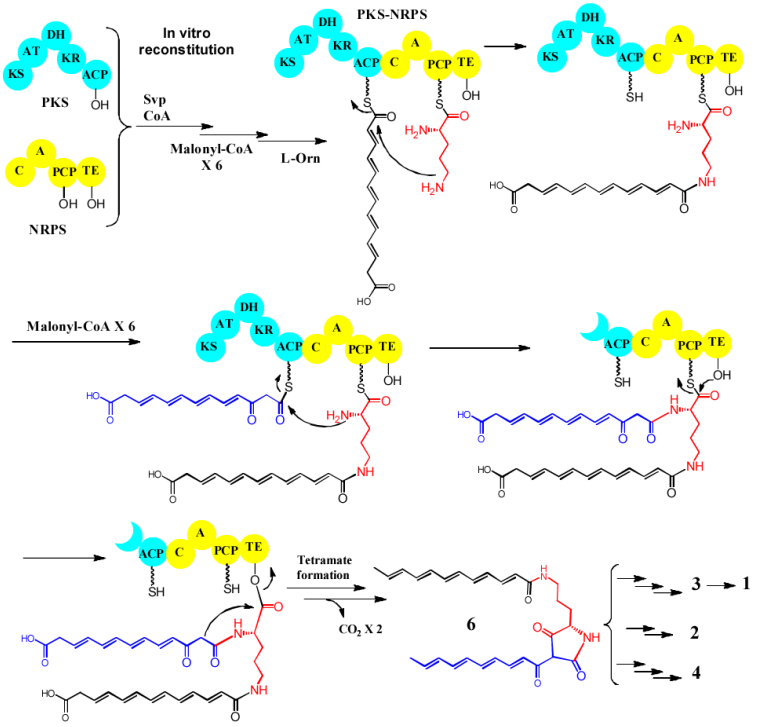


Figure 5.

A proposed biosynthetic mechanism for the iterative bacterial PKS-NRPS in the assembly of the scaffold of HSAF and analogs.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Iterative Assembly of Two Separate Polyketide Chains by the Same Single-Module Bacterial Polyketide Synthase in the Biosynthesis of HSAF**

Yaoyao Li, Haotong Chen, Yanjiao Ding, Yunxuan Xie, Haoxin Wang, Ronald L. Cerny, Yuemao Shen, and Liangcheng Du**

anie_201403500_sm_miscellaneous_information.pdf

Supporting Information

1. Strains, Vectors, Chemicals and Molecular Biology Agents

Escherichia coli DH5α and EPI300 (Epicenter Biotechnologies, Madison, WI) were used as the hosts for general plasmid DNA propagation. *E. coli* SG13009 (pREP4) were used for protein expression. *E. coli* BW25113/pKD46 was used for Red recombination. *E. coli* ET12567/pUZ8002 was used as the conjugal strain. *Streptomyces* sp. LZ35 was isolated from the intertidal soil collected at Jimei, Xiamen, China.[12] Strain SR107 was derived from *Streptomyces* sp. LZ35 by deletion of four PKS gene clusters.[13] Strain ZM12 was derived from *S. coelicolor* by deletion of all ten native PKS and NRPS gene clusters present in its genome.[16] Strains SR107 and ZM12 were used as the hosts for the heterologous production of HSAF. Vector pANT841 was used for general cloning, vector pQE60 was used for protein expression in *E. coli* SG13009 (pREP4), and vector pSET152 was used for integration of the HSAF gene cluster into *Streptomyces* hosts. Chemicals were purchased from Sigma or Fisher Scientific. PCR primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Kits for plasmid preparation and DNA extraction were from Qiagen (Valencia, CA). Standard molecular biology methods were used for all other DNA manipulations.

2. Construction of HSAF expression vectors and heterologous production in *Streptomyces*

To subclone the HSAF gene cluster into the integrative vector pSET152, two restriction sites, *Acl*I and *Bam*HI, were introduced into the vector by insertion of two DNA fragments (Figure S1). The two fragments (1.4 and 1.7 kb) were amplified from the cosmid clone, Cos4'-1, which contains the entire HSAF gene cluster isolated from *Lysobacter enzymogenes* C3.[2] The following two pairs of primers were used for the PCR:

HSAF-LF, Left-*Bgl*II-Forward: 5'-GAA GAT CTC GGG AAA CTC CGA GGC AGA AAT-3'

HSAF-LR, Left-*EcoR*I-Reverse: 5'-AGG AAT TCC GGT GAT GCA AGG GCT GTC GA-3'

HSAF-RF, Right-*EcoR*I-Forward: 5'-AGG AAT TCA ACA CCT GGA GTT CTT CAG CAT CA-3'

HSAF-RR, Right-*Bgl*II-Reverese: 5'-GAA GAT CTA AAT CAG CGC CGG TGG CGT TG-3'

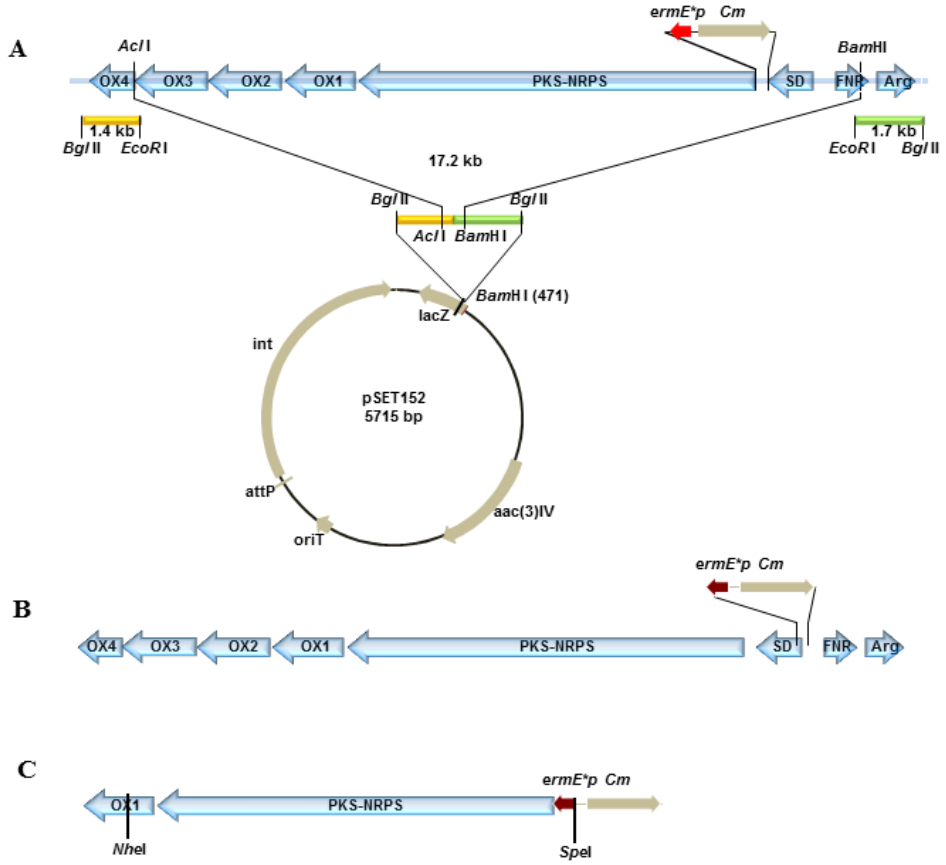


Figure S1. Strategy for constructing the HSAF expression vectors for the production of HSAF analogs in *Streptomyces* hosts. A, construct pSETHSAF3 that resulted in the recombinant *Streptomyces* strains SR107HSAF1 and ZM12HSAF1; B, construct pSETHSAF4 that resulted in the recombinant *Streptomyces* strain SR107HSAF2; C, construct pSETHSAF5 resulting in the recombinant *Streptomyces* strain SR107PKS/NRPS.

The resulting PCR products were digested with *Bgl*II and *EcoR*I, respectively, and ligated to the *BamH*I-digested pSET152 to create pSETHSAF1. Then, a 17.2 kb fragment harboring the bulk of the HSAF gene cluster was cut off from Cos4'-1 by digestion with *Acl*I and *Bam*HI and inserted into the same sites of pSETHSAF1 to yield pSETHSAF2 (Figure S1).

The λ Red mediated recombination in *E. coli* BW25113/pKD46 [11] was used to replace the native promoter of the PKS-NRPS gene or of the SD gene with *Streptomyces ermE** promoter.[9, 10] About 100 ng of pSETHSAF2 was transformed into *E. coli* BW25113/pKD46 by electroporation at 2500 V, 25 μF, and 200 Ohms. Clones were selected on solid LB agar with 100 μg/ml ampicillin and 30 μg/ml apramycin. The transformants were cultured in LB medium at 30 ℃ and induced by addition of 10 mM L-arabinose for the expression of the *red* genes. Two fragments including the chloramphenicol-resistance cassette (*Cm*) and the *ermE** promoter were amplified from pUC19-Cm-PermER* (unpublished data) using the following two pairs of primers, respectively:

ER2HSAFF: 5'-gaa cca gtt tgt tcc aga tca cgg ccg gaa cgc gtt ccg ACA GCT TAT CAT CGA ATT TC-3'

ER2HSAFR: 5'-gcg gca gcc gat gcc gat gat ggc gat gcg gtc ctc cat ATG TCC GCC TCC TTT GGT CG-3'

CmER2SDF: 5'-cgc cgc ggc cta gcc gcg aat att cgc cgg cat tat ctg ACA GCT TAT CAT CGA ATT TC-3'

CmER2SDR: 5'-agc acg cgt ttc caa agt gtg ctc tga taa tcc gtt cat ATG TCC GCC TCC TTT GGT CG-3'

(note: the bases in lower case are the sequence for Red recombination, and the bases in upper case are for amplification of the target genes)

Each of the fragments (~ 200 ng) was electroporated into *E. coli* BW25113/pKD46/pSETHSAF2 to generate pSETHSAF3 and pSETHSAF4, respectively. Transformants were selected for resistance to chloramphenicol at 25 μg/ml. In addition, the 10 kb *Spe*I /*Nhe*I fragment from pSETHSAF3 was inserted into the *Xba*I site of pSET152 to create pSETHSAF5, which only contained the PKS/NRPS gene with the *ermE** promoter. The constructs pSETHSAF3 - 5 were individually transformed into *E. coli* ET12567/pUZ8002, which were used to mate with *Streptomyces* sp. SR107 and ZM12. This intergeneric conjugation resulted in the transfer of three different HSAF expression constructs into *Streptomyces* hosts, construct pSETHSAF3 leading to the recombinant strains SR107HSAF1 and ZM12HSAF1, construct pSETHSAF4 to strain SR107HSAF2, and construct pSETHSAF5 to strain SR107PKS/NRPS.

3. Production and analysis of the metabolites in *Streptomyces* strains

For production of secondary metabolites, all *Streptomyces* strains were inoculated on petri dishes containing the Fermentation Medium (soluble starch 25 g, soybean 15 g, yeast extract 2 g, agar 20 g, pH

7.2, 40 ml) and cultivated at 28 °C for 11 days. The whole solid cultures were diced and extracted with AcOEt/MeOH/AcOH (80 : 15 : 5, v/v/v) at room temperature, and the crude solution was concentrated under reduced pressure. The dried extract was dissolved in 2 ml methanol, and a 20 μl aliquot of each extracts was analyzed by HPLC (Agilent 1200, ZORBAX Eclipse XDB-C18, 4.6×250 mm, 5 μ). For SR107HSAF1-2 and ZM12HSAF1, chromatographic conditions were as follows: solvents-A, water-0.04% trifluoroacetic acid (TFA); solvent-B, acetonitrile-0.04% TFA; solvent gradient from 20% B to 35% B in the first 3 min, increased to 45% at 10 min, to 90% B at 17 min, followed by 3 min with 90% B, to 100% B at 22 min, followed by 4 min with 100% B; flow rate 1 ml/min and UV detection at 380 nm or 318 nm (Figure 1 and S2). For SR107PKS/NRPS, chromatographic conditions were as follows: solvents-A, water-0.1% formic acid (FA); solvent-B, acetonitrile-0.1% FA; solvent gradient from 20% B to 55% B in the first 5 min, increased to 70% at 8 min, followed by 8 min with 70% B, to 100% B at 17 min, followed by 3 min with 100% B, to 20% B at 21 min, followed by 4 min with 20% B; flow rate 1 ml/min and UV detection at 380 nm (Figure 2)

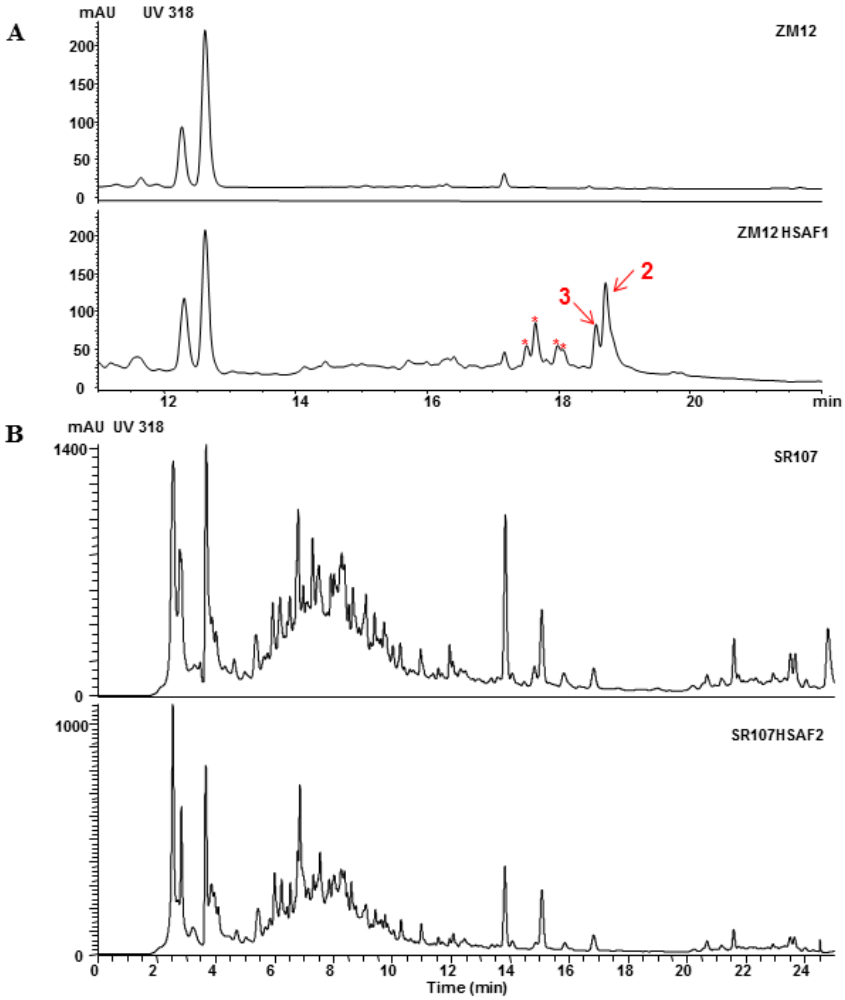


Figure S2. HPLC analysis of metabolites from *Streptomyces* strain ZM12HSAF1 (A) and strain SR107HSAF2 (B). Strain ZM12HSAF1 was generated by introducing the construct pSETHSAF3 (*ermE** promoter placed before the PKS-NRPS gene of the cluster) into *Streptomyces coelicolor* ZM12, in which all the ten native PKS and NRPS gene clusters have been deleted.[16] Strain SR107HSAF2 was generated by introducing the construct pSETHSAF4 (*ermE** promoter placed before the SD gene of the cluster) into *Streptomyces* sp. SR107, in which four PKS genes have been deleted.[12, 13] The metabolites were detected at 318 nm. The chemical structure of compounds **2** and **3** were determined and shown in Figure 1, and the asterisks indicate peaks that were absent in the control.

4. Isolation and structure determination of compounds 2 and 3

SR107HSAF culture (10 liters) was allowed to grow on the Fermentation Medium at 28 °C for 11 days. The whole solid cultures were diced and extracted three times with AcOEt/MeOH/AcOH (80 : 15 : 5,

v/v/v) at room temperature, and the crude extract solution was concentrated under reduced pressure, and the concentrated extract was sequentially solvent partitioned into petroleum ether-soluble extract and MeOH-soluble extract. The MeOH extract was loaded to a Sephadex LH-20 column for separation. The column was eluted with MeOH to obtain six fractions, **Fr.1 - 6**. **Fr.3** was subjected to medium-pressure liquid chromatography (MPLC; 80 g RP-18 silica gel; acetonitrile/H_2O 30%, 50%, 70%, 80%, and 100%, 200 ml each) to afford five subfractions, **Fr.3a – 3e**. **Fr.3c** (60 mg) was subjected to preparative HPLC (Agilent 1200, ZORBAX Eclipse XDB-C18, 9.4×250 mm, 5 μ), using an isocratic solvent of 55% acetonitrile-0.04% TFA, at flow rate of 4 ml/min, detected at UV 360 nm. This afforded compound **2** (10 mg) and compound **3** (4 mg). To determine the structures of the compounds, 1D- and 2D-NMR were performed on a Bruker DRX-600 spectrometer. MS was carried out on an LTQ-Orbitrap-XL mass spectrometer (Thermo Scientific).

Table S1. The ^{1}H- and ^{13}C-NMR spectroscopic data for compound **2**, at 600 MHz and 150 MHz, in DMSO-d_6, respectively; *J* in ppm

Position	^{1}H	^{13}C	Position	^{1}H	^{13}C
1		195.6s	17	1.36 (*m*)	47.0d
2	3.84 (*t*, *J* = 5.4)	61.0d	18	2.10 (*q*, *J* = 9.7)	55.4d
3	1.84 (*m*)	25.7t	19	1.83 (*m*)	55.5d
4	1.38 (*m*) 1.17 (*m*)	20.6t	20	3.78 (*t*, *J* = 9.7)	74.2d
5	3.28 (*m*) 2.45 (*t*, *J* = 9.1)	38.1t	21	6.11 (*m*)	147.1d
6	7.97 (*t*, *J* = 5.4)		22	6.03 (*dd*, *J* = 15.0, 11.1)	128.3d
7		164.8s	23	7.29 (*dd*, *J* = 15.4, 11.1)	142.3d
8	5.90 (*d*, *J* = 14.8)	125.6d*	24	6.83 (*d*, *J* = 15.4)	121.4d
9	6.95 (*t*, *J* = 13.3)	132.9d	25		175.2s
10	6.13 (*m*)	125.5d*	26		100.8s
11	5.80 (*t*, *J* = 11.2)	139.0d	27		171.9s
12	3.07 (*m*)	44.6d	28	8.73 (*s*)	
13	1.55 (*m*)	40.6t	29	1.09 (*d*, *J* = 6.3)	18.6q
14	2.63 (*m*)	41.8d	30	1.59 (*m*) 1.07 (*m*)	26.0t
15	2.05 (*m*) 0.88 (*m*)	40.1t	31	0.86 (*t*, *J* = 7.2)	12.5q
16	1.39 (*m*)	53.7d			

* exchangeable.

5. LC-MS-MS analysis of the metabolites in strain SR107PKS/NRPS

Strain SR107PKS/NRPS containing the construct pSETHSAF5 (PKS-NRPS alone under the control of *ermE** promoter) generated three new peaks as shown in Figure 2. To identify the metabolites, the peaks were individually collected from HPLC, and each of the peaks was further analyzed by UV-Vis spectrometry (Figure S3) and HPLC-HRMS using a Prominence Modular HPLC (Shimadzu Corporation, YMC-Pack Pro C18, 5 μm, 4.6 × 250 mm) coupled to a LTQ Velos Pro HRMS instrument (Thermo Scientific). For peak1, chromatographic condition was 60% acetonitrile/H_2O (containing 0.2% FA) for 10 min; for peaks 2 and 3, chromatographic condition was 85% acetonitrile/H_2O (containing 0.2% FA) for 10 min.

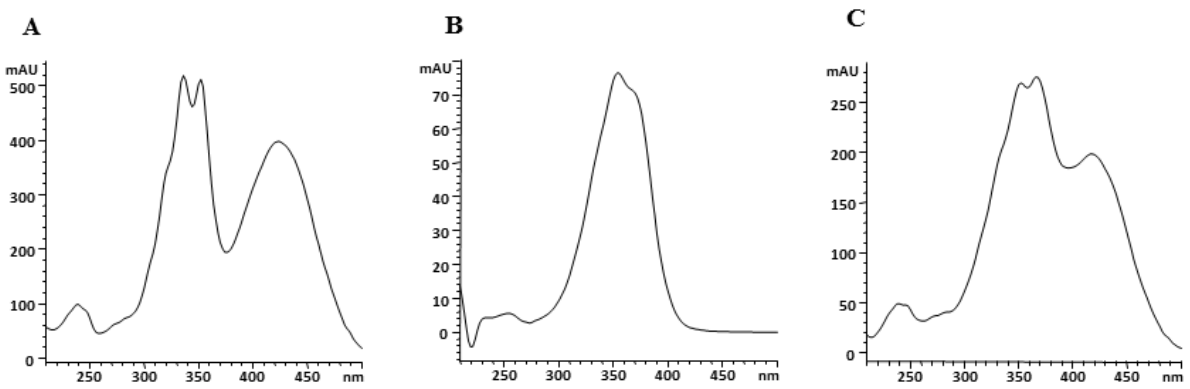


Figure S3. UV-Vis spectra of the three peaks observed in HPLC of the metabolites from *Streptomyces* strain SR107PKS/NRPS that was transformed with PKS-NRPS alone. A, peak at 11.7 min; B, peak at 16.3 min; C, peak at 20.8 min (polyene tetramate **6**).

6. HSAF polyketide synthase (PKS) expression in *E. coli*

The procedure for constructing the entire PKS gene (5526 bp) for expression in *E. coli* is illustrated in Figure S4. Two of the three pieces that make up the complete PKS gene were amplified by PCR using Cos4'-1 as the template. The following two pairs of primers were used for the PCR:

P1, KS-*Nco*I-Forward: 5'-GAG ACC ATG GAG GAC CGC ATC GCC A-3'

P2, KS-*Eco*RI-Reverse: 5'- CAT CAT CGT GCC GGT GGC GGT G-3'

P3, ACP-*Kpn*I-Forward: 5'- ACC GCG GCG ATG CGG TCG AAC-3'

P4, ACP-*Bam*HI-Reverse: 5'- GGG AAG GAT CCC AGC GCG TTC TGG T-3'

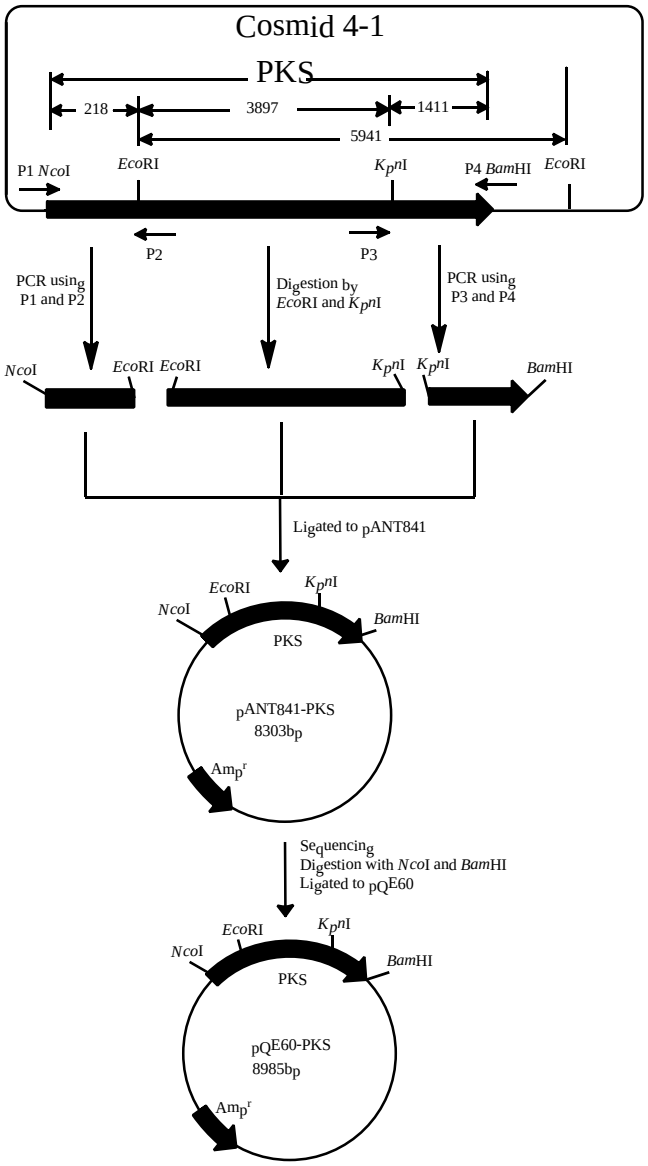


Figure S4. Strategy for constructing the PKS expression vector for the production of the enzyme in *E. coli*.

The KS-fragment (218 bp) amplified by primers P1/P2 was digested with *Nco*I and *Eco*RI; the ACP-fragment (1411 bp) amplified by primers P3/P4 was digested with *Kpn*I and *Bam*HI (Figure S4). Both fragments were cloned into pANT841 at the respective restriction sites. The middle fragment of the PKS was obtained by digesting Cos4'-1 with *Eco*RI, followed by digestion of the resulted fragment (5941 bp) with *Kpn*I. The larger *Eco*RI/*Kpn*I fragment (3897 bp) was cloned into pANT841 at the same sites. The three pieces were ligated together to generate pANT841-PKS (8303 bp). The construct was sequenced to

verify the identity of the PKS gene. Finally, the PKS gene was released from pANT841-PKS as an *Nco*I/*Bam*HI fragment and cloned into the expression vector pQE60 (Figure S4 and S5).

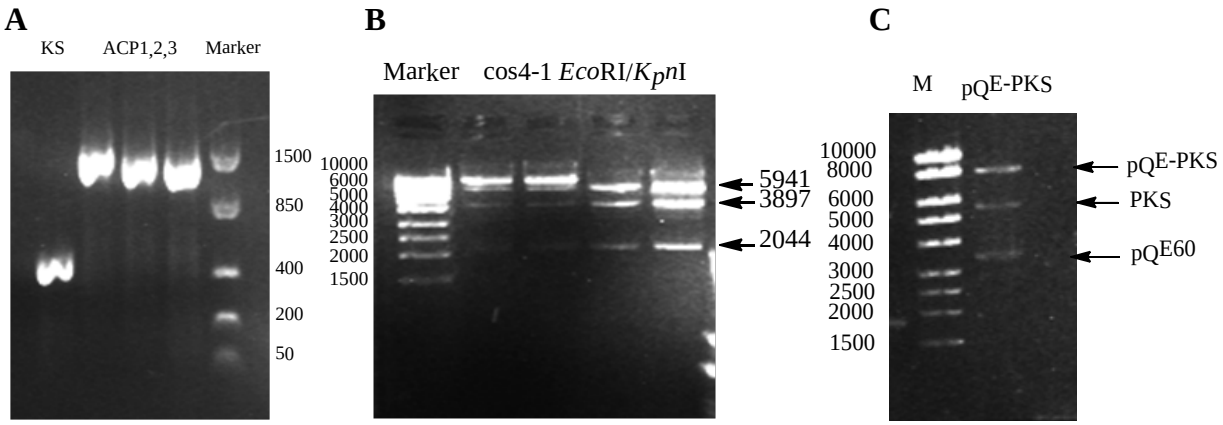


Figure S5. Gel electrophoresis to check the constructs for PKS expression in *E. coli*. A, PCR amplification of the KS-fragment and the ACP-fragment; B, the *Eco*RI/*Kpn*I fragment (3897 bp) of the PKS obtained by enzyme digestion of the cosmid clone, Cos4'-1; C, *Nco*I/*Bam*HI digestion of pQE60-PKS to confirm the final construct for expression.

To express the PKS, pQE60-PKS was introduced into *E. coli* SG13009. Individual single colonies were inoculated in 3 ml of liquid LB medium containing ampicillin (50 µg/ml), and the culture was incubated in a shaker (250 rpm) at 37 °C overnight. The overnight culture was added to 50 ml fresh LB medium and continued in a shaker (250 rpm) at 37 °C until the cell density (OD_{600nm}) reached 0.6. To induce the expression of the PKS, IPTG (final concentration 0.3 mM) was added to the culture, and the cells were allowed to grow (250 rpm) at the room temperature for another 12 hours. To extract proteins, the cells were harvested and resuspended in 2 ml of PBS buffer (500 mM NaCl, 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.8). The cell suspension was sonicated on ice, and the soluble fraction of protein extracts was loaded onto a Ni-NTA column, which was previously calibrated with the PBS buffer containing 10 mM imidazole. The column was washed three times with the buffer containing 20 mM imidazole, and the C-His_6-tagged PKS protein (199.8 kDa) was purified by using an imidazole step-gradient as described by the manufacturer's protocol. The purity of the protein was analyzed by SDS-PAGE (Figure S6), and the fractions containing the purified PKS were pooled, concentrated, and dialyzed against a buffer containing

50 mM NaCl, 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.8, and 15% glycerol. Finally, the protein preparation was frozen in liquid nitrogen and stored at -80 °C until use.

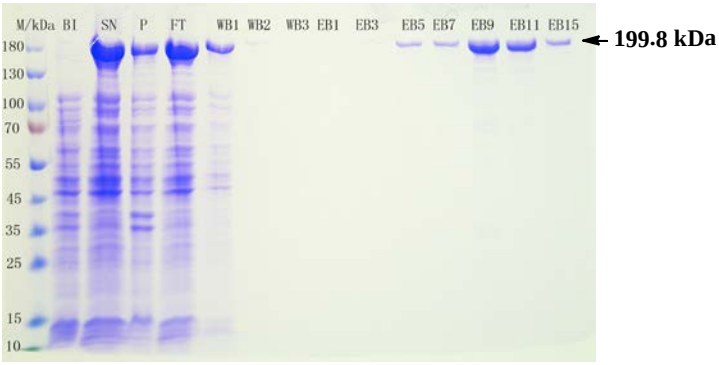


Figure S6. SDS-PAGE of the HSAF PKS expressed in *E. coli*. M, markers; BI, total proteins before IPTG induction; SN, total soluble proteins after sonication; P, total precipitated proteins after sonication; FT, flow-through fraction of the Ni-NTA column; WB1-3, wash-off fractions of the Ni-NTA column; EB1-15, imidazole-eluted fractions of the Ni-NTA column. The purified PKS is indicated with an arrow.

7. *In vitro* assay for the activity of the PKS module

The first step was to convert the purified PKS to its holo form, through 4'-phospho-pantetheinylation at the active site serine residue of its ACP domain. This was carried by Svp,[17] a promiscuous 4'-phosphopantetheinyl transferase (PPTase). The reaction contained the PKS protein (25 μM), CoA (2 mM) and Svp (10 μM), in a 45 μl solution of Tris-HCl (100 mM, pH 8.0), $MgCl_2$ (10 mM), and TCEP (0.5 mM). After incubation at 37 °C for 2 h, the reaction was stopped by boiling for 10 min. Due to the huge size (199.8 kDa) of the PKS protein, a trypsin digestion procedure was employed prior to the ESI-Q-TOF analysis of the mass of the PKS. The protein was loaded to a SDS-PAGE (7.5%), and the PKS band was excised from the gel and subjected to trypsin digestion.[20] Briefly, gel pieces were digested by trypsin (no.V5111, Promega, Madison, WI) and digested peptides were extracted in 5% formic acid / 50% acetonitrile and separated using C18 reversed phase LC column (75 micron × 15 cm, BEH 130 , 1.7 micron Waters, Milford, MA). A Q-TOF Ultima tandem mass spectrometer couple with a Nanoaquity HPLC system (Waters) with electrospray ionization was used to analyze the eluting peptides. The system was user controlled employing MassLynx software (v 4.1, Waters) in data-dependent acquisition mode

with the following parameters: 0.9-sec survey scan (380–1900 Da) followed by up to three 1.4-sec MS/MS acquisitions (60–1900 Da). The instrument was operated at a mass resolution of 8000. The instrument was calibrated using the fragment ion masses of doubly protonated Glu-fibrinopeptide. The peak lists of MS/MS data were generated using Distiller (Matrix Science, London, UK) using charge state recognition and de-isotoping with the other default parameters for Q-TOF data. Data base searches of the acquired MS/MS spectra were performed using Mascot (Matrix Science, v2.2.0, London, UK). The NCBI database (20100701) was used in the searches. Search parameters used were: no restrictions on protein molecular weight or pI, enzymatic specificity was set to trypsin with up to 3 missed cleavage sites, carbamidomethylation of C was selected as a fixed modification, and methionine oxidaztion was selected as a variable modification. Mass accuracy settings were 0.15 daltons for peptide mass and 0.12 daltons for fragment ion masses. The apo-PKS, holo-PKS, and the various forms of acylated PKS were analyzed by the MS method, which had been successfully used in other studies.[18, 19]

The trypsin digestion of this apo-PKS is expected to release a 26-residue fragment, VKPEQ IDADA SLNAL GLD**S**L LAMELR, which contains the active site serine residue (underlined) of the ACP domain of the PKS (Table S2, shown the mass of a number of selected ions, both the predicted and the observed). Q-TOF MS data showed that the tryptic fragment from the apo-PKS isolated from *E. coli* had *m*/*z* 928.1374 for $[M+3H]^{3+}$ (calculated 928.1604) (Figure 3). This fragment was subject to tandem MS-MS analysis (Figure S7). The result showed that this tryptic fragment had the expected amino acid sequence, confirming that the 199.8 kDa protein purified from *E. coli* was HSAF PKS.

Table S2. The predicted mass and observed mass of the tryptic PKS fragment containing the ACP domain.

Substrate	Predicted product	Predicted Mass (Da)		Observed Mass (Da)		Error (Da)
		M+3H	M+4H	M+3H	M+4H	M+1H
-	apo-PKS	928.1604	696.3722	928.1347	NA	0.0613
CoA	holo-PKS	1041.5221	781.3936	1041.5487	781.4270	0.0798
Mal-CoA	PKS-polyene	1113.5533	835.4170	1113.4847	835.4001	0.2058

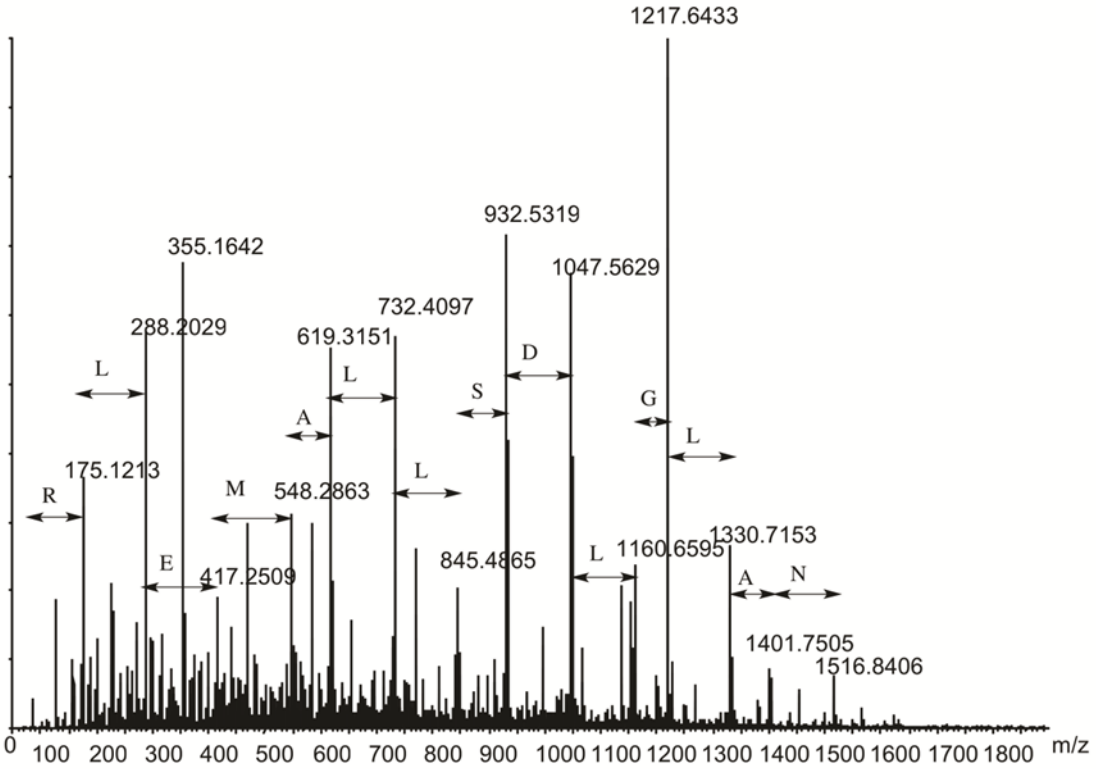
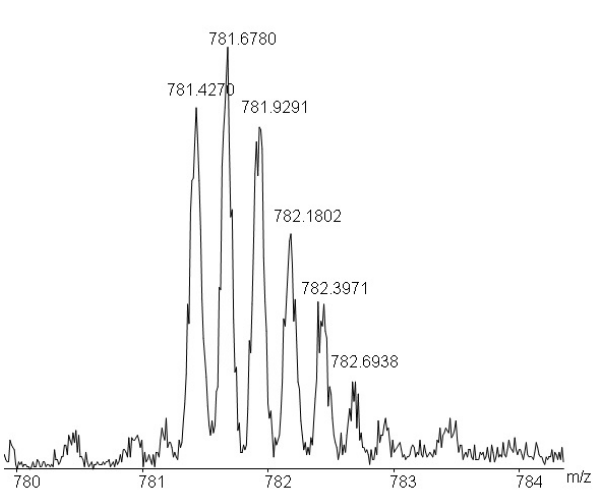
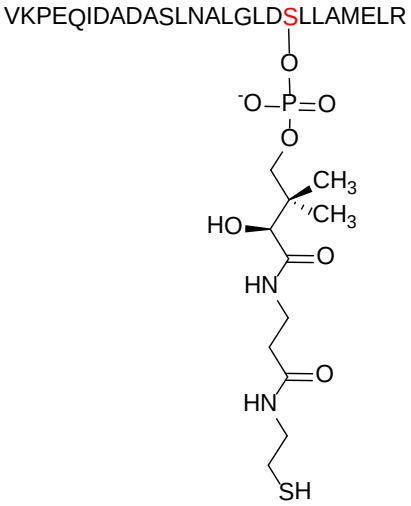
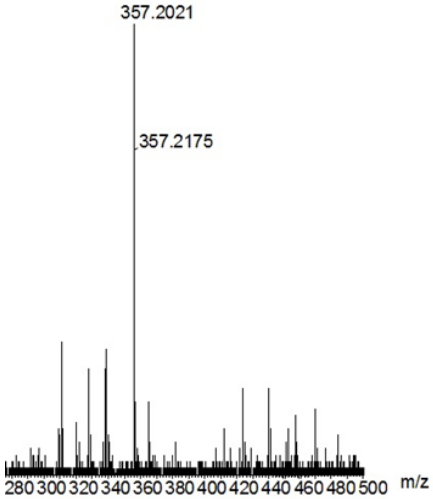


Figure S7. Tandem MS-MS data for the tryptic PKS fragment that harbors the serine active site of the ACP domain, VKPEQ IDADA SLNAL GLD**S**L LAMELR.

Upon 4'-phosphopantetheinylation, the tryptic fragment from the holo-PKS is expected to have a 340.1 mass increase. Q-TOF MS data showed *m*/*z* 1041.5487 for $[M+3H]^{3+}$ (calculated 1041.5221) (Figure 3) and *m*/*z* 781.4270 for $[M+4H]^{4+}$ (calculated 781.3936) (Figure S8, top panel). In addition, the 4'-phosphopantetheinylation of the PKS was examined using PPT ejection assay.[21] The expected compound was detected from the holo-PKS (Figure S8, bottom panel). The results demonstrated that the apo-form of the PKS expressed in *E. coli* had been converted into its holo-form *in vitro*.







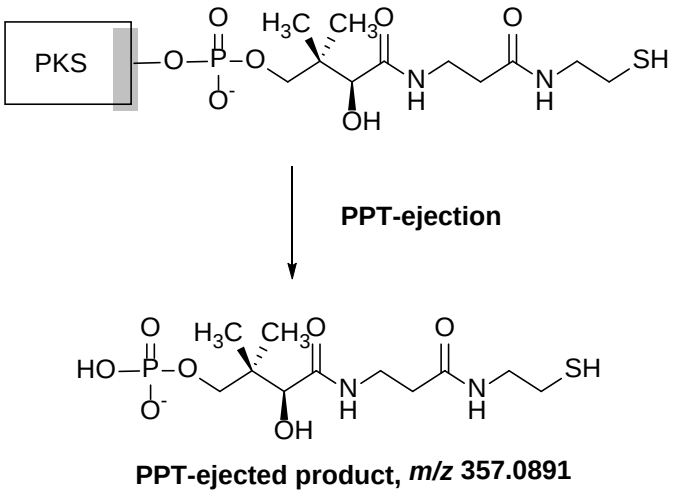


Figure S8. Top panel, Q-TOF MS detection of the 26-residue tryptic fragment from the 4'-phosphopantetheinylated PKS, with *m*/*z* 781.4270 for $[M+4H]^{4+}$ (calculated 781.3936); bottom panel, MS detection of the compound resulted from a PPT ejection, with *m*/*z* 357.2021 for $[M]^-$ (calculated 357.0891).

To test the activity of the PKS, the reaction mixture (45 µl) containing the holo-PKS was added with malonyl-CoA (10 mM), NADPH (10 mM) and acetyl-CoA (2 mM), and the final volume was brought to 50 µl. The reactions were incubated at 37 °C for 20 h. A reaction containing boiled PKS protein served as the control. The reactions were stopped by boiling for 10 min, and a 10 µl aliquot of the reaction mixtures was loaded on SDS-PAGE (7.5%), and the PKS band was cut out for trypsin digestion. The 26-residue tryptic fragment was then analyzed by MS. Since the acyl chain synthesized by the PKS could have various chain lengths and oxidation levels, the samples were subject thorough search for various potential acylated products. However, hexakitide polyene appeared to be the only one that was detectable under the experimental conditions (Figure 3 and S9).

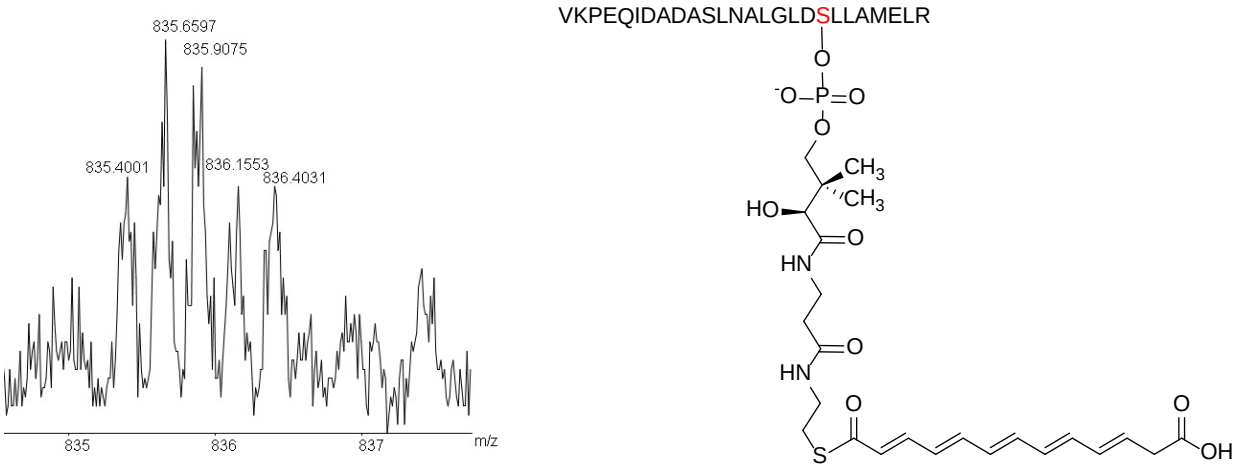


Figure S9. Q-TOF MS detection of the 26-residue tryptic fragment from acylated PKS, with *m*/*z* 835.4001 for $[M+4H]^{4+}$ (calculated 835.4170), showing the hexaketide polyene synthesized by the single-module PKS (also see Figure 3).

8. Expression of PKS with a point-mutated KS domain in *E. coli* and activity assay of the synthase

The *in vitro* data obtained from the purified PKS suggest that the synthase may use malonyl-CoA as both the starter and the extender in the polyketide chain synthesis because a carboxylate appeared present in the intermediate (Figure 3). To probe the starter specificity, we generated another PKS expression construct, in which the active site cysteine in the KS domain of this PKS was mutated to alanine (C176A,

with the protein accession number ABL86391). Primer extension PCR was performed to generate a 2151-bp fragment of PKS gene with C176A mutation. The following two pairs of primers were used to amplify a 359-bp fragment and a 1832-bp fragment (Figure S10):

KS(C/A) up-fw: 5'-CGC GGC GGC TAC ATC GAC GGC TTC GAC G-3'

KS(C/A) up-rv: 5'-GGC GAC CAG GGA GGA ACT AGC GGC GGT GTC GAT CGA CAG GCT-3'

KS(C/A) down-fw: 5'-AGC CTG TCG ATC GAC ACC GCC GCT AGT TCC TCC CTG GTC GCC-3'

KS(C/A) down-rv: 5'-GTA GAG CTC GTG GCC GGC GGC GAT GCC-3'

After purification of the PCR fragments, an overlapping PCR was carried out using these 2 fragments as template. The resulted 2151-bp fragment was digested with *Eco*RI and *Sac*I, and cloned into previously constructed pANT841-PKS (Figure S4) at the same sites. The mutated PKS gene (mPKS) was released from pANT841 as a *Bam*HI/*Nco*I fragment and cloned into expression vector pQE60 (Figure S10). The subsequent expression and purification are the same as for the wild type PKS.

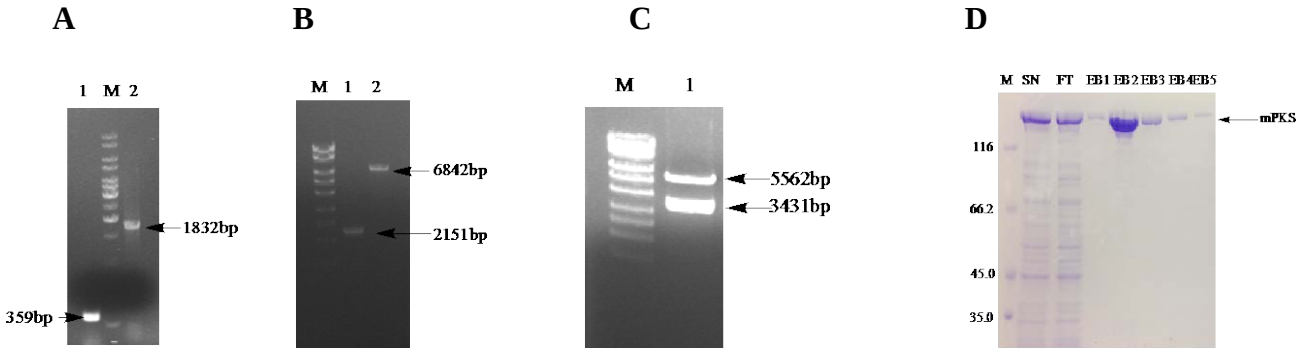


Figure S10. Construction of the expression vector for the KS active-site mutated PKS (mPKS) and expression of the enzyme in *E. coli*. A, PCR amplification of the KS upstream fragment (359-bp) and downstream fragment (1832-bp). These two fragments were used as templates in overlapping PCR. The marker is the 1 kb ladder from Fermentas. B, Overlapping PCR. Lane 1 is the overlapping PCR fragment (2151-bp), which harbors the C/A mutation site; lane 2 is the pANT841-PKS (original) vector digested by *Eco*RI and *Sac*I. Marker is the high range ladder No.393 from Fermentas. C, analysis of the expression construct pQE60-mPKS. Lane 1 is *Nco*I and *Bam*HI double digested pQE60-mPKS, resulted in 2 bands of the expected lengths (5562-bp and 3431-bp). D, SDS-PAGE of the mPKS expressed in *E. coli*. M, markers; SN, total soluble proteins after sonication; FT, flow-through fraction of the Ni-NTA column; EB1-5, imidazole-eluted fractions of the Ni-NTA column. The purified PKS is indicated with an arrow.

To assay the activity, mPKS (25 µM) was first converted into its holo form, by incubating with CoA (2 mM) and Svp (10 µM) in a 50 µl reaction containing Tris-HCl (100 mM, pH 8.0), $MgCl_2$ (10 mM), and TCEP (0.5 mM). After incubation at 37 °C for 2 h, the reaction mixture was added with 10 mM malonyl-CoA, 2 mM acetyl-CoA, and 10 mM NADPH. A reaction containing the wild type PKS served as the control. After continual incubation at 37°C for 20 h, the reactions were stopped by boiling. A 10 µl aliquot of the reaction mixtures was loaded on SDS-PAGE for analysis. The mPKS band was cut from the gel and subject to trypsin digestion and Q-TOF MS analysis as described above.

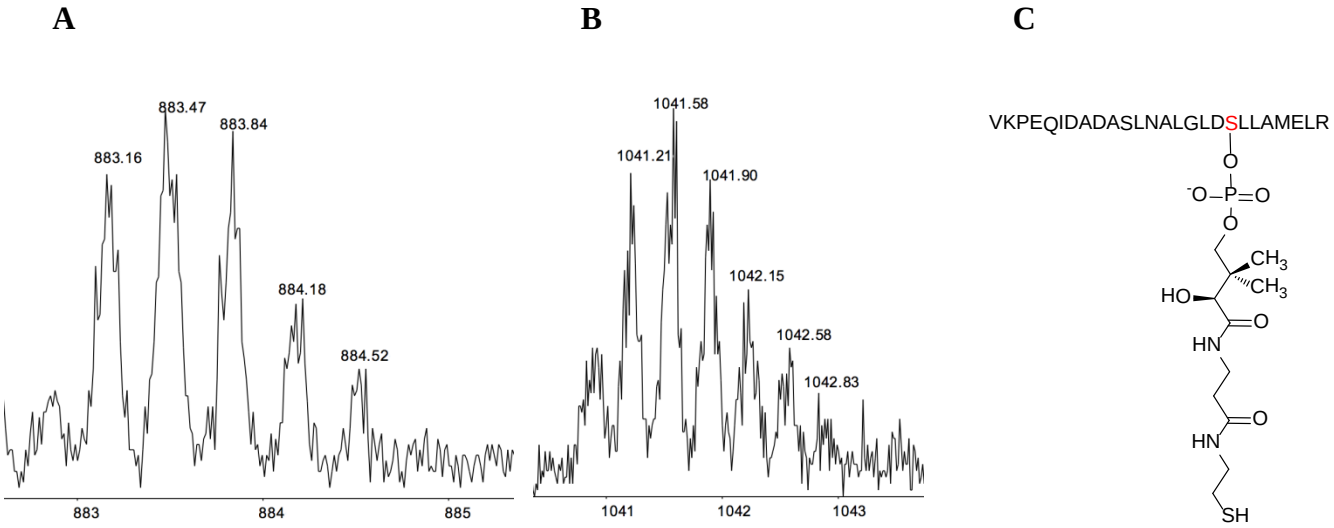


Figure S11. Q-TOF MS data for the tryptic fragment of mPKS. A, MS data confirming the C176A mutation in the tryptic fragment from the KS domain, GPSLSIDTA**A**SSSLVAVHLACHSLRR (underlined residue to indicate the C176A mutation), with *m/z* 883.16 for $[M+3H]^{3+}$ (calculated 883.45). B and C, MS data to show that the KS point-mutated PKS is active in 4'-phosphopantetheinylation by Svp, using CoA as substrate; the 26-residue tryptic fragment from the 4'-phosphopantetheinylated ACP domain gave *m/z* 1041.21 for $[M+3H]^{3+}$ (calculated mass: 1041.52).

Q-TOF MS confirmed the C176A mutation in the KS domain of mPKS and the 4'-phosphopantetheinylation of the ACP domain (Figure 11S). When holo-mPKS was incubated with both

acetyl-CoA and malonyl-CoA in the presence of NADPH, we detected the tryptic fragment with *m*/*z* 1069.89 for $[M+3H]^{3+}$ (calculated 1070.19 for malonyl-*S*-ACP, 1055.52 for acetyl-*S*-ACP) and *m*/*z* 802.69 for $[M+4H]^{4+}$ (calculated 802.87 for malonyl-*S*-ACP, 791.86 for acetyl-*S*-ACP) (Figure S12). The results showed that the PKS prefers malonyl-CoA over acetyl-CoA as the starter.

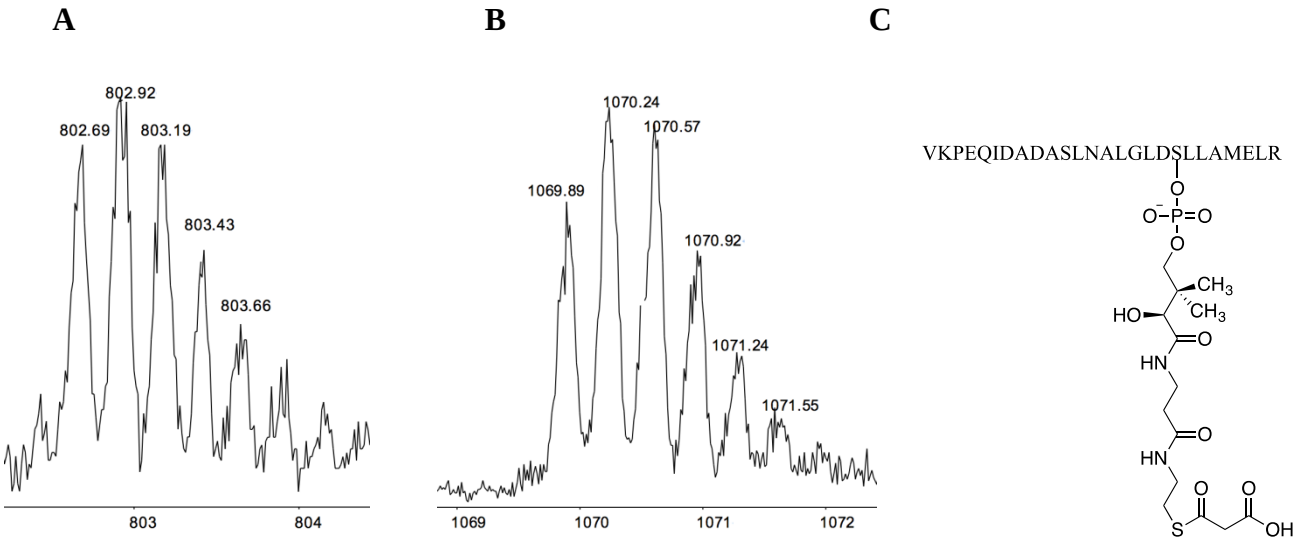


Figure S12. Q-TOF MS data for the tryptic fragment of holo mPKS, after incubating with malonyl-CoA, acetyl-CoA and NADPH. A, *m/z* 1069.89 for $[M+3H]^{3+}$ (calculated 1070.19 for tryptic malonyl-*S*-ACP, 1055.52 for tryptic acetyl-*S*-ACP); B, *m/z* 802.69 for $[M+4H]^{4+}$ (calculated 802.87 for tryptic malonyl-*S*-ACP, 791.86 for tryptic acetyl-*S*-ACP); C, the 26-residue tryptic fragment from malonyl-*S*-ACP.

9. *In vitro* reconstitution of the PKS module and the NRPS module

The PKS module was activated to its holo-form and incubated with proper substrates as described in the reactions above. The expression and purification of the NRPS module had been described previously.[3] The purified NRPS module was also converted to its holo form by incubating with CoA (0.83 mM) with Svp (5.6 μM) in a 60 μl reaction containing Tris-HCl (100 mM, pH 8.0), $MgCl_2$ (10 mM), and TCEP (0.5 mM). The reactions were incubated at 37 °C for 2 h. The NRPS reaction mixture and the PKS reaction mixture were then combined into a tube, which had a 40 μl solution containing L-Orn (1.5 mM), ATP (3 mM), Tris-HCl (100 mM, pH 8.0), $MgCl_2$ (10 mM), NaCl (50 mM), EDTA (0.1 mM), and TCEP (0.5 mM). For control, the reaction contained all the components of the reconstitution reaction except that

NRPS was replaced with the total proteins of the *E. coli* host strain that was transformed only with the blank expression vector pQE60. After continual incubation for 5 h at 37 °C, the reactions were stopped by adding 150 µl of 0.2 mM TCA in methanol and were frozen at -20 °C for 30 min. The mixtures were centrifuged at 13,200 rpm for 20 min in a desktop Eppendorf centrifuge, and the supernatants were transferred to new tubes. The solutions were dried in a Speed-Vac, and the residues in the tubes were re-dissolved in 150 µl methanol. The methanol extracts were centrifuged, and supernatants were transferred to new tubes and dried again. Finally, the residues in the tubes were re-dissolved in 300 µl 50% methanol and analyzed by LC-MS. The sample (50 µl) was loaded to a Waters BEH column (C18 1.7 µm, 1.0 × 100 mm), and eluted with the following solvent program: Solvent A was H_2O containing 0.1% formic acid; Solvent B was acetonitrile containing 0.1% formic acid. The flow rate was 50 µl/min with a binary gradient system (0-15 min, 5%-45% B gradient; 15-35 min, 45% B to 90% B gradient; 35-46 min, 90% B to 100% B gradient; 46-55 min, 100% B; 55-56 min, 100% B to 5% B gradient; 56-66 min, 5% B). The mass spectrometer was Mass Spec Waters Snapt G2-S Q-TOF, operated in positive ion high resolution mode with lock mass, mass range 100-1200 amu, 1sec scans.

The LC-MS was set to scan for various masses of potential products that could be synthesized by the reconstituted PKS-NRPS and then released into the medium by the thioesterase (TE) domain in the NRPS module.[3, 4, 22, 23] One eminent peak at 15.48 min was detected in the reconstituted reaction, but no in the control reaction, when the scan was at *m*/*z* 475.2597 (Figure 4). This peak was subject to HR-ESI-MS analysis and gave *m*/*z* 475.2572 (Figure S13). This mass is coincident with the expected $[M+H]^+$ for the polyene tetramate **6**.

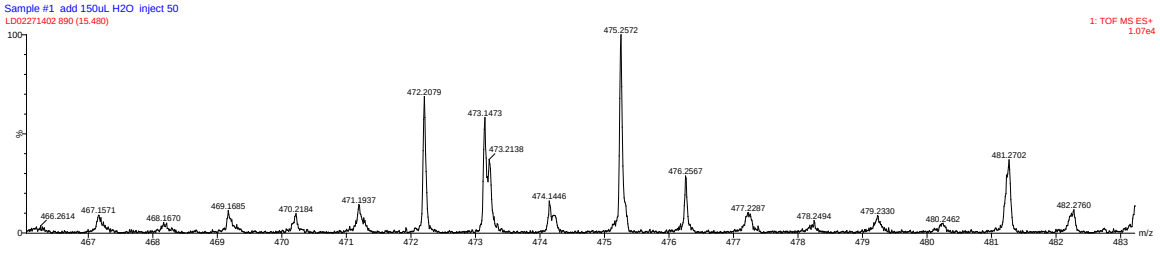


Figure S13. Q-TOF MS analysis of the polyene tetramate (**6**) produced in the *in vitro* reconstituted PKS and NRPS reaction. The *m/z* 475.2572 is consistent with $[M+H]^+$ of compound **6** (calculated 475.2597).

Below are NMR data for Compounds **2** and **3** whose structures are shown in Figure 1.

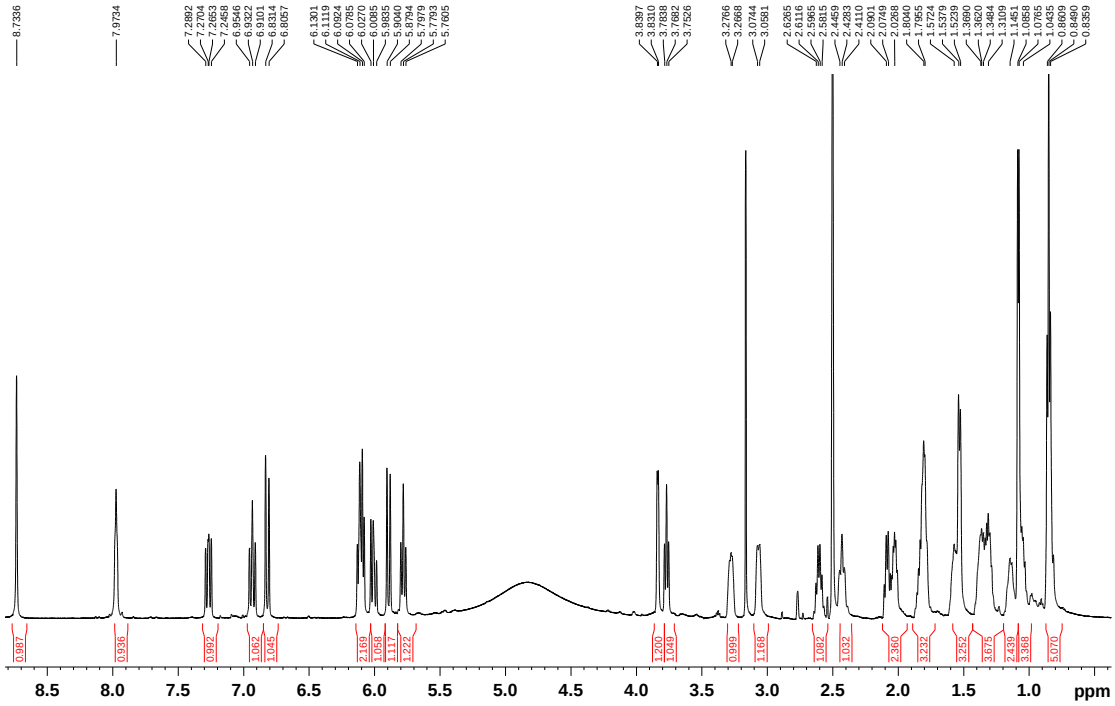


Figure S14. ^{1}H NMR spectrum of **2** in DMSO-d_6.

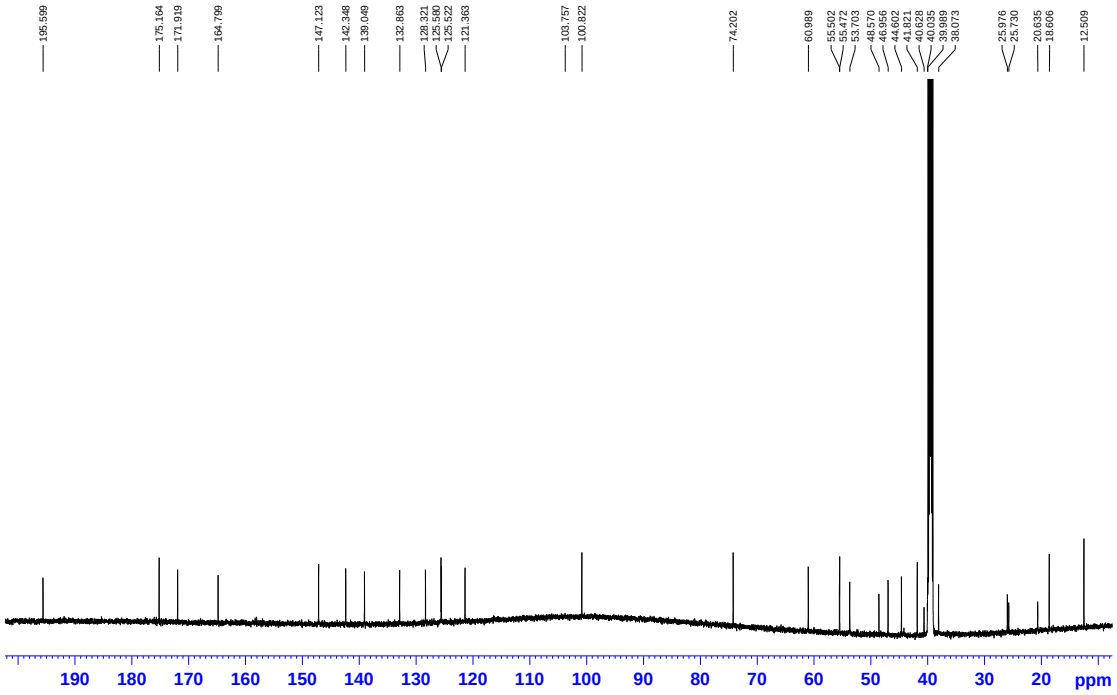


Figure S15. ^{13}C NMR spectrum of **2** in DMSO-d_6.

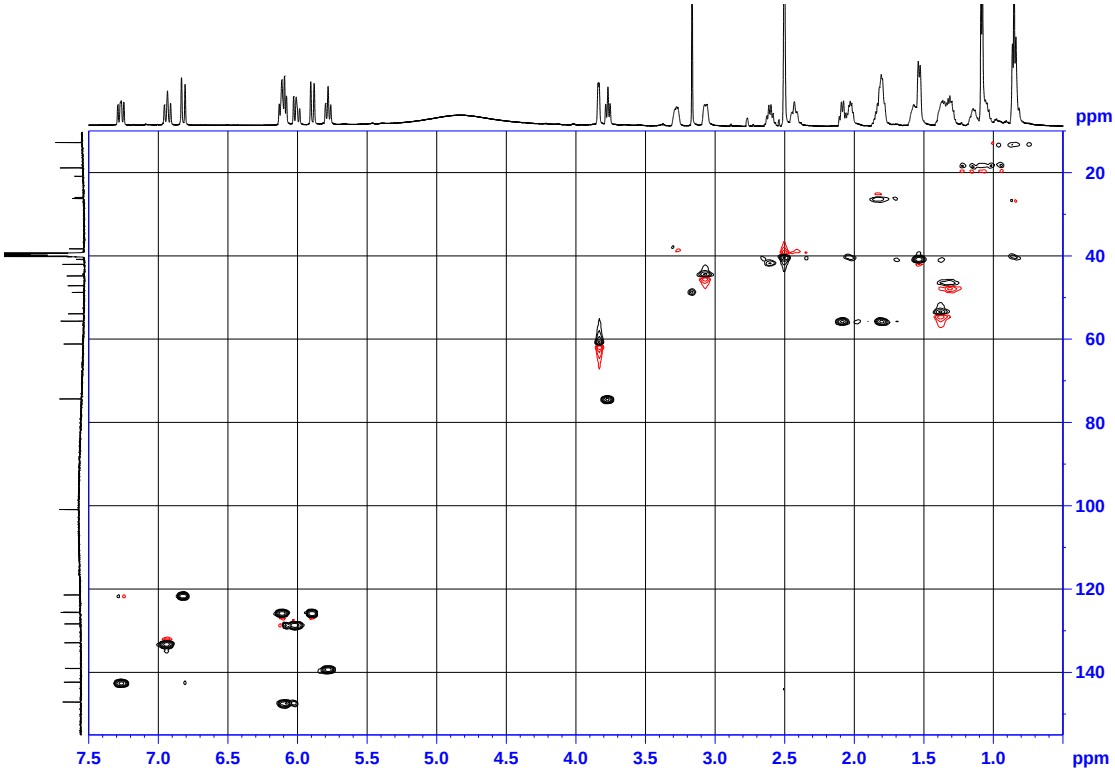


Figure S16. The HSQC spectrum of **2** in DMSO-d_6.

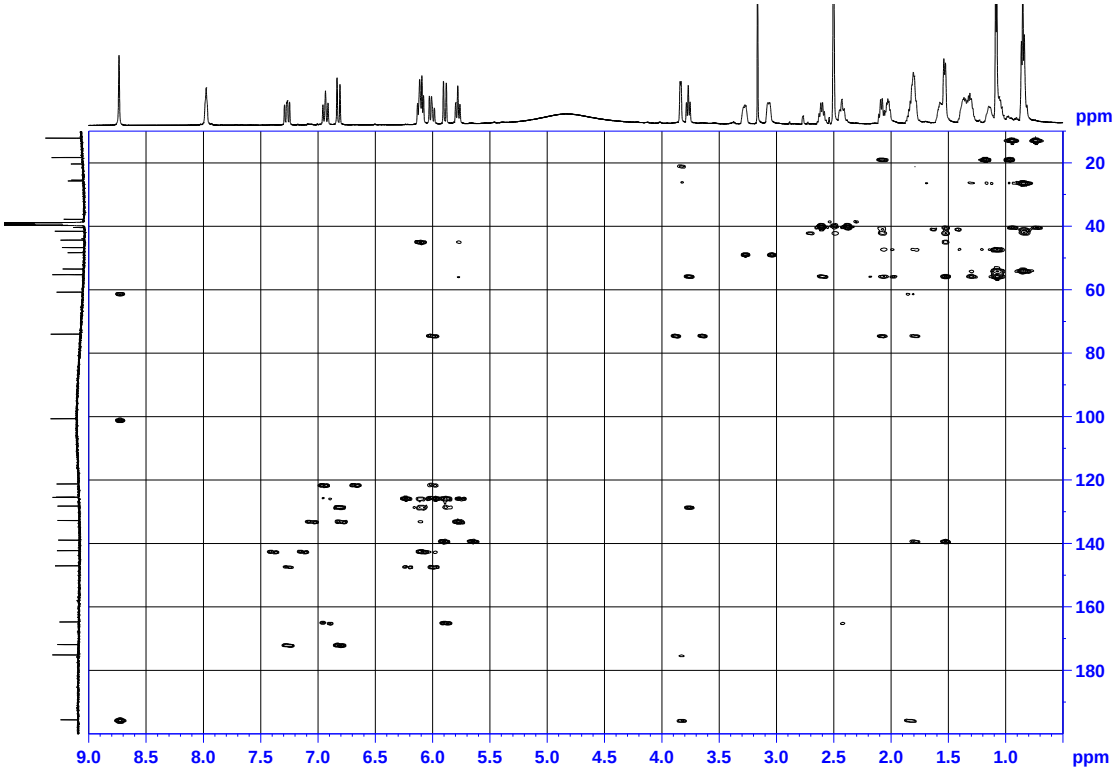


Figure S17. The HMBC spectrum of **2** in DMSO-d_6.

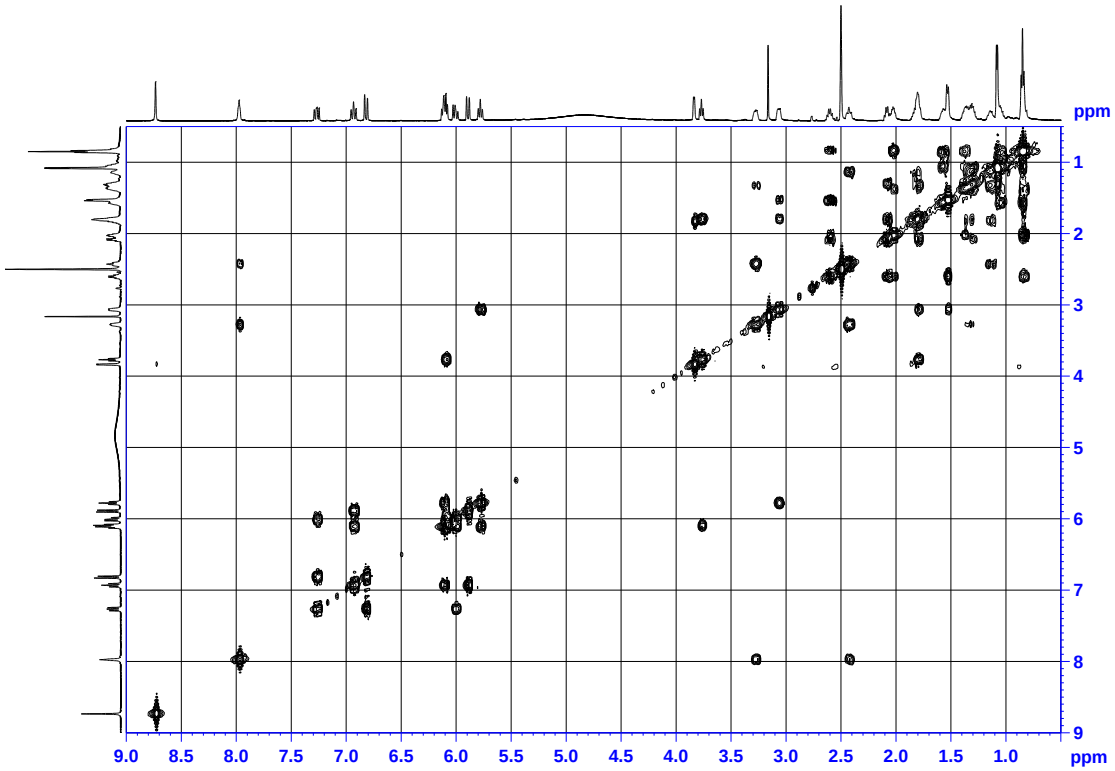


Figure S18. The ^{1}H - ^{1}H COSY spectrum of **2** in DMSO-d_6.

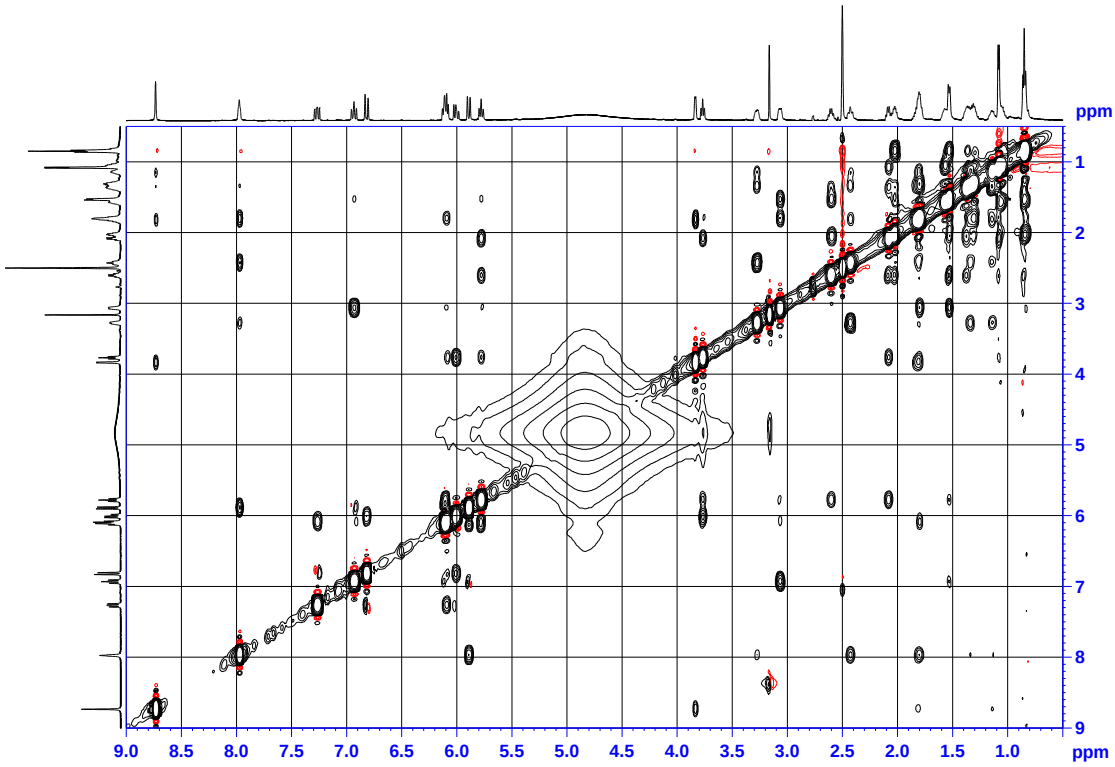


Figure S19. The NOESY spectrum of **2** in DMSO-d_6.

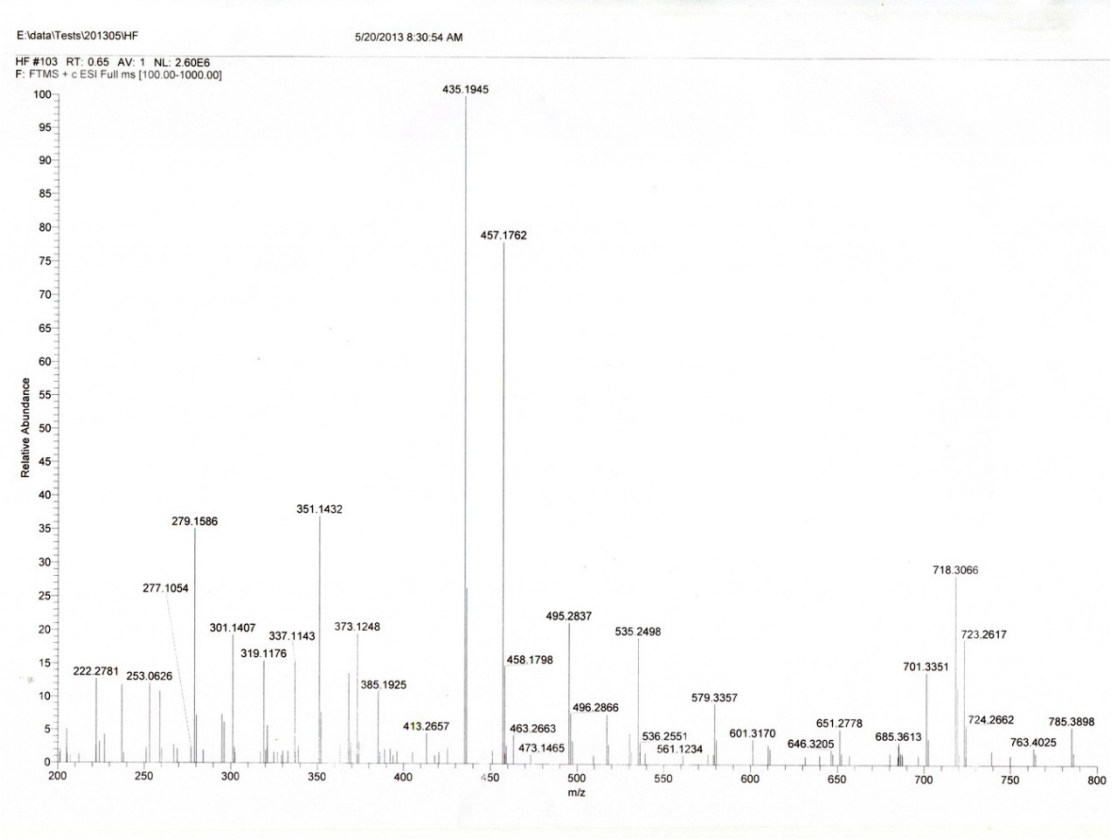


Figure S20. HR-ESI mass spectrum of **2**.

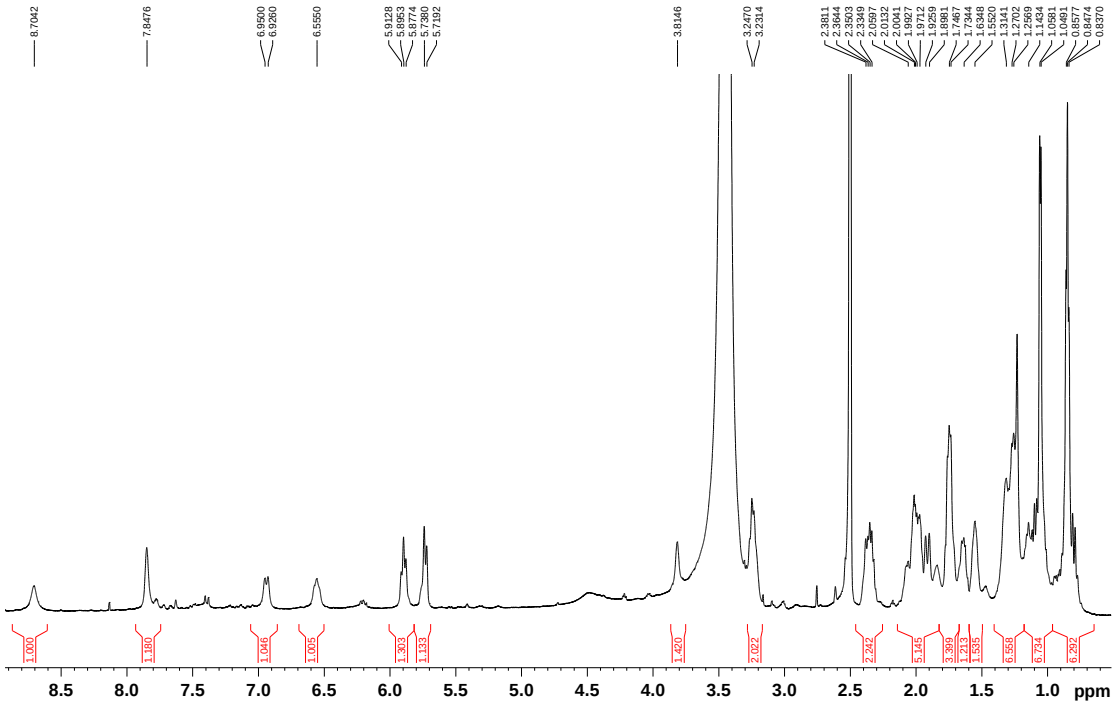


Figure S21. ^{1}H NMR spectrum of **3** in DMSO-d_6.

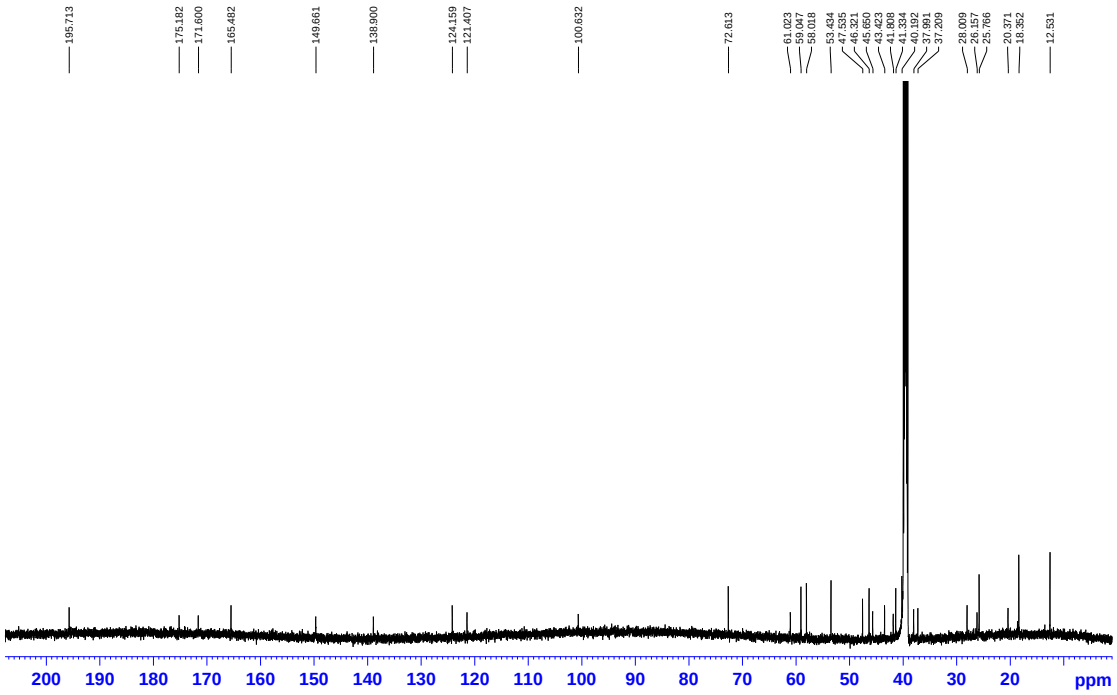


Figure S22. ^{13}C NMR spectrum of **3** in DMSO-d_6.

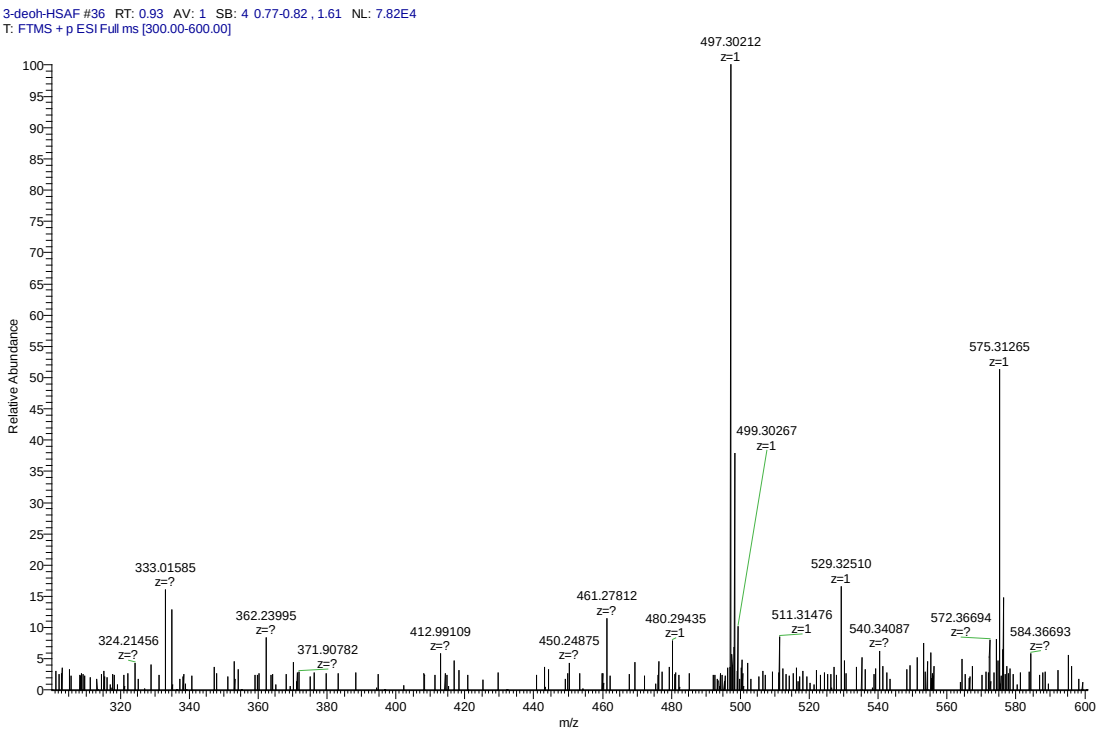


Figure S23. HR-ESI mass spectrum of **3**.

References

[1] S. Li, L. Du, G. Yuen, S. D. Harris, *Mol Biol Cell* **2006**, *17*, 1218-1227.

[2] F. Yu, K. Zaleta-Rivera, X. Zhu, J. Huffman, J. C. Millet, S. D. Harris, G. Yuen, X. C. Li, L. Du, *Antimicrob Agents Chemother* **2007**, *51*, 64-72.

[3] L. Lou, G. Qian, Y. Xie, J. Hang, H. Chen, K. Zaleta-Rivera, Y. Li, Y. Shen, P. H. Dussault, F. Liu, L. Du, *J Am Chem Soc* **2011**, *133*, 643-645.

[4] L. Lou, H. Chen, R. L. Cerny, Y. Li, Y. Shen, L. Du, *Biochemistry* **2012**, *51*, 4-6.

[5] J. A. Blodgett, D. C. Oh, S. Cao, C. R. Currie, R. Kolter, J. Clardy, *Proc Natl Acad Sci U S A* **2010**, *107*, 11692-11697.

[6] S. Cao, J. A. Blodgett, J. Clardy, *Organic Letters* **2010**, *12*, 4652-4654.

[7] J. Antosch, F. Schaefers, T. A. Gulder, *Angew Chem Int Ed Engl* **2014**, *53*, 3011-3014.

[8] G. Zhang, W. Zhang, Q. Zhang, T. Shi, L. Ma, Y. Zhu, S. Li, H. Zhang, Y. Zhao, R. Shi, C. Zhang, *Angew Chem Int Ed Engl* **2014**, *53*, in press.

[9] M. J. Bibb, G. R. Janssen, J. M. Ward, *Gene* **1985**, *38*, 215-216.

[10] C. J. Wilkinson, Z. A. Hughes-Thomas, C. J. Martin, I. Bohm, T. Mironenko, M. Deacon, M. Wheatcroft, G. Wirtz, J. Staunton, P. F. Leadlay, *J Mol Microbiol Biotechnol* **2002**, *4*, 417-426.

[11] K. A. Datsenko, B. L. Wanner, *Proc Natl Acad Sci U S A* **2000**, *97*, 6640-6645.

[12] Y. Jiang, H. Wang, C. Lu, Y. Ding, Y. Li, Y. Shen, *ChemBiochem* **2013**, *14*, 1468-1475.

[13] G. S. Zhao, S. R. Li, Y. Y. Wang, H. L. Hao, Y. M. Shen, C. H. Lu, *Drug Discov Ther* **2013**, *7*, 185-188.

[14] H. Shigemori, M. A. Bae, K. Yazawa, T. Sasaki, J. Kobayashi, *Journal of Organic Chemistry* **1992**, *57*, 4317-4320.

[15] Y. Li, J. Huffman, Y. Li, L. Du, Y. Shen, *MedChemComm* **2012**, *9*, 982-986.

[16] M. Zhou, X. Jing, P. Xie, W. Chen, T. Wang, H. Xia, Z. Qin, *FEMS Microbiology Letters* **2012**, *333*, 169-179.

[17] C. Sanchez, L. Du, D. J. Edwards, M. D. Toney, B. Shen, *Chemistry & Biology* **2001**, *8*, 725-738.

[18] T. S. Hitchman, J. Crosby, K. J. Byrom, R. J. Cox, T. J. Simpson, *Chem Biol* **1998**, *5*, 35-47.

[19] J. Crosby, K. J. Byrom, T. S. Hitchman, R. J. Cox, M. P. Crump, I. S. Findlow, M. J. Bibb, T. J. Simpson, *FEBS Lett* **1998**, *433*, 132-138.

[20] J. P. Kayser, J. L. Vallet, R. L. Cerny, *J Biomol Tech* **2004**, *15*, 285-295.

[21] P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh, N. L. Kelleher, *Biochem* **2006**, *45*, 12756-12766.

[22] L. Du, L. Lou, *Nat Prod Rep* **2010**, *27*, 255-278.

[23] Y. Xie, S. Wright, Y. Shen, L. Du, *Nat Prod Rep* **2012**, *19*, 1277-1287.

[24] A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson, R. J. Cox, *Journal of the American Chemical Society* **2011**, *133*, 10990-10998.

[25] J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas, C. R. Hutchinson, *Science* **1999**, *284*, 1368-1372.

[26] S. M. Ma, J. W. Li, J. W. Choi, H. Zhou, K. K. Lee, V. A. Moorthie, X. Xie, J. T. Kealey, N. A. Da Silva, J. C. Vederas, Y. Tang, *Science* **2009**, *326*, 589-592.

[27] Y. Abe, T. Suzuki, C. Ono, K. Iwamoto, M. Hosobuchi, H. Yoshikawa, *Mol Genet Genomics* **2002**, *267*, 636-646.

[28] Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb, H. Zhao, *Nat Commun* **2014**, *4*, in press.