Supporting Information

Zhang *et al.* 10.1073/pnas.0809084105

SI Text

Cloning of a PKS4 KS_MAT Didomain. The KS_MAT didomain was cloned as a standalone protein with primer pair pks4-KS-F-NdeI and pks4-AT-R-NotI-S-EcoRI, flanking the didomain with an NdeI restriction site preceding the domain and appending the domain with a NotI restriction site, the stop codon, and an EcoRI restriction site. The plasmid pSMa76 was used as the template. The PCR product was digested with NdeI and EcoRI and inserted into pET28a to yield pWJ249.

KS_MAT Sequence (Active Site of KS Emphasized). TLESLTKDRA-EPNLAPGDENKIAIIGFSGRFPEADNLDEFWDLLIRG-LDVHKPVPEERFARDHYDPTGQRKNTSQVQYGCWLKS-AGYFDTQFFHMSPKEAMQTDPAQRLALLTAYEALEMA-GVVPDRTPSTQRNRVGVYYGTTSNDWGEVNSSQDVDT-YYIPGANRAFIPGRVNYFFKFTGPSIAVDTA*C*SSSLAAINL-AITSLKNRDCNTAIAGGTNVMTNPDNFAGLDRGHFLSR-TGNCKAFNDGADGYCRADGIGTLILKRLPDAIADSDPIF-GVILGAHTNHSAESVSITRPLADAQEYLFKKLLNETGIHP-HDVSYVEMHGTGTQAGDAVEMRSVLNSFAFDHSRPRD-KSLYLGSVKANVGHAESASGVLAIIKVLLMMQKNTIPPH-CGIKTKINQGFPKDLDHRGVRIALKDSVDWSRPEGGKR-RVLVNNFSAAGGNTSLLLEDGPAVHPARQHQDGDART-EHVVAVSARSTKALEENLKALEAYIANSWAPEGELLSQL-SYTTTARRVHHSRRVAFVTNGLDDLRKSLLKAATDAG-QVKGIPAVSPKVGFLFTGQGAQETAMAIGYYKSFSSFRS-DIHQLDSIATLQGLPSVLPLIHGTTPVEDLSAVVVQLGTC-IIQISLARFWISLGITPQYVIGHSLGEYAALQIAGVLSVND-AIFLCGHRAALLDKKCTAYTHGMVAVKAADDLRQH-ISSDLKVEIACVNGAEDTVLSGPNADIESLCGKLTQAGY-KLHKLEIPFAFHSSQVDPILDDLEELASQVGFHEPKLPIV-SPLLRTLLTGDTLGPQYIRRHCRETVDFLGAIKMAESQG-IMDRSGMCIEIGAHPILTRMVKSIIGQDFRCLASLRRKED-HFKTLADSLCALHLAGFSVNWDEYHRDFASSRNVLQLP-KYSWQLANYWMQYKYSWCLTKGDAPVENGPVGAVV-QARALRLSDSVHNVIEQVHGDKRSSITVESDMHDPS-LLAIAQNHRVNGLT

Cloning of a Stand-Alone PKS4 ACP Domain. The ACP domain was cloned as a stand-alone protein with primer pair pks4-ACP-F-NdeI and pks4-ACP-R-NotI-S-EcoRI, flanking the didomain with an NdeI restriction site preceding the domain and appending the domain with an EcoRI restriction site. The plasmid pSMa76 was used as the template. The PCR product was digested with NdeI and EcoRI and inserted into pET28a to yield pWJ248.

ACP Sequence. SVGTTSPPEPTESPVGSASGLIQKALEIIAD-EIGVDISQLTDTTLLADLGVDSLMSLTILGNFREELDLD-IPAAQFYEFSTVQDLKSFLGANDQDFSSSNSEAESSA-SSAASTSPS

Cloning of PKS_WJ. The ACP domain was recloned with primer pair PKS4ACP-F and PKS4ACP-R, and the linker was cloned with primer pair DEBS2M3-ACPL-F and DEBS2M3-R from *eryII* (encoding DEBS2). The 2 PCR products were assembled together by overlapping extension PCR and were amplified with primer pair DEBS2M3-ACPL-F and PKS4ACP-R. This SOE PCR product flanked with NotI sites was then digested with NotI, inserted into pWJ249 digested with NotI, and selected for the correct orientation to yield pWJ282.

Linker Sequence. LPGGAVDDGYLRERGLRSLSADRAMR-TWERVLAAGPVSVAVADVDWPVLSEGFAATRPTALFA-ELAGRGGQAEAEPDSGPTGEPAQR

Site-Directed Mutagenesis. Site-directed mutagenesis was used to generate the point mutants of the KS_MAT didomain. The mutagenic primers are PKS4KSS2A and PKS4KSS2A-r (shown in Table S1). The expression plasmid pWJ249 was used as the template for KS mutagenesis. Restriction analysis of the plasmid product was first performed to identify mutant, followed by DNA sequencing of the entire gene to verify the desired mutation.

Protein Expression and Purification. All proteins purified in this work contain N terminus hexahistidine tags. Transcription of the genes were regulated under T7 promoter and controlled by LacO operator. The expression plasmid was transformed into *E. coli* BL21(DE3) strain or BAP1 strain (contains a chromosomal copy of the *sfp* gene for phosphopantetheine transfer to apoACP) for protein expression. For 500 ml of liquid culture, the cells were grown at 37 °C in LB medium with appropriate concentration of antibiotics to an OD_{600} of 0.4–0.6, at which time the cells were incubated on ice for 10 min, and then induced with 0.1 mM isopropyl thio-β-D-galactoside (IPTG) for 16 h at 16 °C. The cells were harvested by centrifugation (3,500 rpm, 10 min, 4 °C), resuspended in 30 ml of lysis buffer (20 mM Tris·HCl, pH = 7.9, 0.5 M NaCl, 10 mM imidazole) and lysed by using sonication on ice. Cellular debris was removed by centrifugation (15,000 × *g*, 1 h, 4 °C). Ni-NTA agarose resin was added to the supernatant (4 ml/L of culture) and the solution was stirred at 4°C for at least 2 h. The protein resin mixture was loaded into a gravity flow column and proteins were purified with increasing concentration of imidazole in buffer A (20 mM Tris·HCl, pH = 7.9, 0.5 M NaCl, 5 mM imidazole, 2 mM DTT). Purified proteins were concentrated and buffered exchanged into Tris Buffer (50 mM Tris·HCl, pH = 7.9, 2 mM EDTA, 2 mM DTT) + 10% glycerol with Centriprep filter devices (Amicon). The final protein was concentrated, aliquoted, and flash frozen. Protein concentrations were determined with the Bradford assay by using BSA as a standard. Approximate yields: KS_MAT (3 mg/L), ACP (2 mg/L), PKS_WJ (1.6 mg/L).

Large-Scale Polyketide Synthesis and NMR. During the scaled-up reaction, malonyl-CoA was made in situ with MatB by using either sodium malonate or ^{13}C-3-labeled malonic acid. In a typical reaction, the following reagents were added to phosphate buffer (pH 7.4): 100 mM sodium malonate, 2 mM DTT, 7 mM $MgCl_2$, 20 mM ATP, 5 mM CoA, 20 μM MatB, 10 μM KS_MAT and 50 μM ACP or 10 μM PKS_WJ. 40 μM ZhuI was added for compound **6** production. The reaction mixture was shaken gently at room temperature and the reaction progress was followed by HPLC. After the product level had maximized, the reaction mixture was extracted 3 times with 99% EA/1% AcOH. The resultant organic extracts were combined and evaporated to dryness, redissolved in methanol. Compounds **3** and **5** were purified by reverse-phase HPLC (Alltima C18 5u, 250 mm × 10.0 mm) on a linear gradient of 30 to 50% CH_3CN (vol/vol) over 30 min and 95% CH_3CN (vol/vol) further for 15 min in H_2O supplemented with 0.1% (vol/vol) trifluoroacetic acid at a flow rate of 2 ml/min. Compound **6** was purified by reverse-phase HPLC (Alltech Apollo 5u, 250 mm × 4.6 mm) on a linear gradient of 10–50% CH_3CN (vol/vol) over 45 min and 95% CH_3CN (vol/vol) further for 15 min in H_2O supplemented with

0.1% (vol/vol) trifluoroacetic acid at a flow rate of 1 ml/min. NMR were performed on Bruker DRX-500 spectrometer by using CD_3OD as solvent.

Transacylation Activity Assay. The transfer of malonyl group from malonyl-CoA to acceptor ACPs was performed in phosphate buffer (pH 7.4). Holo-ACP domain was each added to a final concentration of 100 μM. KS^0_MAT (C514A) was added to a final concentration between 10 and 50 nM. [2-^{14}C]malonyl-CoA (55 nCi/nmol) was added to initiate the reaction at a final concentration of 180 μM. The reaction was performed at room temperature for 30, 60, and 90 s. A 10-μl aliquot was quenched with 10 μl of 2× SDS loading dye at each time point, separated on SDS/PAGE and the radiolabeled ACP bands were visualized by autoradiography.

High-density F1 fed-batch fermentation. Methods for F1 fed-batch fermentation and medium composition were adopted from methods described previously by Pfeifer *et al.* [Pfeifer B, Hu Z, Licari P, Khosla C (2002) Process and metabolic strategies for improved production of Escherichia coli-derived 6-deoxyerythronolide B. *Appl Environ Microbiol* 68:3287–3292]. A starter culture was grown overnight in 5 ml of LB medium at 37 °C and 250 rpm, and 1 ml was used to inoculate a 100-ml shake flask seed with F1 medium. Ten milliliters of the seed F1 culture was used to inoculate a 2-L Applikon Biobundle vessel containing 1 L of F1 medium. Fermentation was started at 37 °C, and the pH was maintained at 7.0 throughout the experiment with 1 M HCl and half-concentrated NH_4OH. Aeration was controlled at 0.2 to 0.5 L/min, and agitation was maintained at 1,400 rpm. When the OD_{600} reached ≈15–20, the temperature of the fermentation was reduced to 16°C, followed by the addition of 0.5 mM IPTG to induce protein expression. At the same time, a peristaltic pump delivered 0.1 ml/min of the feed solution to the fermentor. At each time point, a 0.5-ml aliquot was removed from the culture and the products were analyzed by HPLC.

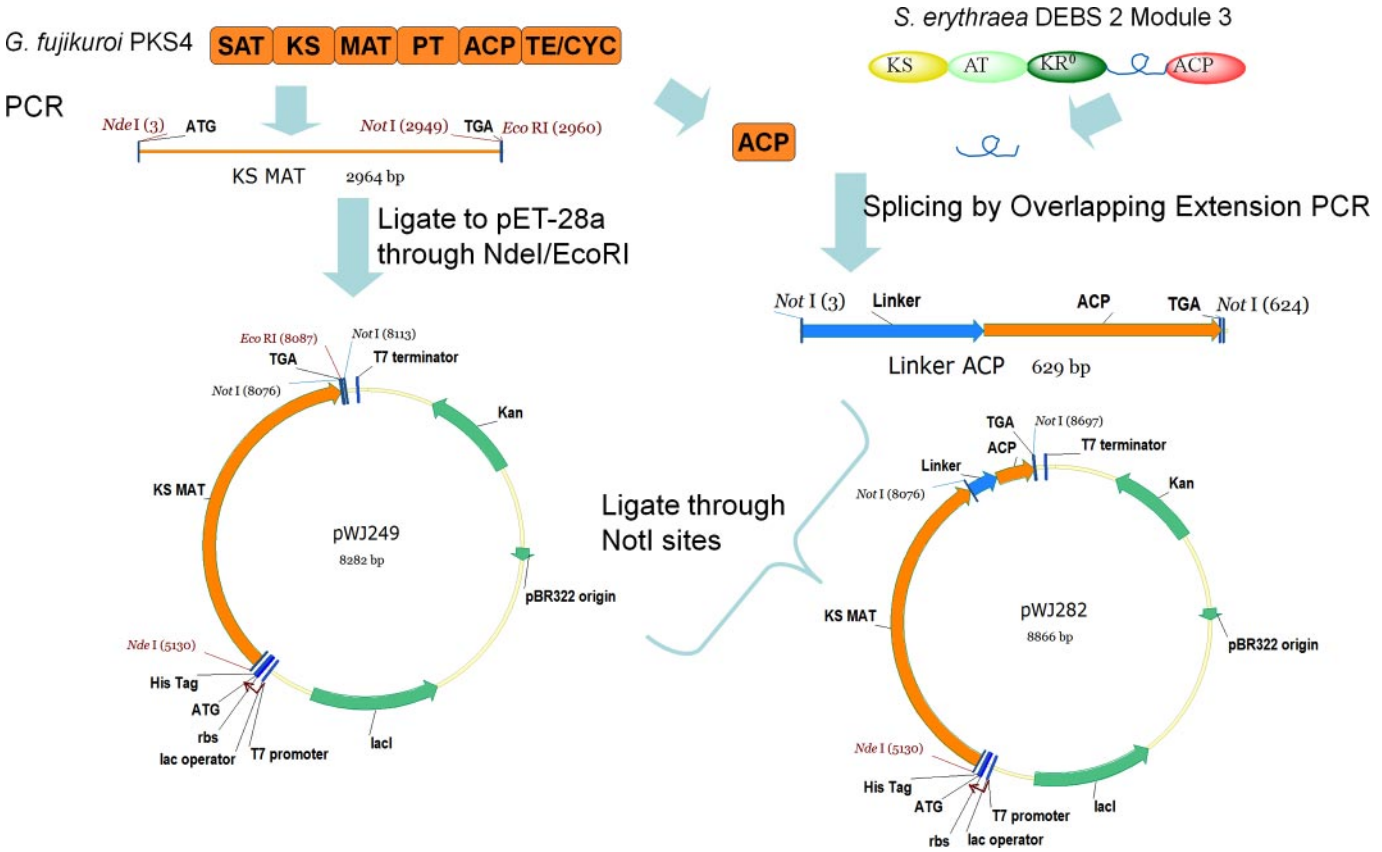

Fig. S1. Construction of PKS_WJ expression plasmid (pWJ282). PKS4 KS_MAT didomain was first inserted into pET-28a vector, followed by insertion of Linker_ACP didomain in the same ORF through NotI site.

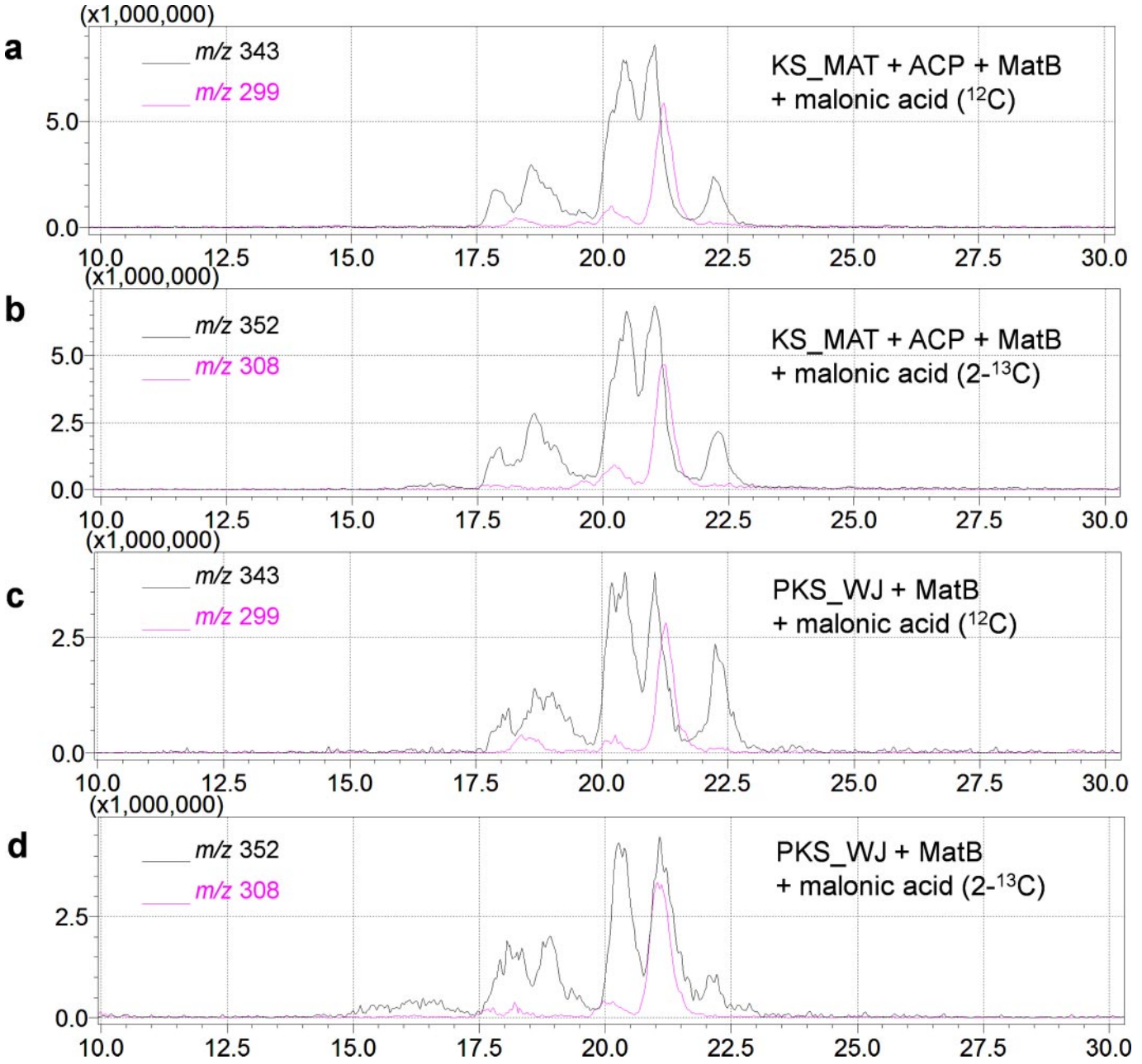

Fig. S2. LC/MS analysis and selected ion monitoring of nonaketides produced by KS_MAT + ACP and PKS_WJ in vitro. The molecular weight of 342 corresponds to nonaketides that have been subjected to 2 dehydration reactions, such as PK8 **3**. The molecular weight of 298 corresponds to nonaketides subjected to 3 dehydration reactions followed by one decarboxylation reaction, such as **5**. When ^{12}C-malonic acid was replaced by 2-^{13}C-malonic acid, 2-^{13}C-malonyl-CoA was synthesized by MatB, which would then be incorporated into polyketide products leading to a mass increase of + 1 for each ketide unit. For example, there should be a mass increase of 9 for nonaketides. The observed mass increase in *b* and *d* therefore confirmed the detected peaks were nonaketides. The traces in *a*–*d* also indicated that the products of KS_MAT and PKS_WJ were nearly identical during in vitro biosynthesis.

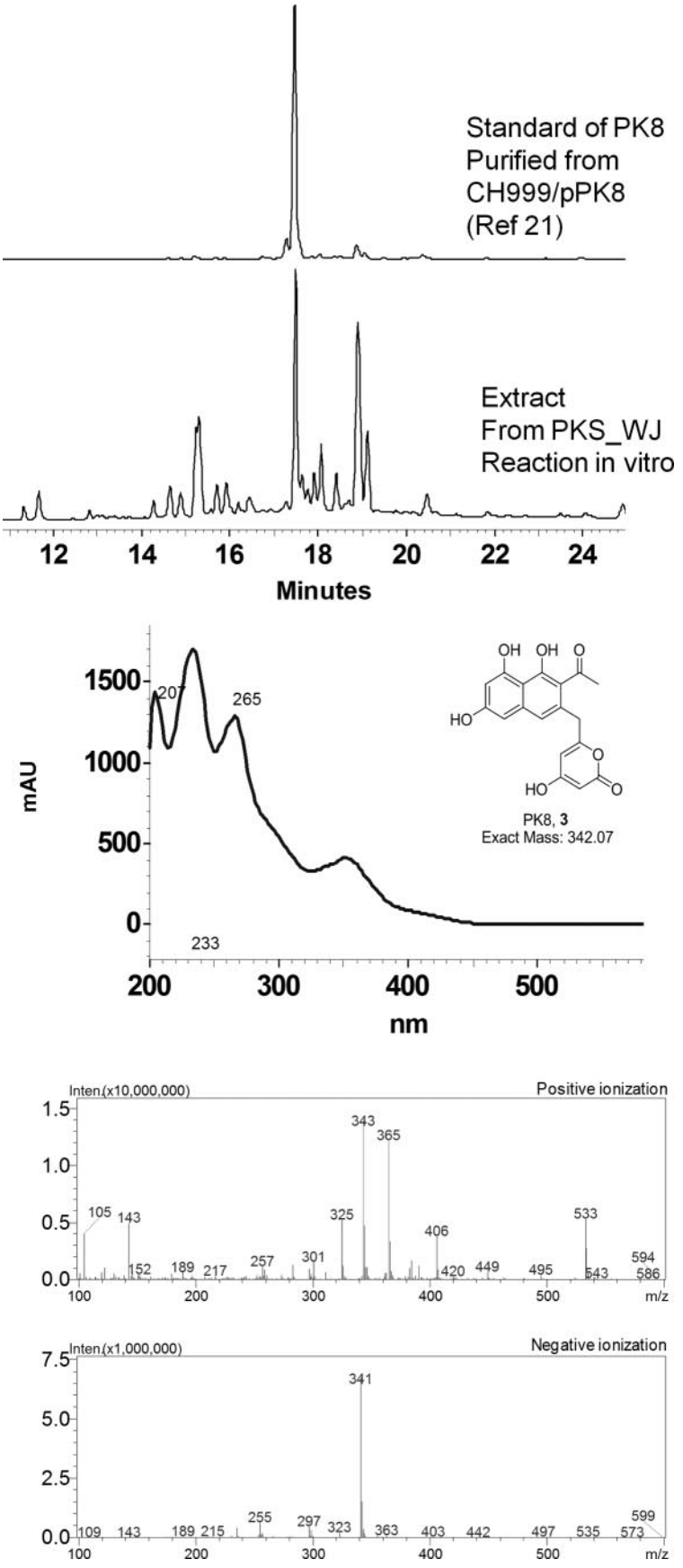

Fig. S3. HPLC comparison with standard PK8, UV spectrum, and MS of **3**.

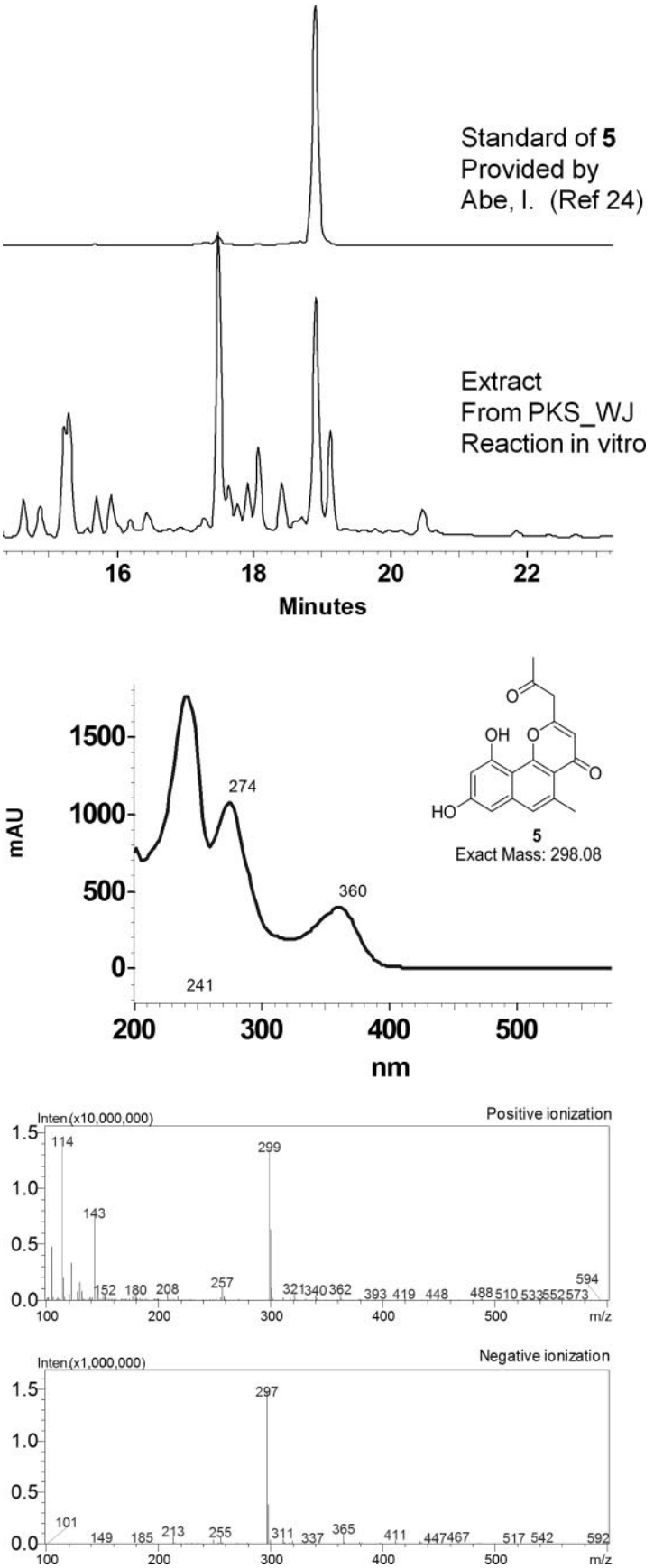

Fig. S4. HPLC comparison with standard, UV spectrum, and MS of **5**.

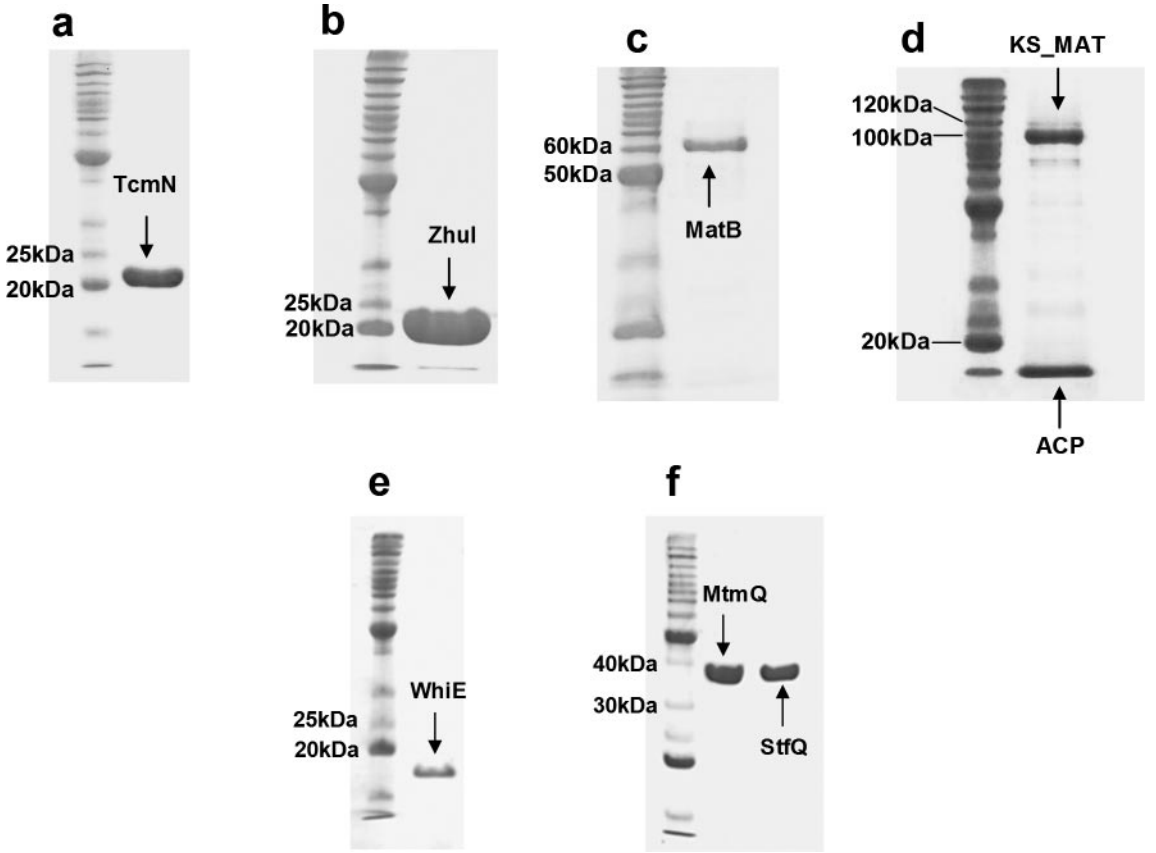

Fig. S5. SDS/PAGE analysis of the purified proteins. (*a*) TcmN cyclase domain (21 kDa), (*b*) ZhuI (19 kDa), (*c*) MatB (56 kDa), (*d*) PKS4 KS_MAT (108 kDa) and PKS4 ACP (13 kDa), (*e*) WhiE (19 kDa), and (*f*) MtmQ (36 kDa) and StfQ (36 kDa). All proteins have N-terminal hexahistidine tags. KS_MAT and ACP were co-expressed on the same plasmid under separate T7 promoters.

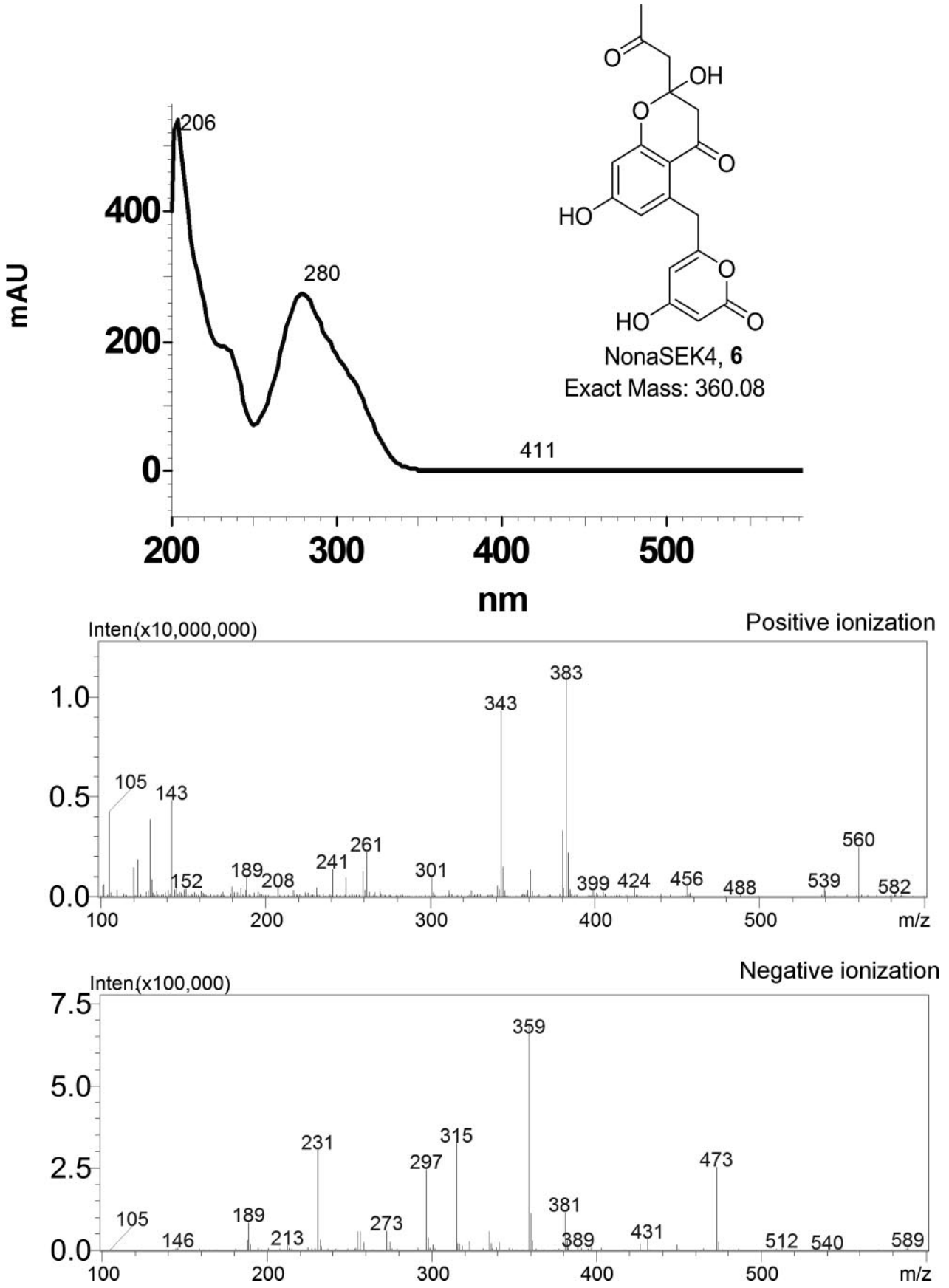

Fig. S6. UV spectrum and MS of **6**.

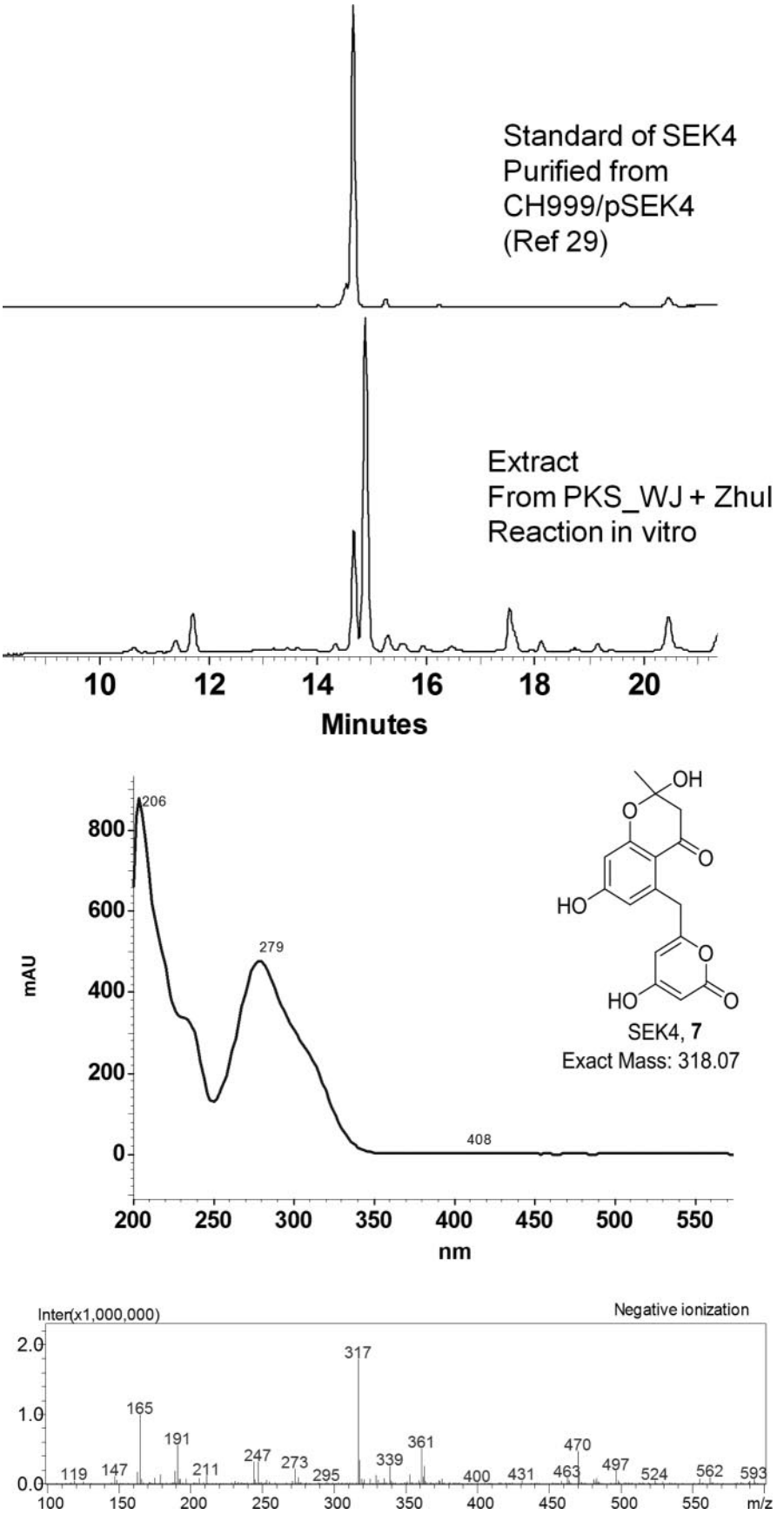

Fig. S7. HPLC comparison with standard, UV spectrum and MS of **7**.

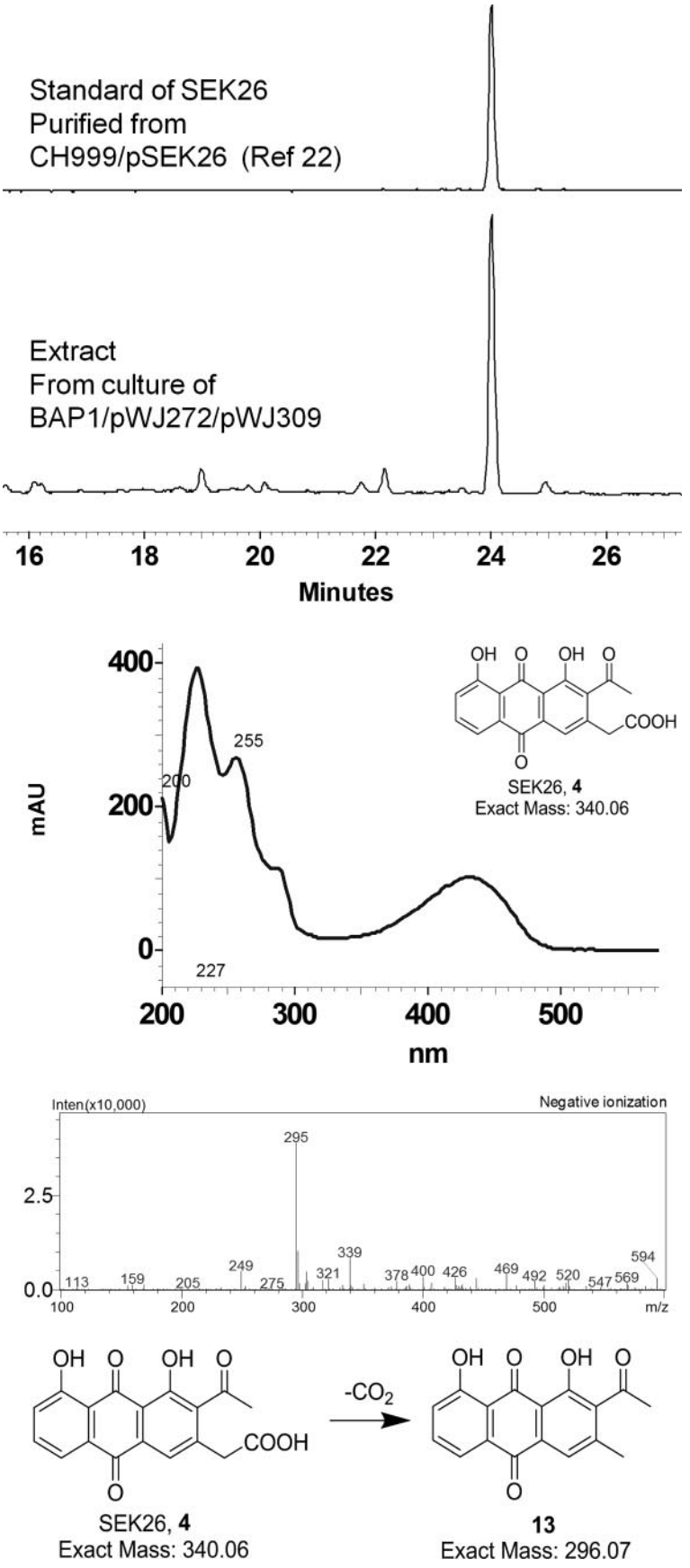

Fig. S8. HPLC comparison with standard, UV spectrum, and MS of **4**.

Table S1. Primers used in this study

Primer	Sequence
pks4-KS-F-NdeI	5′-AAACATATGACCTTGGAGAGCCTGACCAAA-3′
pks4-AT-R-NotI-S-EcoRI	5′-AAAGAATTCCTATGCGGCCGCAGTCAATCCGTTGACGCGATG-3′
pks4-ACP-F-NdeI	5′-AAACATATGAGCGTCGGCACAACATCACCT-3′
pks4-ACP-R-NotI-S-EcoRI	5′-AAAGAATTCCTATGCGGCCGCACTAGGGGATGTTGAAGCAGC-3′
DEBS2M3-ACPL-F	5′-AAGCGGCCGCACTGCCGGGCGGGGCGGTGG-3′
DEBS2M3-R	5′-GGTGATGTTGTGCCGACGCTCCGTTGTGCCGGCTCGCCGG-3′
PKS4ACP-F	5′-CCGGCGAGCCGGCACAACGGAGCGTCGGCACAACATCACC-3′
PKS4ACP-R	5′-AAGCGGCCGCTCAACTAGGGGATGTTGAAGCAG-3′
GrisARO-NdeI-His	5′-AAAACATATGAGCAGCCATCATCATCATCATCACAGCAGCATGTCGCAGCCCGGCCTGCG-3′
GrisARO-XhoI	5′-AAAAACTCGAGTCAGCCCCGCCGGCCCGCGG-3′
OxyN_5_NcoI	5′-AAAAAACCATGGCAATGCGCATCATCGATCTGTC-3′
OxyN_3_HindIII	5′-AAAAAAAAGCTTCTACTCCTCCACCACCGCC-3′
MtmQ_5_NdeI	5′-AAGTTTAAACGGAGGAGCCCATATGCCGAACAGCCAGCAGTACT-3′
MtmQ_3_SpeI	5′- AAACTAGTCAGACACCGGCGCGCTCCTC-3′
StfQ_5_NdeI	5′-AGTTTAAACGGAGGAGCCCATATGCCGACTCTGGGCAAAGA-3′
StfQ_3_NheI	5′- AGCTAGCTCAGACACCGGCGCGCTCCT-3′
PKS4KSS2A	5′-AGTATCGCTGTTGACACTGCAGCCTCTTCGAGTCTTGCTGCGA-3′
PKS4KSS2A-r	5′- TCGCAGCAAGACTCGAAGAGGCTGCAGTGTCAACAGCGATACT-3′

Restriction sites are underlined.

Table S2. Expression plasmids used in this study

Plasmid	Derived from	Function
pSMa99	pET28a	Expression plasmid for PKS4–99, the ″TE-less″ PKS4
pWJ248	pET28a	Expression plasmid for PKS4 ACP domain
pWJ249	pET28a	Expression plasmid for PKS4 KS_MAT didomain
pWJ282	pET28a	Expression plasmid for PKS_WJ (KS_MAT_Linker_ACP)
pWJ221	pET28a	Expression plasmid for TcmN cyclase domain
pWJ291	pET28a	Expression plasmid for ZhuI
pYR30	pET28a	Expression plasmid for MtmQ
pYR31	pET28a	Expression plasmid for StfQ
pETwhiE	pET28a	Expression plasmid for WhiE
pWJ292	pET28a	Expression plasmid for MatB
pWJ269	pET28a	Coexpression plasmid for KS_MAT and ACP
pWJ309	pET28a	Coexpression plasmid for PKS_WJ and *act* KR
pWJ272	pCDFDuet-1	Coexpression plasmid for *oxy* CYC2 and *gris* CYC1
pWJ268	pACYCDuet-1	Coexpression plasmid for ACP and *act* KR
pWJ298	pET28a	Coexpression plasmid for PKS_WJ and ZhuI

Table S3. NMR characterization of 6

No.	^{13}C δ, ppm	^{1}H δ, ppm (m, area, J_{HH} (Hz))
1	163.6	–
2	89.3	5.28(d, 1H, 2.0)
3	173.4	–
4	101.5	5.71(d, 1H, 2.0)
5	168.9	–
6	39.4	4.16(d, 1H, 16.0)4.22(d, 1H, 16.0)
7	140.1	–
8	115.0	6.40(d, 1H, 2.2)
9	165.2	–
10	104.3	6.32(d, 1H, 2.3)
11	167.4	–
12	113.1	–
13	196.7	–
14	50.1	2.66(d, 1H, 16.0)2.89(d, 1H, 16.0)
15	103.5	–
16	52.9	3.06 (s, 2H)
17	207.8	–
18	30.4	2.26 (s, 3H)

Spectra were obtained at 500 MHz for proton and 125 MHz for carbon and were recorded in CD_3OD. For carbon numbering, see Fig. 4*C*.