# Supporting Information

# Partial *In Vitro* Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis

James Kuo[a], Stephen R. Lynch[b,‡], Corey W. Liu[c,‡], Xirui Xiao[b] and Chaitan Khosla[a,b,d,*]
[a]Department of Chemical Engineering, [b]Department of Chemistry,
[c]Stanford Magnetic Resonance Laboratory, [d]Stanford ChEM-H,
Stanford University, Stanford, CA 94305 USA

## Contents

**Materials and Methods** ..... 2
- ***In Silico Analysis*** ..... 2
- ***Plasmid Construction*** ..... 2
- ***Protein Expression and Purification*** ..... 2
- ***In Vitro Assays*** ..... 3
- ***LC-MS Q-TOF*** ..... 3
- ***XCMS and Other Analysis*** ..... 4
- ***Radioassays – Radio-SDS-PAGE*** ..... 4
- ***Large Scale Reactions and HPLC Purification*** ..... 5
- ***NMR, 800 MHz*** ..... 5

**Supplementary Tables S1-S5** ..... 6

**Supplementary Data and Figures** ..... 11
- ***In Silico analysis of the NOCAP synthase gene cluster*** ..... 11
- ***Design and expression of proteins harboring Modules 4-8 and trans-AT-TE*** ..... 13
- ***Acyltransferase (AT) Substrate Specificity*** ..... 14
- ***Starter Unit Analysis*** ..... 15
- ***LC-MS, UV*** ..... 16
- ***NMR Spectra*** ..... 21
- ***MS/MS*** ..... 48

**SI References** ..... 56

## Methods and Materials

### *In Silico Analysis*

A list of 800+ previously identified assembly line PKS clusters [1] was manually evaluated. The antiSMASH 2.0 results [2] were used as a starting point for cluster composition. Chosen cluster sequences were aligned with five relatively well-characterized modules with canonical KS-AT-KR-DH-ER-ACP domain architectures: DEBS Module 4, Modules 4 and 5 of the epothilone synthetase, and Modules H and K of the curamycin synthase. Other domains (e.g., methyltransferase) were confirmed through BLAST searches. Key catalytic residues were identified for each domain to predict domain existence and likely activity.

### *Plasmid Construction*

Genes were amplified from *Nocardia* genomic DNA (*N. pneumoniae* DSM 44730; *N. abscessus* DSM 44432; *N. araoensis* DSM 44729; DSMZ Germany) by PCR and inserted into pET21 or pET28 vectors (Novagen) using Gibson assembly [3], or restriction enzymes (New England Biolabs) and T4 DNA ligase (Invitrogen). Oligonucleotides are listed in Table S5. Cloning was done in *E.coli* DH5α. Codon-optimized DNA for *N. pneumoniae* Modules 1 and 2 from Gen9, Inc. (Cambridge, MA) were used in comparison to genomic DNA constructs – protein expression was similar for both sets of DNA.

### *Protein Expression and Purification*

All proteins were expressed and purified according to procedures described previously [4]. Plasmids were introduced into *E. coli* BL21(DE3) or BAP1 [5] by electroporation. Single colonies were grown in 11 mL seed cultures of LB-Miller broth containing 100 µg/mL carbenicillin or 50 µg/mL kanamycin, and shaken at 37°C. Seed cultures were then inoculated into 1 L cultures grown under the same conditions. Cultures were transferred to 18°C at $OD_{600}$~0.5, induced with 100 µM IPTG at $OD_{600}$~0.5-0.8, and grown for an additional 17-20 h. Cells were harvested by centrifugation at 5,500 rpm (Beckman-Coulter Avanti J-E, JA-10 rotor) and resuspended in 30 mL Buffer A (50 mM sodium phosphate, 10% glycerol, pH 7.5) with 200 mM NaCl and 10 mM imidazole. Resuspended cells were lysed by sonication for 3-5 rounds of 45 second cycles with 1 pulse/sec (Branson Sonifier 450). Lysates were centrifuged at 15,000

rpm (JA-18 rotor) and incubated with Ni-NTA resin for ~1-2 h. Resin was applied to a column and washed with 25-40 mM imidazole in Buffer A and eluted with 300-500 mM imidazole in Buffer A. The eluted protein was further purified by anion exchange chromatography (HiTrapQ, GE Healthcare) using a linear gradient of 0-500 mM or 0-1 M NaCl in Buffer A. Fractions were analyzed by SDS-PAGE on TGX gels (Bio-Rad). Desired fractions were pooled, concentrated and exchanged into 100 mM phosphate pH 7.5 with Amicon Ultra centrifugal filters (Millipore). Aliquots were flash frozen with liquid nitrogen and stored at -80°C.

Proteins with ACP domains were produced in *E. coli* BAP1 to ensure phosphopantetheinylation to the active *holo* form [5], and verified by MALDI-TOF for stand-alone ACPs or by radioactive labeling experiments. Protein concentrations were determined by the BCA assay (Thermo Scientific).

***In Vitro Assays***

Purified proteins were mixed in 300 mM sodium phosphate pH 7.5 and 5 mM TCEP. When applicable, malonyl-CoA and/or NADPH were added in excess, along with S-adenosyl methionine (SAM) for assays involving Module 6 *C*-methyltransferase. For [$^{13}$C]-labeled malonyl-CoA, *S. coelicolor* malonyl-CoA synthetase MatB [6] was added at 2 μM, along with 6 mM ATP, 6 mM malonic acid and 2-4 mM CoASH. Proteins were assayed in the 3-5 μM concentration range in 1.7 mL Eppendorf tubes. For LC-MS analysis, reactions performed on 25-50 μL scale were extracted with ethyl acetate (3x 250 μL), dried by Speedvac, and reconstituted in methanol. For $UV_{340}$ measurements of NADPH cofactor consumption, reactions were mixed and initiated via addition of cofactors.

Kinetic analysis of AT proteins was performed using a coupled assay with malonyl-CoA, methylmalonyl-CoA or ethylmalonyl-CoA as substrates [7].

***LC-MS Q-TOF (Elizabeth Sattely Lab, Stanford University)***

Dried extracts from *in vitro* reactions were resuspended in 88 μL methanol (HPLC-grade, Fisher Scientific) with 5 μL injected into an LC-MS (Agilent 1260 HPLC with Agilent 6520 Accurate-Mass Q-TOF ESI mass spectrometer). A ZORBAX Eclipse Plus Phenyl-Hexyl (3.5 μm, 50 x 2.1 mm, Agilent) or Gemini NX-C18 (5 μm, 100 x 2.00 mm, Phenomenex) column was used, with separate analysis in positive ion (water and acetonitrile + 0.1% formic acid) and

negative ion (0.5 mM ammonium acetate and acetonitrile, no additives) modes. A 3-97% acetonitrile gradient over 20 min, followed by a 10 min 97% acetonitrile hold, was used.

$MS^n$ analysis involved 10, 20 or 40 V applied for fragmentation, with all voltages analyzed for major fragments.

***XCMS and Other Analysis***

XCMS analysis of MS data [8] was used in R to reveal differences between individual samples. Briefly, a continuous wavelet transform (CWT) peak finding method, *centWave*, was applied, which is superior to a matched filter Gaussian fitting approach in its ability to find features in spectra or to resolve nearby peaks. (The matched filter peak finding method was also used.) Detected features were subjected to two rounds of retention time correction, realigning traces to account for small drift variations between samples. Only traces collected in the same experimental set were directly compared this way. Feature lists were compared pairwise, filtering based on fold difference, p-value across replicates, and signal intensity to identify significant masses. Using the processed list, extracted ion chromatograms were obtained from the LC-MS traces and verified for ion signal intensity and peak shapes, evaluating the likeliness as an actual molecular feature. Lists of significant m/z values were compiled in this manner.

For each m/z value, possible molecular formulae were found using "Molecular Weight Calculator" by Matthew Monroe, calculating from $[M+H]^+$ or $[M+Na]^+$ forms for positive mode and only $[M-H]^-$ forms for negative mode, searching for molecules comprised of only C, H, O atoms. Formulae with the lowest calculated ppm difference between experimental and theoretical values were assigned to the peak, and were usually a much better fit than alternative formulae.

For [$^{13}$C]-labeled samples, lists of identified candidates were compared for 0-16 extensions (+1.003355 Da/neutron), identifying polyketides produced by the system.

***Radio-SDS-PAGE Analysis***

[$^{14}$C]-labeled substrates were substituted for unlabeled components in radioactive assays. This included various [$^{14}$C]-acyl-CoA species as well as [2-$^{14}$C]-malonyl-CoA with unlabeled acyl-CoAs. Starter units (acetyl-CoA, propionyl-CoA, butyryl-CoA, isobutyryl-CoA, isovaleryl-CoA, hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, lauroyl-CoA, myristoyl-CoA or palmitoyl-CoA) were pre-incubated for 1-2 h at room temperature. Experiments involving *trans*-AT proteins and

radiolabeled malonyl extender units were performed on ice for 10-20 min. Reactions were quenched with an equal volume of Laemmli sample buffer (Bio-Rad), and analyzed on 4-20% TGX gels (Bio-Rad) with SDS. Gels were washed with water, stained with SimplyBlue SafeStain (Invitrogen), dried (Bio-Rad) for 2 h at 80°C, and imaged with a Rita Star TLC Analyzer (Raytest).

***Large Scale Reactions and HPLC Purification***

*In vitro* reactions were scaled up to 12 mL total for isolation of **1**, **2** and **4** in aliquots of 1 mL in 5 mL Eppendorf tubes for 5 h. Modules 4-8 were added along with 3 µM *trans*-AT-TE and 2 µM *S. coelicolor* MatB. In reactions to produce methylated products **2** and **4**, 1.6 mM of SAM was also added. Each 1 mL reaction was extracted 3 times with 3 mL ethyl acetate, vortexed vigorously and spun down at 2,000 x *g* for 4 min. Ethyl acetate layers were pooled into a 30 mL scintillation vial and dried on a rotary evaporator. The residue was reconstituted in 700 µL methanol for HPLC (Agilent 1100 with 1260 diode array detector). Samples were separated on a Clipeus $C_{18}$ column (250 x 4.6 mm; Higgins Analytical, Inc.) with water (A)/acetonitrile (B) with 0.1% TFA. Gradient elutions were used: desmethyl heptaketide, 47-80%B over 15 min; methyl octaketide mix, 59-92%B over 15 min. Peaks were identified from characteristic 362 nm absorption, verified by LC-MS, and pooled. Purified compound was evaporated to remove most acetonitrile, then diluted to < 25% acetonitrile with water and lyophilized.

***NMR, 800 MHz (Stanford Magnetic Resonance Laboratory)***

All NMR spectra were acquired on an 800 MHz Agilent VNMRS spectrometer (Stanford Magnetic Resonance Laboratory). All samples were prepared in 250 µL deuterated solvent ($d_6$-DMSO or $CDCl_3$) in a matched symmetrical Shigemi tube (5 mm). Data shown was acquired at 25°C. $^1$H 1-D experiments tested at 37°C showed no appreciable sharpening of linewidths.

**Table S1.** List of *Nocardia* species whose genomes harbor the NOCAP synthase gene cluster. All organisms were isolated from patients diagnosed with nocardiosis [9-17].

| ***Nocardia* Species** | **NBRC** | **Genbank Accession ID (main PKS contig)** |
|---|---|---|
| *N. araoensis* | 100135 | BAFR01000107.1 |
| *N. abscessus* | 100374 | BAFP01000080.1 |
| *N. pneumoniae* | 100136 | BAGF01000151.1 |
| *N. exalbida* | 100660 | BAFZ01000063.1 |
| *N. asiatica* | 100129 | BAFS01000454.1 |
| *N. paucivorans* | 100373 | BAGE01000052.1 |
| *N. arthritidis* | 100137 | BDBB01000017.1 |
| *N. gamkensis* | 108242 | BDBM01000012.1 |
| *N. niwae* | 108934 | BDCK01000022.1 |

**Table S2.** Expression plasmids used in this study. DNA was either from *N. pneumoniae* (N151) or from *N. araoensis* (N107) in those cases where *N. pneumoniae* DNA could not be expressed as soluble proteins in *E. coli*. All genes were either amplified from genomic DNA or assembled from synthetic DNA by Gen9, Inc. (Cambridge, MA). Each gene was expressed from the T7 promoter using pET21 or pET28 vectors. The N- and C-termini of modules were flanked with docking domains from the 6-deoxyerythronolide B synthase (DEBS). A number of additional expression constructs were prepared and evaluated; only those used in experiments described in this report are listed here. KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; TE, thioesterase; TR, thioester reductase.

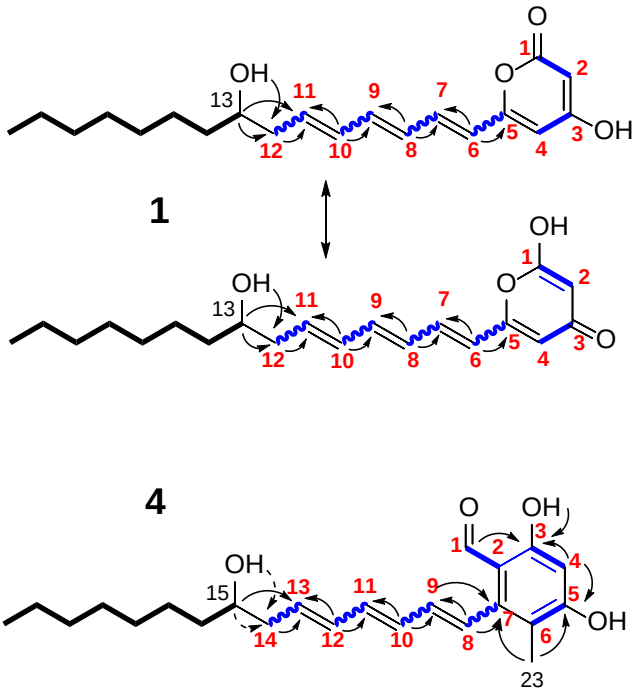
| **Name** | **Vector** | **Gene** | **Construct Notes** | **Resistance** |
|---|---|---|---|---|
| pJK21 | pET21 | N151_m1 full - DEBS(4) linker | Gibson assembly | Carbenicillin |
| pJK22 | pET21 | DEBS(5) - N151_m2 - DEBS(4) | Gibson assembly | Carbenicillin |
| pJK24 | pET21 | Gen9 N151_m1 full - DEBS(4) | Gibson assembly | Carbenicillin |
| pJK25 | pET21 | Gen9 DEBS(5) - N151_m2 - DEBS(4) | Gibson assembly | Carbenicillin |
| pJK26 | pET21 | Gen9 N151_m1-m2 - DEBS(4) | Gibson assembly | Carbenicillin |
| pJK32 | pET21 | N151_m1_ACP (C-6xH) | NdeI/HindIII | Carbenicillin |
| pJK33 | pET21 | DEBS(5) - N151_m3 - DEBS(4) | Gibson assembly | Carbenicillin |
| pJK35 | pET21 | N151_m5dh-m6 - DEBS(2) | Gibson assembly | Carbenicillin |
| pJK49 | pET21 | N151_transAT-TE (C-6xH) | NdeI/HindIII | Carbenicillin |
| pJK50 | pET21 | N151_transAT (C-6xH) | NdeI/HindIII | Carbenicillin |
| pJK52 | pET28 | N151_transAT-TE (N-6xH) | NdeI/HindIII | Kanamycin |
| pJK65 | pET21 | N151_m5dh-ACP-KR | NheI/HindIII | Carbenicillin |
| pJK66 | pET21 | N151_m5dh-ACP | NheI/HindIII | Carbenicillin |
| pJK71 | pET28 | N151_m2-ACP (N-6xH) | NdeI/EcoRI | Kanamycin |
| pJK72 | pET21 | N151_m2-ACP (C-6xH) | NdeI/HindIII | Carbenicillin |
| pJK94 | pET28 | DEBS(5) - N107_m4-5ks (N-6xH) | Gibson assembly | Kanamycin |
| pJK97 | pET21 | DEBS(3) - N107_m7-8-TR (C-6xH) | Gibson assembly | Carbenicillin |

**Table S3.** Extracted ions observed in reaction mixtures containing alternative acyl-CoA primer units of 8-16 carbon lengths.

| | Calcd [M+H]+ | Observed [M+H]+ | Δm/z (ppm) | Formula | Compound |
|---|---|---|---|---|---|
| C8 | 333.2061 | 333.2060 | 0.3 | $C_{20}H_{28}O_4$ | 1 |
| | 347.2217 | 347.2219 | -0.6 | $C_{21}H_{30}O_4$ | 2 |
| | 359.2217 | 359.2212 | 1.4 | $C_{22}H_{30}O_4$ | 3 |
| | 373.2373 | 373.2373 | 0.1 | $C_{23}H_{32}O_4$ | 4 |
| | 211.1329 | 211.1330 | -0.5 | $C_{12}H_{18}O_3$ | 5 |
| C10 | 361.2373 | 361.2372 | 0.3 | $C_{22}H_{32}O_4$ | 1b |
| | 375.2530 | 375.2525 | 1.2 | $C_{23}H_{34}O_4$ | 2b |
| | 387.2530 | 387.2511 | 4.8 | $C_{24}H_{34}O_4$ | 3b |
| | 239.1642 | 239.1644 | -0.8 | $C_{14}H_{22}O_3$ | 5b |
| C12 | 389.2686 | 389.2681 | 1.2 | $C_{24}H_{36}O_4$ | 1c |
| | 415.2843 | 415.2822 | 5.1 | $C_{26}H_{38}O_4$ | 3c |
| | 429.2999 | 429.2995 | 0.9 | $C_{27}H_{40}O_4$ | 4c |
| | 267.1955 | 267.1960 | -2.1 | $C_{16}H_{26}O_3$ | 5c |
| C14 | 417.2999 | 417.2993 | 1.4 | $C_{26}H_{40}O_4$ | 1d |
| | 295.2268 | 295.2266 | 0.7 | $C_{18}H_{30}O_3$ | 5d |
| C16 | 323.2581 | 323.2582 | -0.5 | $C_{20}H_{34}O_3$ | 5e |

**Table S4**. $^{1}H$ and $^{13}C$ chemical shifts (ppm) for polyketides **1, 2** and **4.** n.d., not determined.

| | Desmethyl Heptaketide (1), $d_6$-DMSO | | Methyl Heptaketide (2), $d_6$-DMSO | | Methyl Octaketide (4), $d_6$-DMSO | | Methyl Octaketide (4), $CDCl_3$ | |
|---|---|---|---|---|---|---|---|---|
| **Carbon** | **δH** | **δC** | **δH** | **δC** | **δH** | **δC** | **δH** | **δC** |
| **1** | - | n.d. | n.d. | n.d. | 9.54 | 191.6 | 9.92 | 189.9 |
| **2** | n.d. | n.d. | n.d. | n.d. | - | n.d. | - | n.d. |
| **3** | n.d. | n.d. | - | (177.8) | (12.52) | n.d. | (12.28) | 158.2 |
| **4** | (5.59) | (109.4) | - | n.d. | 6.01 | n.d. | 6.30 | n.d. |
| **5** | - | 155.3 | - | 150.8 | n.d. | n.d. | n.d. | 155.5 |
| **6** | 6.14 | 125.0 | 6.50 | n.d. | - | n.d. | - | n.d. |
| **7** | 7.15 | 126.1 | 7.11 | 125.4 | - | 142.2 | - | 138.8 |
| **8** | 6.01 | 126.0 | 6.08 | n.d. | 6.73 | n.d. | 6.67 | n.d. |
| **9** | 6.14 | 132.2 | 6.08 | 131.2 | 6.73 | 132.8 | 6.66 | 129.0 |
| **10** | 6.64 | 126.8 | 6.60 | n.d. | 6.12 | n.d. | 6.15 | n.d. |
| **11** | 5.89 | 135.1 | 5.87 | 134.6 | 6.10 | 131.2 | 6.16 | 126.5 |
| **12** | 2.27 | 40.9 | 2.24 | n.d. | 6.63 | n.d. | 6.57 | n.d. |
| **13** | 4.45 OH, 3.47 H | n.d. | 4.45 OH, 3.46 H | n.d. | 5.83 | 134.4 | 5.86 | 128.4 |
| **14** | n.d. | n.d. | n.d. | n.d. | 2.17 | n.d. | 2.25 | n.d. |
| **15** | n.d. | n.d. | n.d. | n.d. | 4.44 OH, 3.44 H | n.d. | 3.68 H | n.d. |
| **21** | - | - | 1.77 | n.d. | n.d. | n.d. | n.d. | n.d. |
| **23** | - | - | - | - | 1.94 | n.d. | 2.12 | n.d. |



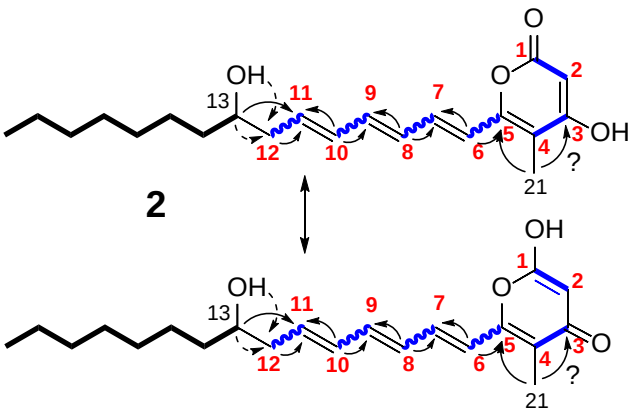



**Table S5.** Primers used for the construction of plasmids. Some primers, ex. pJK21-DEBS(4)-F, were used for multiple plasmids.

| Plasmid | Primer Name | Sequence (5' -> 3') |
|---|---|---|
| pJK21 | pJK21-N151m1(4)-F | TTTGTTTAACTTTAAGAAGGAGATATACAT ATGGCGGATGATGGGTACGA |
| | pJK21-N151m1(4)-R | CGATGTCCACCGCCGGTGAGGCCGCGAAGCTCAACCGGGCGATCT |
| | pJK21-DEBS(4)-F | TTCGCGGCCTCACCGGCGGT |
| | pJK21-DEBS(4)-R | AAGCTTGTCGACGGAGCTCGAATTCGGCAGGTCCTCTCCCCCG |
| | pJK21-(4)pET21-F | GAGGACCTGCCGAATTCGAGCTCCGTC |
| | pJK21-(4)pET21-R | CATGCAGCTCCCGGAGACG |
| | pJK21-pET21-F | CGTCTCCGGGAGCTGCATG |
| | pJK21-pET21-R | CCGCCATATGTATATCTCCTTCTTAAAG |
| pJK22 | pJK22-(5)N151m2(4)-F | TGCGTGAAGTCGAGCACCGGGCCGGTGACGCGATCGCGATCAT |
| | pJK22-(5)N151m2(4)-R | CGATGTCCACCGCCGGTGAGGCCGCGAAAGATTTGCGCTGCAAG |
| | pJK22-(5)pET21-R | ACCGGCCCGGTGCTCGACTT |
| pJK24 | pJK24-G9N151_f1-F | TTTGTTTAACTTTAAGAAGGAGATATACAT ATGGCAGACGACGGTTATGA |
| | pJK24-G9N151_f1-R | CGCGGCGGAATTACCGCGGT |
| | pJK24-G9N151_f2-F: | CAGGCGCTGATATTGATTGGAC |
| | pJK24-G9N151_f2-R | GCCGATGTCCACCGCCGGTGAGGCCGCGAAGCTTAAACGTGCGATCTCTT |
| | pJK24-(4)pET-F | GCAGTCAGTGAAGAGATCGCACGTTTAAGCTTCGCGGCCTCACCG |
| | pJK24-(4)pET-R | TACACGCAAATCATAACCGTCGTCTGCCATATGTATATCTCCTTC |
| pJK25 | pJK25-G9N151m2_f3-F | TGCGTGAAGTCGAGCACCGGGCCGGTGATGCCATCGCCATCATC |
| | pJK25-G9N151m2_f3-R | GCCACTGGGGTGTTGGCGGCCAAG |
| | pJK25-G9N151m2_f4-F | TGAGCGGCGTATGCCTCT |
| | pJK25-G9N151m2_f4-R | GCCGATGTCCACCGCCGGTGAGGCCGCGAATGACTTGCGCTGCAAGAA |
| | pJK25-(4)pET(5)-F | GATATTGCCCGCTTCTTGCAGCGCAAGTCATTCGCGGCCTCACCG |
| pJK26 | pJK26-G9N151_f2-R | GCTTAAACGTGCGATCTCTT |
| | pJK26-G9N151m2_f3-F | AACGGTTGCTGCAGTCAGTG |
| pJK32 | pJK32-N151m1ACP-R | GCCG AAGCTT GCTTAAACGTGCGATCTCTTCA |
| pJK33 | pJK33-N151m3-F | CGTGAAGTCGAGCACCGGGCCGGTGACCCGATCGTGATCGTCGG |
| | pJK33-N151m3-R | CGATGTCCACCGCCGGTGAGGCCGCGAACCCCAACTGCACCAGCAGAT |
| pJK35 | pJK35-N151m5dh6-F | CTTTAAGAAGGAGATATACATATGAGGGGAATCGACCTGGACA |
| | pJK35-N151m5dh6-R | CCTCCCCCCCGGACCTCGGTGCCGCCGTCTTCGGCCATACGCCTGC |
| | pJK34-(2)pET21-F | GCAGGCGTATGGCCGAAGACGGCGGCACCGAGGTCCGGGGGGAGG |
| | pJK34-(2)pET21-R | CCAGGTCGATTCCCCTCATATGTATATCTCCTTCTTAAAGTT |
| pJK49 | pJK49-ATTE-leader-F | ATATATCAT ATGAGCACCGAGCCTTCCGA |
| | pJK27-trAT-TE-R | CCGCAAGCTTAGCGGGTACTACGTCGGGGT |
| | pJK28-trATonly-R | CCGCAAGCTTTCCGGCCTGGATCGGAAGAGCC |
| | pJK29-trATlink-R | CCGCAAGCTTCAGGCTCGGCGTGCCCGCGGTC |
| pJK65 | pJK65-N151m5dhACPKR-F | ATATAT GCTAGC AGGGGAATCGACCTGGACAGCTTCG |
| | pJK65-N151m5dhACPKR-R | GCCG AAGCTT CTCGACGCCGACCGGTCTCGCCGG |
| pJK66 | pJK66-N151m5dhACP-R | GCCG AAGCTT GACTGTCTCGGCGAGGTGACCGGCCAGC |
| pJK71 | pJK71-N151m2-ACP-F | ATATAT CATATG TCTAAAGCACCTACCGATTCTGC |

| | | |
|---|---|---|
| | pJK71-N151m2-ACP-R | ATAT GAATTC CTATGACTTGCGCTGCAAGAAGCGG |
| pJK72 | pJK72-N151m2-ACP-R | GCCG AAGCTT TGACTTGCGCTGCAAGAAGCGG |
| pJK94 | pJK94-N107m45ks-F | GAAGTCGAGCACCGGGCCGGTGGGGACATCGCGATCGTGGGTGTC |
| | pJK94-N107m45ks-R | GCTCGAGTGCGGCCGCAAGCTTCATCGCAGCGGCTCCACCCGGAG |
| | pJK94-pET28-107-F | CTCCGGGTGGAGCCGCTGCGATGAAGCTTGCGGCCGCACTCGAGC |
| | pJK94-DEBS(5)107-R | GACACCCACGATCGCGATGTCCCCACCGGCCCGGTGCTCGACTTC |
| pJK97 | pJK97-N107m78-F | GCATCCGCGAGCTGGAATCCATGGACATCGCCATCATCGGC |
| | pJK97-DEBS(3)107-R | GCCGATGATGGCGATGTCCATGGATTCCAGCTCGCGGATGC |
| | pJK97-pET21-107-F | CGACCGCGGAGGTGGTGCGGCCGAATTCGAGCTCCGTCGACAAGC |
| | pJK97-N107m78-R | GCTTGTCGACGGAGCTCGAATTCGGCCGCACCACCTCCGCGGTCG |

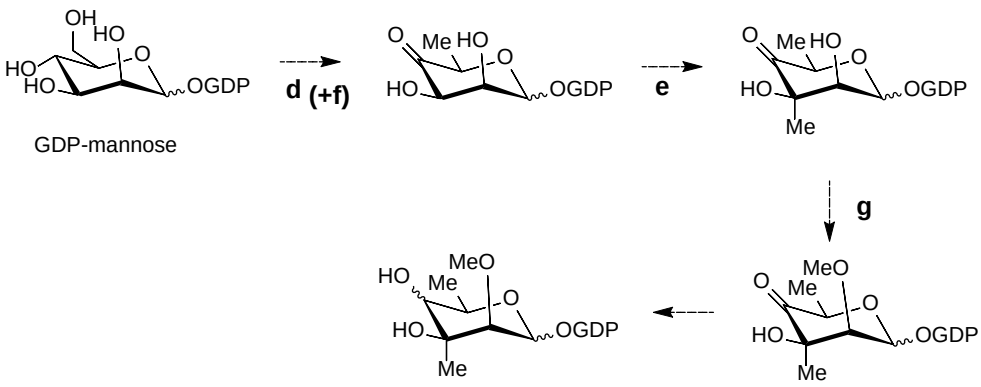
| **Plasmid** | **Primer Name** | **Sequence (5' -> 3')** |
|---|---|---|

### *In Silico Analysis of the NOCAP Synthase Gene Cluster*

In addition to the three genes (ORFS 1-3, Figure S1) encoding the NOCAP synthase, the gene cluster also encodes a unimodular protein (Module X) comprised only of KS and ACP domains. The two PKS gene sets are separated by several genes encoding putative tailoring enzymes, principally associated with sugar biosynthesis and transfer. None of these genes have close homologs in *Nocardia* or other actinomyces strains other than those listed in Table S1.


| ORF | Description | |
|---|---|---|
| | ***PKS Genes*** | |
| **1** | *nocap_PKS1* (Modules 1-KS5) | |
| **2** | *nocap_PKS2* (Modules 5(b)-8) | |
| **3** | *nocap_PKS3* (trans-AT-TE) | |
| **4** | *nocap_PKS4* (Module X (KS-ACP)) | |
| | ***Non-PKS Genes (Putative Functions)*** | ***Closest Characterized Homolog (UniProtKB)*** |
| **a** | Cupin-like superfamily, mannose-6-phosphate isomerase | Conserved domains (smart00835; COG0662) |
| **b** | NAD dependent epimerase/dehydratase family, nucleoside-diphosphate-sugar epimerase family | Conserved domains (pfam01370; COG0451) [18] |
| **c** | Dolichyl-phosphate beta-D-mannosyltransferase | *Mycobacterium tuberculosis* (A0QZ12.1)[19] |
| **d** | GDP-mannose 4,6-dehydratase | *Escherichia coli* O157:H7 (P0AC90.1)[20] |
| **e** | C-Methyltransferase | *Streptomyces niveus* novU (novobiocin cluster) (Q9L9E7.2)[21]; *Streptomyces fradiae* tylCIII (tylosin biosynthesis, not in UniProtKB) (AAD41823.1) [22] |
| **f** | GDP-mannose 4,6-dehydratase | *Arabidopsis thaliana* (Q9SNY3.1)[23] |
| **g** | O-Methyltransferase, sugar | *Streptomyces fradiae* tylE (tylosin biosynthesis) (Q9ZHQ4.1)[24] |
| **h** | Glycosyltransferase | *Streptomyces olivaceus* (Q9F2F9.2)[25,26] |
| **i** | Mannosyltransferase, dolichyl-phosphate-mannose-protein mannosyltransferase, transmembrane (pimE) | Conserved domains (pfam09594; PRK13375) [27] |
| **j** | Putative flavin-containing monooxygenase | *Mus musculus* (Q8K2I3.3)[28] |
| **k** | Cytochrome P450 (P450-derived glycosyltransferase activator) | *Streptomyces coelicolor* A3(2) (Q9K498.1)[29]; Conserved domain (TIGR04515) |
| **m** | Sugar translocase (GtrA superfamily) | Conserved domain (pfam04138; COG2246) [30] |

**Figure S1.** The NOCAP synthase gene cluster consists of 4 PKS-related and 12 non-PKS genes, found in all *Nocardia* species of Table S1. The cluster shown is from *N. araoensis*.




**Scheme S1.** Proposed deoxysugar biosynthesis by enzymes listed in Figure S1. The products of ORFs **a** and **b** (not shown) may convert a metabolically available hexose precursor into GDP-mannose. A putative heterodimer encoded by ORFs **d** and **f** is thought to be a GDP-mannose 4,6-dehydratase. Based on their homology to enzymes involved in tylosin biosynthesis, ORF **e** presumably encodes a C-3 C-methyltransferase, whereas ORF **g** is predicted to be an O-methyltransferase. The glycosyltransferase **h** is homologous to elmGT involved in elloramycin biosynthesis [25], while ORF **k** encodes a P450 homolog that has a conserved domain of a glycosyltranferase activator. An unidentified gene product is anticipated to reduce the 4-keto-6-deoxysugar.

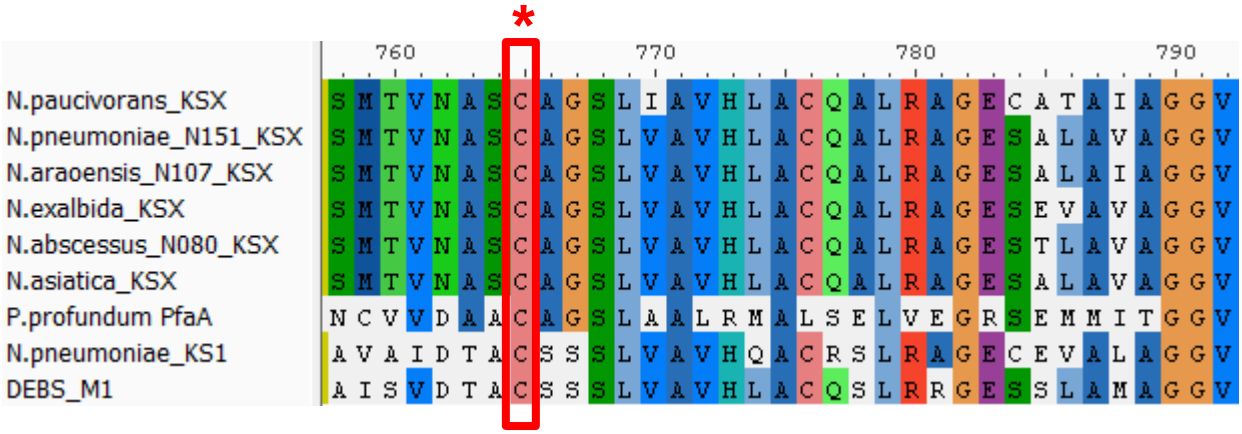



**Figure S2.** Sequence alignment of the active site ketosynthase (KS) domains of Module X from six *Nocardia* strains listed in Table S1. The active site Cys residue (position 765, *N. pneumoniae* numbering) is highlighted in a red box, and is part of an unusual S**C**AGS motif found in fewer than 200 of ~32,500 KS domains. This set of KS domains is most closely related to those found in polyunsaturated fatty acid synthases, such as the PfaA PKS found in *P. profundum* (AAL01060.1) [31]. For comparison, the sequences of the KS domains from Module 1 of the *N. pneumoniae* NOCAP synthase and Module 1 of the 6-deoxyerythronolide B synthase (AAV51820.1) are shown. The linker region between the KS and ACP domains of Module X may possibly be involved in AT-docking [32], but they are distinct from the other putative *trans*-AT docking regions found Modules 2, 4, 6, 7 and 8 of the NOCAP synthase.

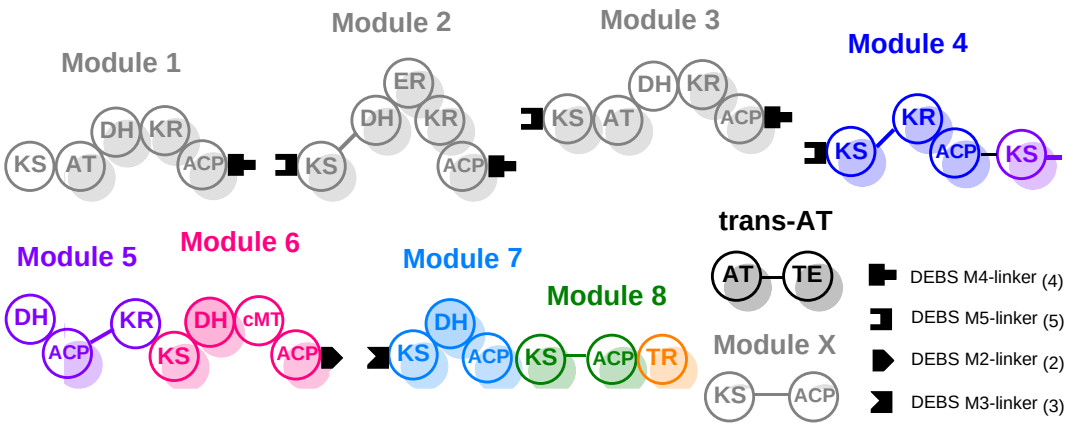



**Figure S3.** Proteins comprising the reconstituted pentamodular derivative of the NOCAP synthase harboring Modules 4-8 and the *trans*-AT-TE (highlighted in color) and other pieces (shown in gray). Docking domains from the 6-deoxyerythronolide B synthase (shown as black tabs) were fused onto the C-terminus of ACP6 and the N-termini of KS4 and KS7. KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; ACP, acyl carrier protein; cMT, C-methyltransferase; TR, NAD-dependent thioester reductase; TE, thioesterase.

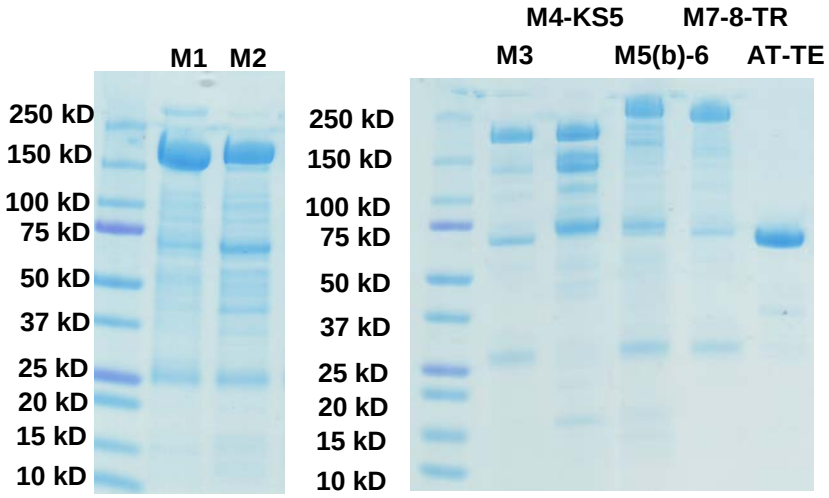



**Figure S4.** SDS-PAGE of expressed *Nocardia* PKS proteins, with module number shown. M1, pJK21; M2, pJK25; M3, pJK33; M4-KS5, pJK94; M5(b)-6, pJK35; M7-8-TR, pJK97; AT-TE, pJK49. For proteins larger than 200 kDa, putative proteolytic fragments are observed to co-purify with the intact protein on ion exchange and gel filtration chromatography columns, suggesting that the fragments remain associated even after proteolysis.

***Acyltransferase (AT) Substrate Specificity***

Malonyl-, methylmalonyl-, or ethylmalonyl- CoA were tested as AT substrates using the α-ketoglutarate dehydrogenase coupled kinetic assay [7]. Different *trans*-AT constructs (pJK49, 50, 52) were tested with ACP2 (pJK71, 72). In all cases, the AT only showed transacylase activity with malonyl-CoA (Fig. S5). No activity was observed with methylmalonyl- or ethylmalonyl- CoAs (data not shown). The *trans*-AT-TE protein had considerably lower activity in the presence of ACP1 (pJK32). Both N- and C- terminal $His_6$-tagged forms of *trans*-AT-TE showed comparable kinetics (Fig. S5). The overall turnover rates of the ATs were in line with those of other assembly-line PKSs [33]. The coupled assay results confirm malonyl-CoA selectivity for the ATs in the *Nocardia* system, with no evidence of methylmalonyl- or ethylmalonyl-CoA usage.

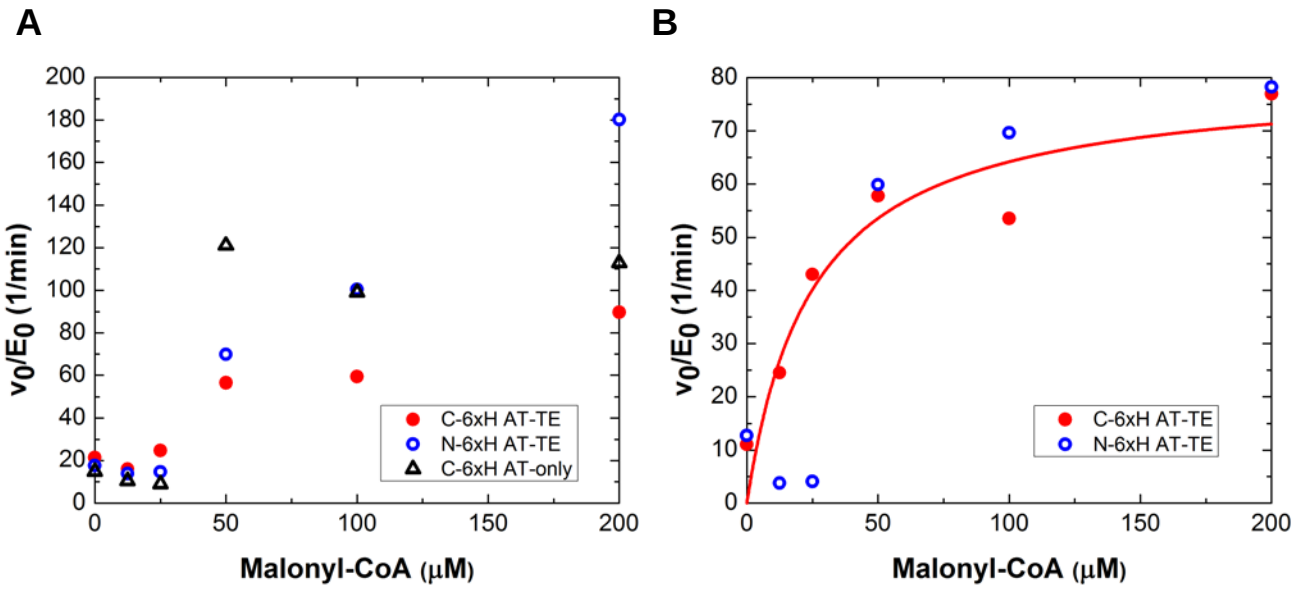



**Figure S5.** Coupled AT assay results for *trans*-AT-TE, with partner ACP2 ($His_6$ tag at (A) N-terminus or (B) C-terminus) or without TE domain. Michaelis-Menten fit parameters: Vmax ~ 80 $min^{-1}$, Km~25 μM.

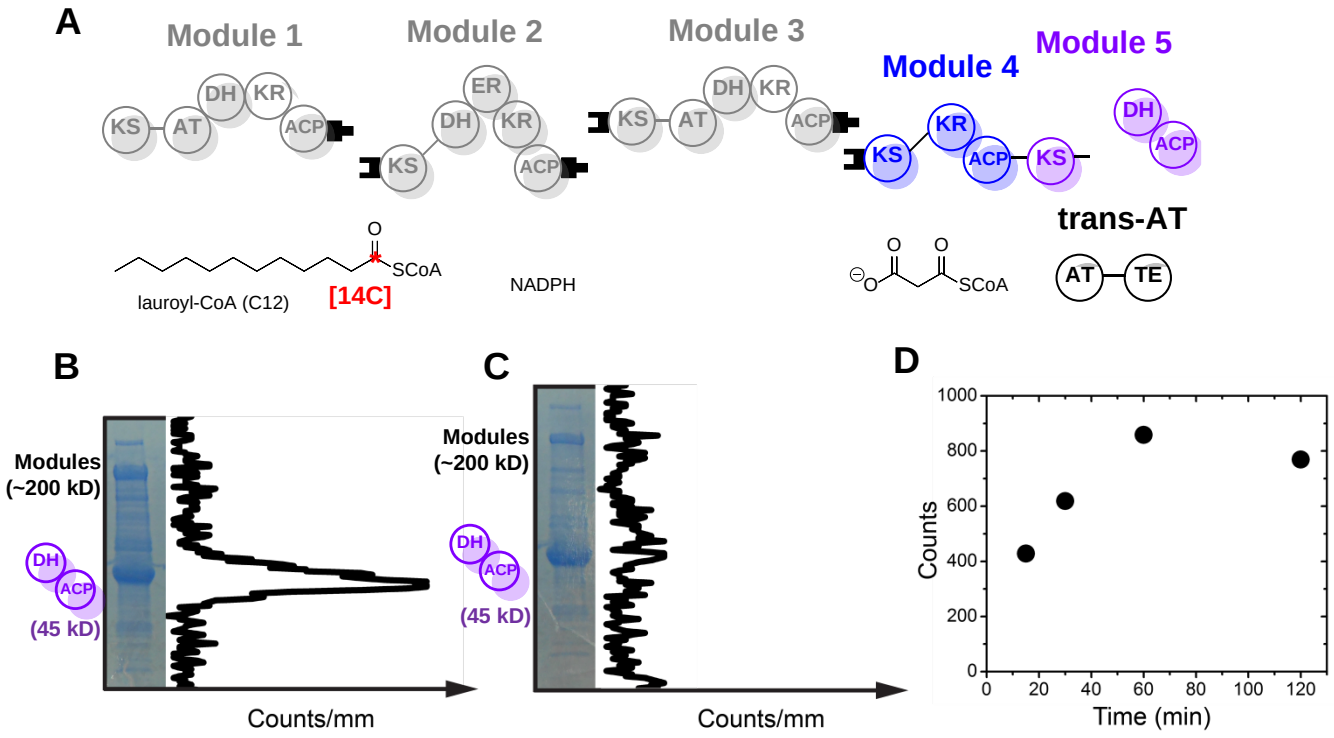



**Figure S6. Module 5(b) transfer requires Module 4-KS5.** Module 1-5 *in vitro* partial system, using Module 5(b) DH-ACP didomain as an endpoint. Reaction included [1-$^{14}$C]-lauroyl-CoA (C12), with incorporation read by radio-SDS PAGE. (A) Modules 1-3 and the trans-AT-TE were incubated (B) with Module 4-KS5 protein or (C) without Module 4-KS5. Labeling was also dependent on malonyl-CoA and NADPH being present, but not on Modules 1-3 (data not shown). (D) Labeling increased over time with Module 4-KS5 present.

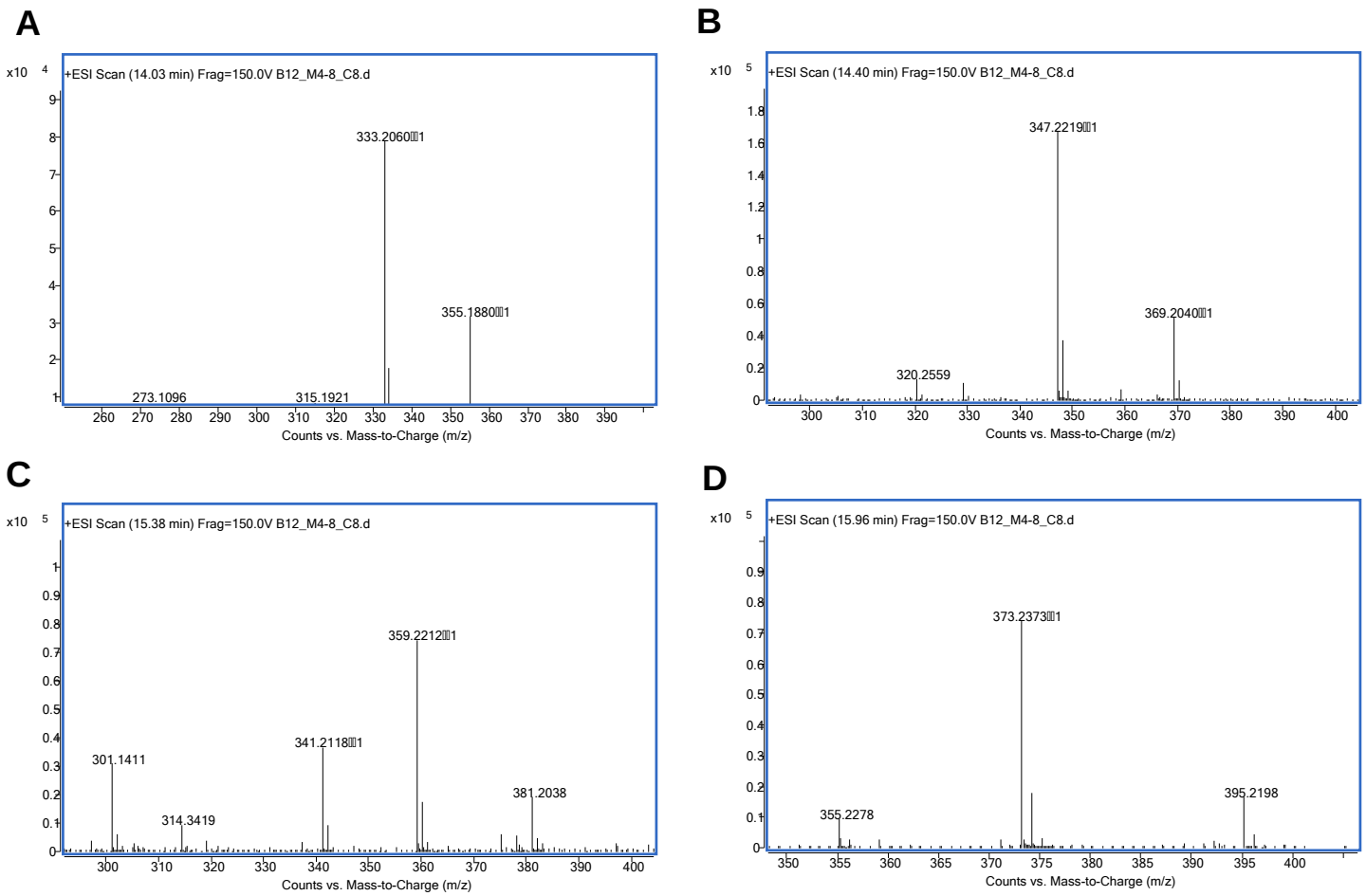



**Figure S7. Exact mass spectra of octanoyl-CoA derived polyketides 1-4**: (A) desmethyl heptaketide (theoretical $[M+H]^+$ = 333.2061), (B) methyl heptaketide (th. $[M+H]^+$ = 347.2217), (C) desmethyl octaketide (th. $[M+H]^+$ = 359.2217), (D) methyl octaketide (th. $[M+H]^+$ = 373.2373).

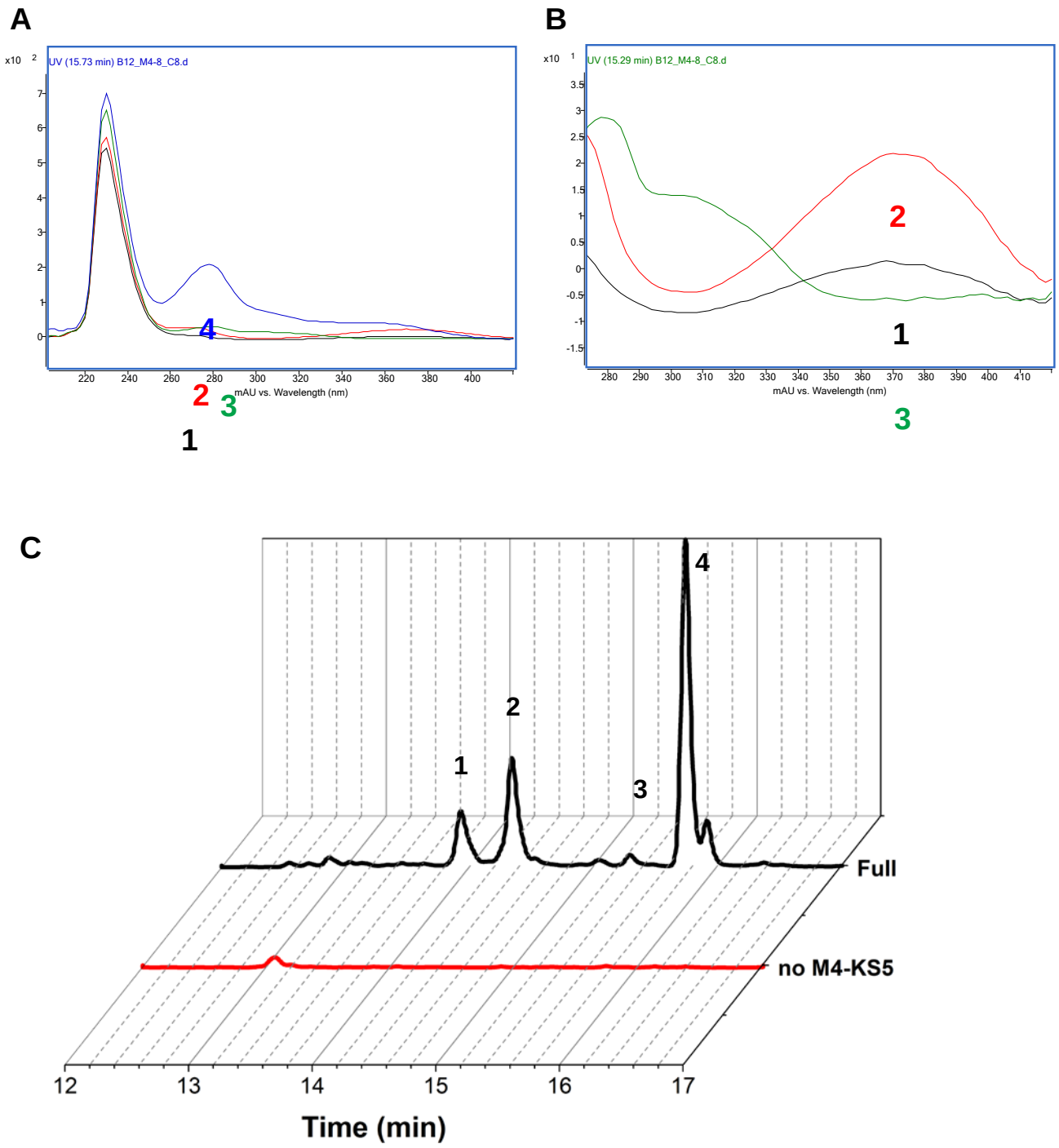



**Figure S8. UV spectra of polyketides 1-4**. Full UV (A) and zoom-in (B) spectra are shown for compounds **1** (black), **2** (red), **3** (green) and **4** (blue). (C) Absorbance at 350.8 nm for the Full Module 4-8 *in vitro* system and with no Module 4-KS5 protein. Conditions same as in Figure 2.

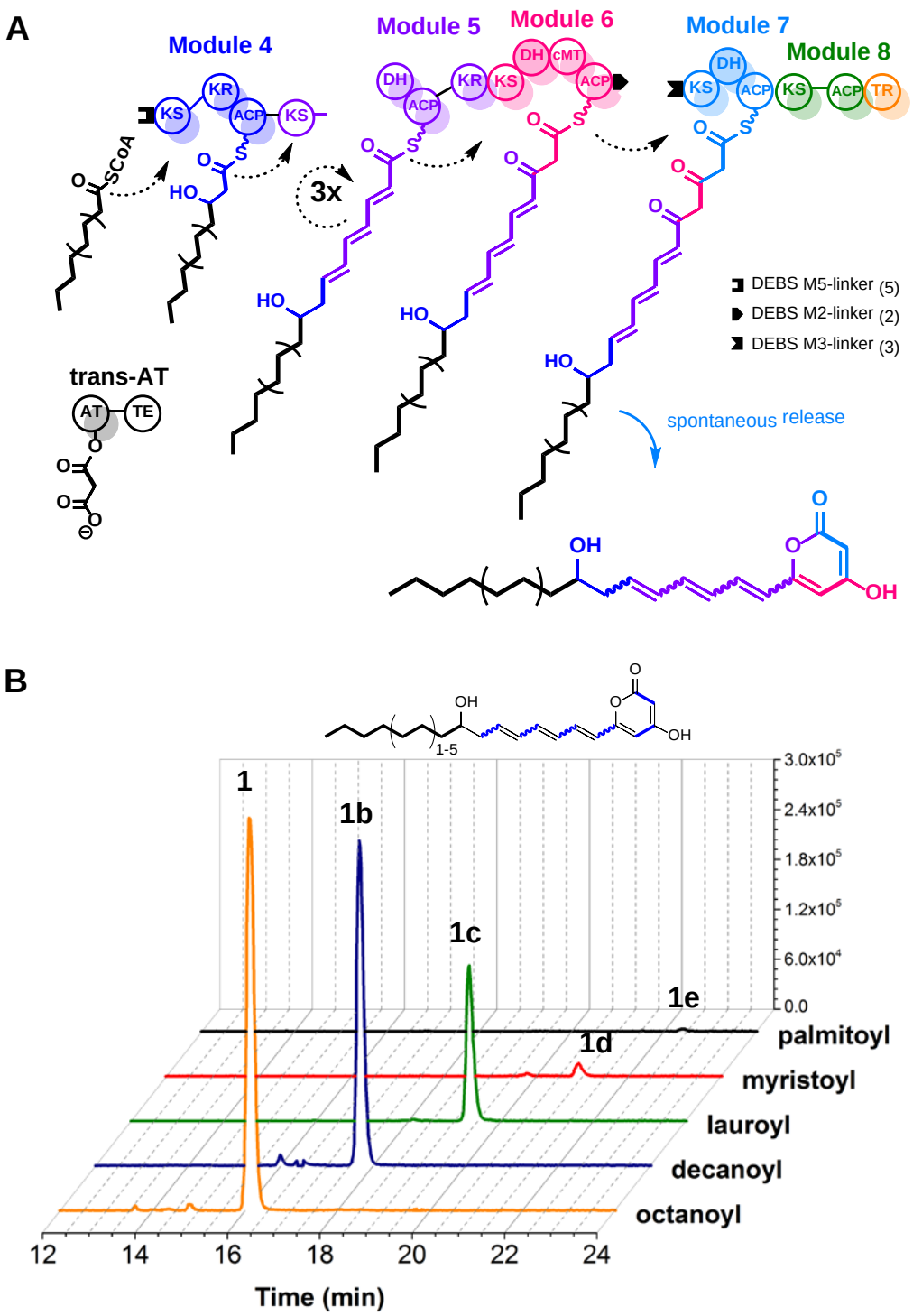



**Figure S9. Heptaketides synthesized by Modules 4-8 in the presence of alternative primer units**. (A) Desmethyl heptaketides are produced by Modules 4-7, with spontaneous release after Module 7 extension. (B) Representative LC-MS traces are shown for analogs of the desmethyl heptaketide **1**, which appears as the major product in the absence of SAM. Extracted ion chromatograms ($[M+H]^+$ and $[M+Na]^+$) of corresponding heptaketides from acyl-CoA primers in the $C_8$-$C_{16}$ range.

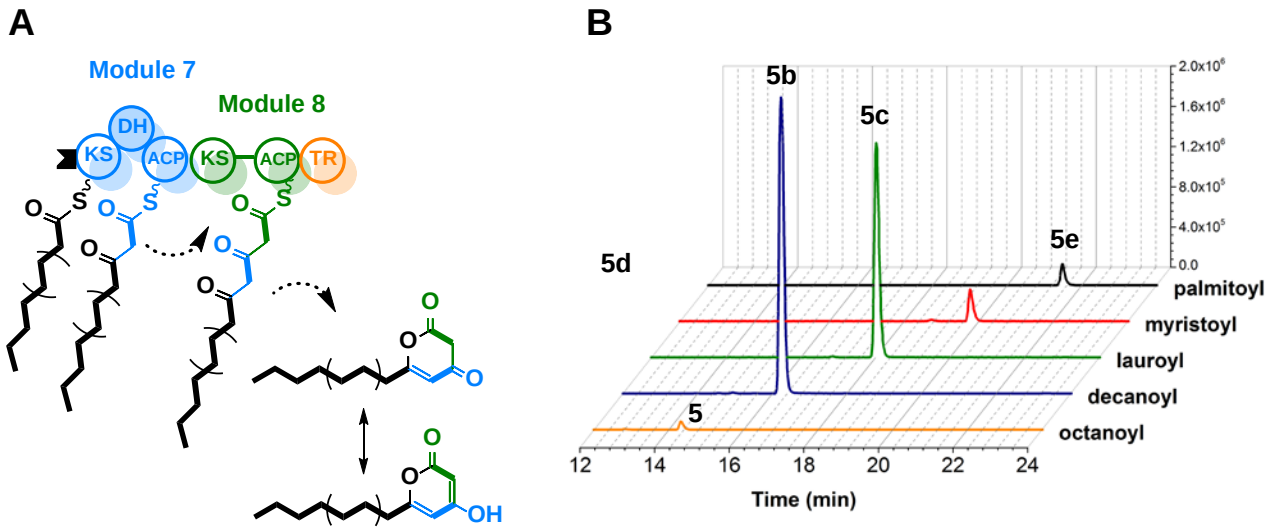



**Figure S10. Triketide side products found by LC-MS**. (A) Modules 7-8 produce a fatty α-pyrone triketide (**5** with octanoyl unit) through downstream priming of Module 7 with acyl-CoA. (B) Extracted ion chromatograms ($[M+H]^+$ and $[M+Na]^+$) of corresponding triketides from acyl-CoA primers in the $C_8$-$C_{16}$ range.

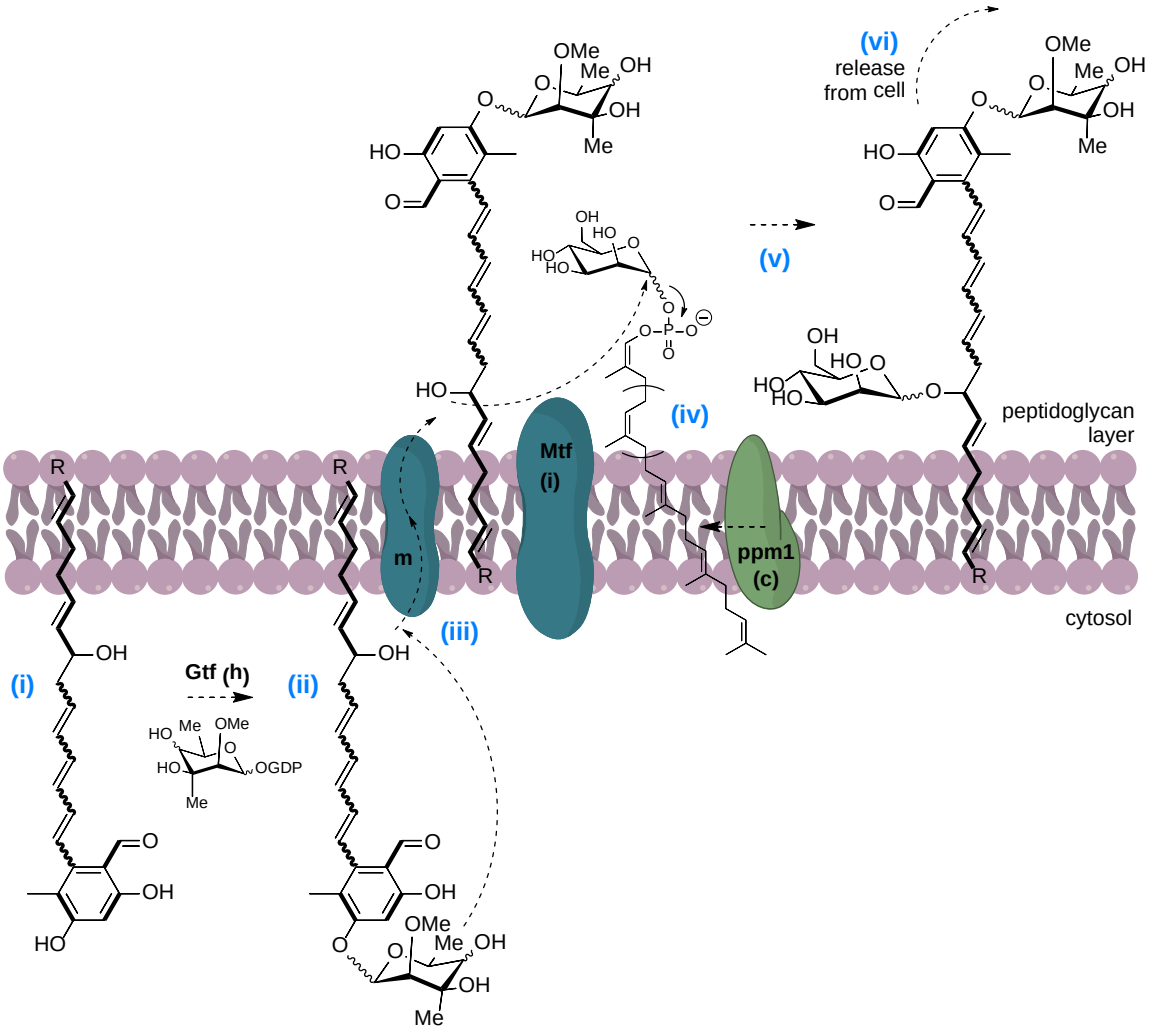



**Figure S11. Proposed polyketide decoration and export.** From *in silico* analysis of the sugar-related genes in the cluster, (i) the resorcylaldehyde product of the NOCAP synthase may partition into the inner leaflet of the *Nocardia* cell membrane. (ii) A deoxysugar (Scheme S1) may be attached to a resorcinol hydroxyl by a glycosyltransferase (Gtf). (iii) Glycosylation allows the product to be transported across the membrane by a sugar translocase (m). (iv) Meanwhile, polyprenol monophosphomannose synthase (ppm1) synthesizes dolichyl mannosyl phosphate, which is "flipped". (v) It acts as a sugar donor to the monoglycosylated polyketide in a reaction catalyzed by membrane-bound mannosyltransferase (Mtf). (vi) The diglycosylated product is ultimately released. The resorcylaldehyde has multiple nucleophilic sites (3-, 5- and 15-OH), two of which may be glycosylated while the third may be acylated by the polyketide product of Module X. Dual glycosylations with mannose-derived sugars could be envisioned, the second of which occurs only after export by a dedicated translocase and is catalyzed by a transmembrane transglycosylase that uses a dolichyl sugar substrate (Figure 4). Such a biosynthetic strategy could be particularly effective at "weaponizing" the aglycone core after it exits the producer cell.

## NMR Spectra

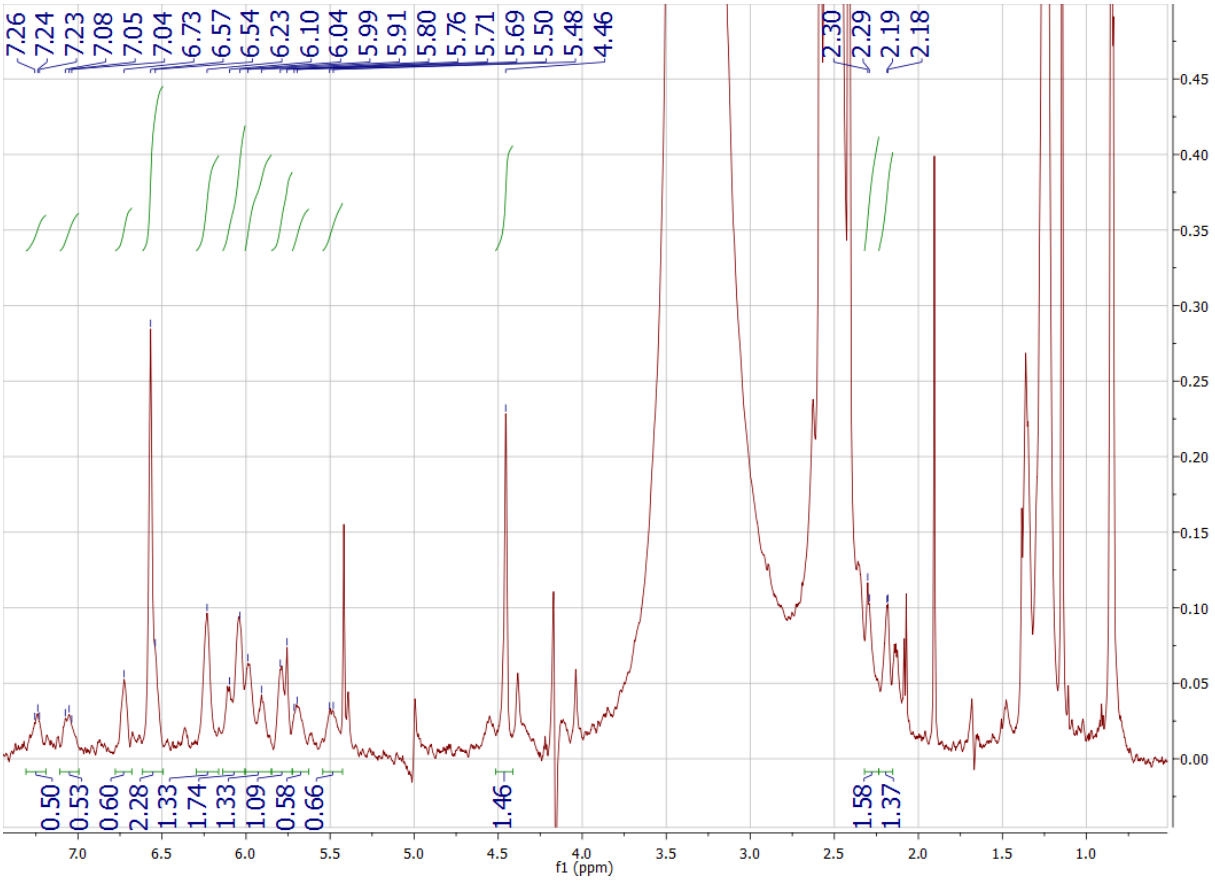



**Figure S12.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, $^{1}$H NMR (800 MHz, $d_6$-DMSO).

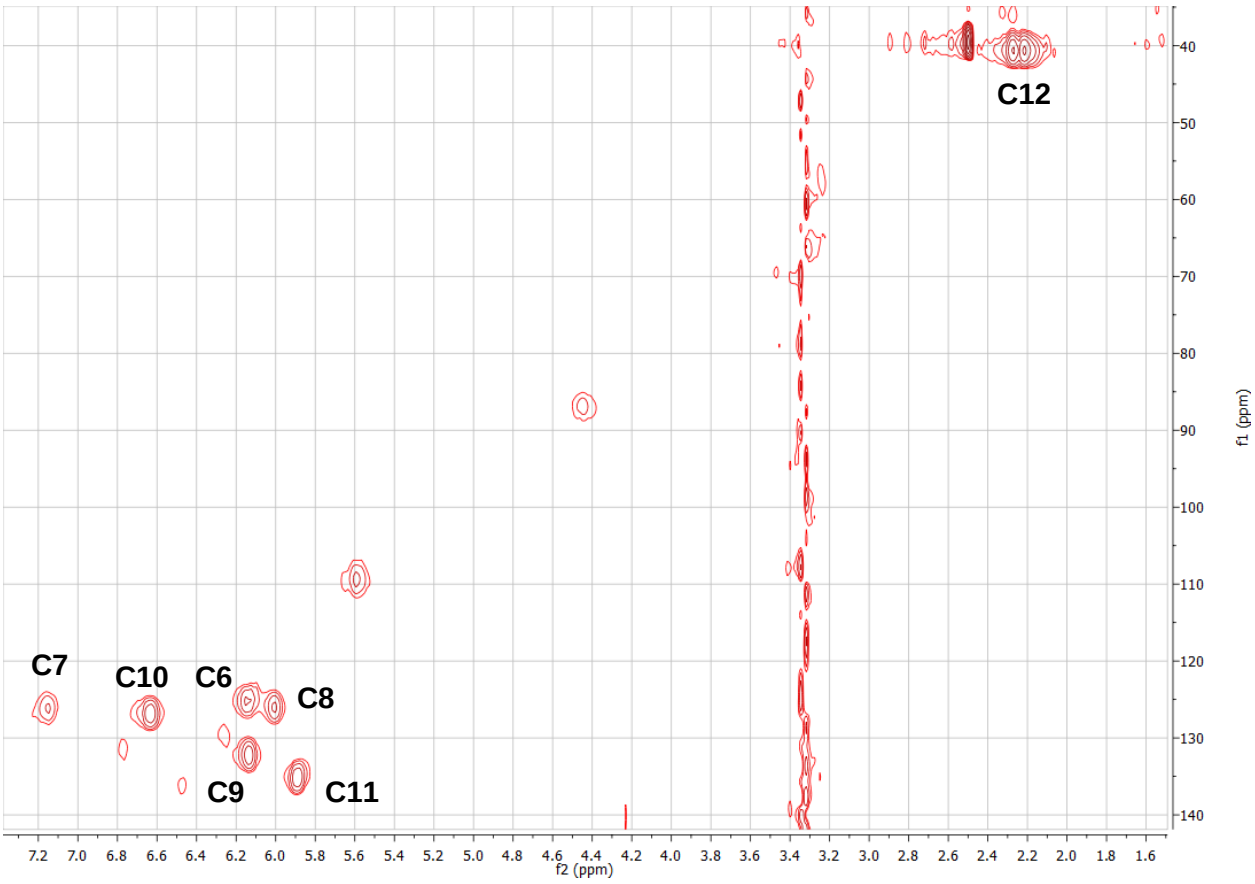



**Figure S13.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, HMQC NMR (800 MHz, $d_6$-DMSO).

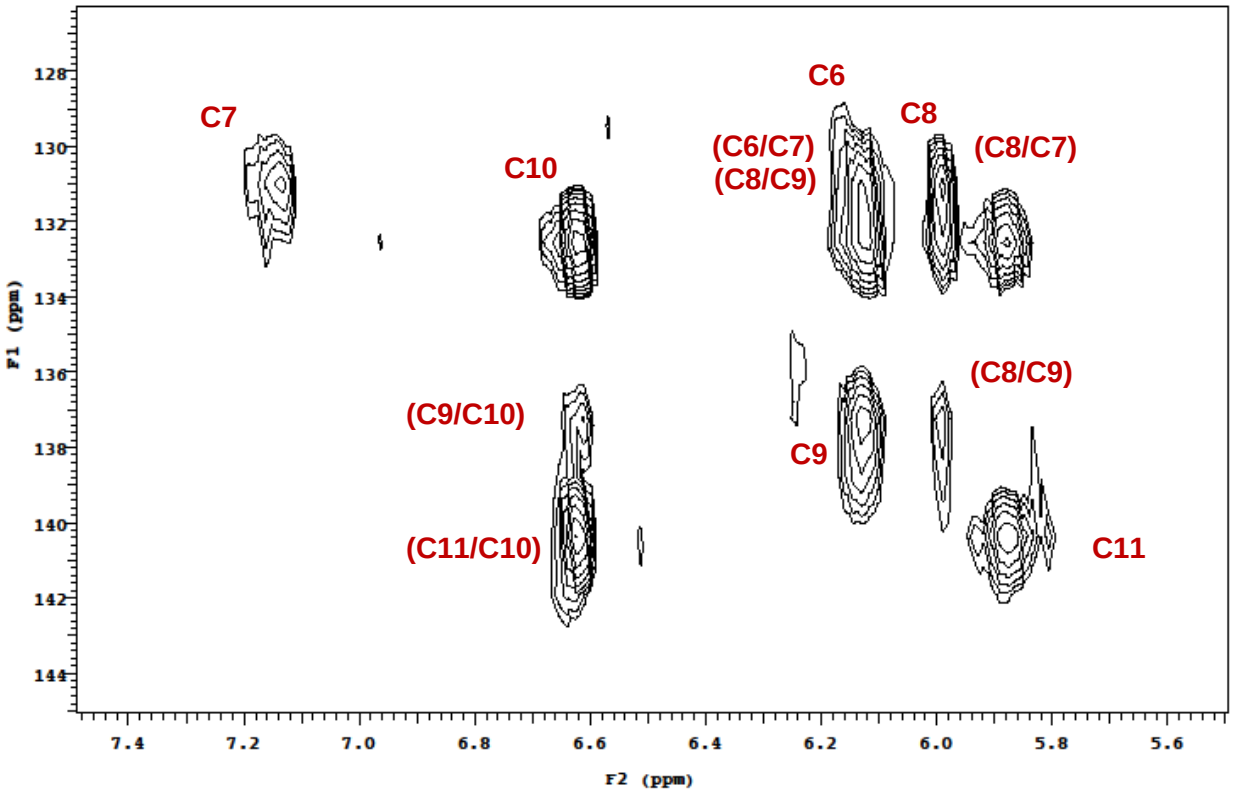



**Figure S14.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, HCCH-COSY NMR (800 MHz, $d_6$-DMSO), $^{1}$H-$^{13}$C plane.

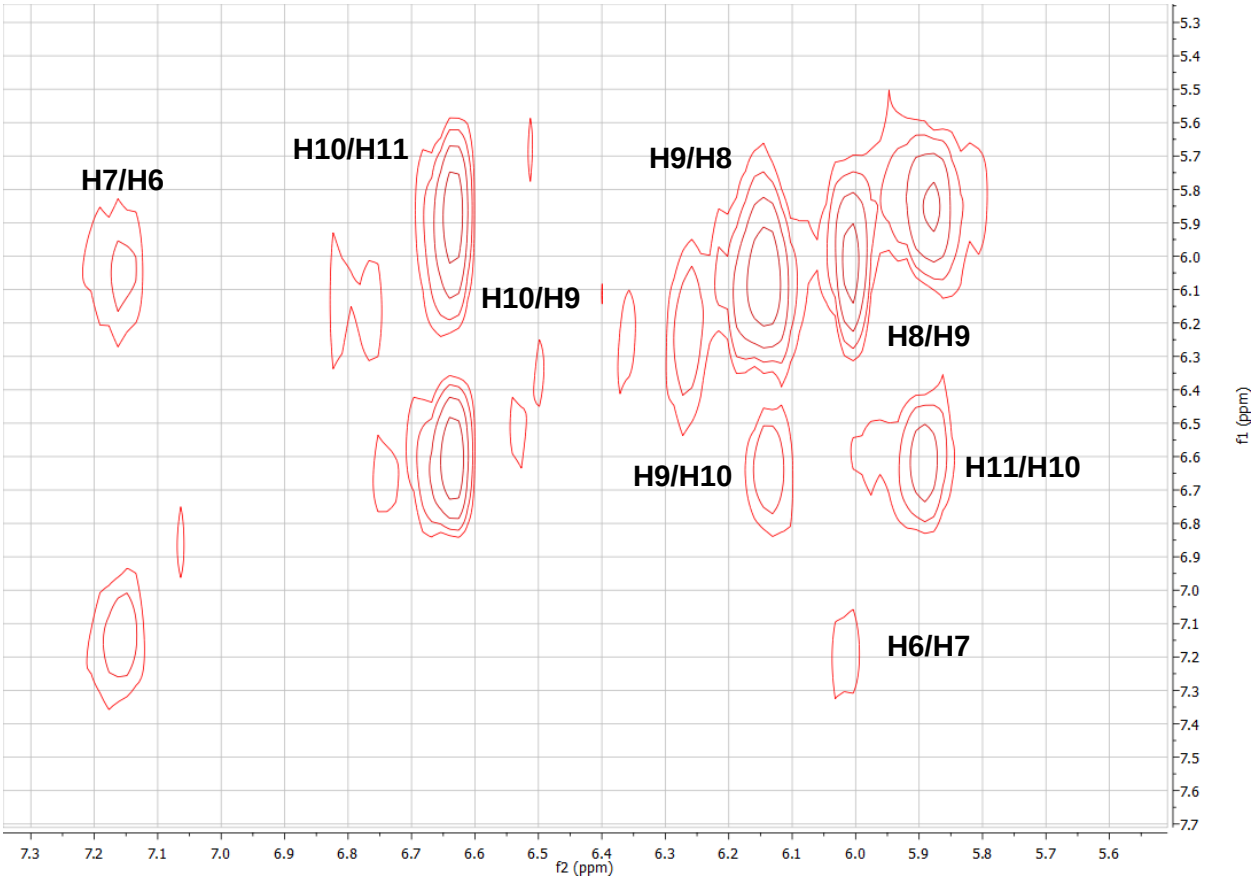



**Figure S15.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, HCCH-COSY NMR (800 MHz, *d*$_6$-DMSO), $^{1}$H-$^{1}$H plane.

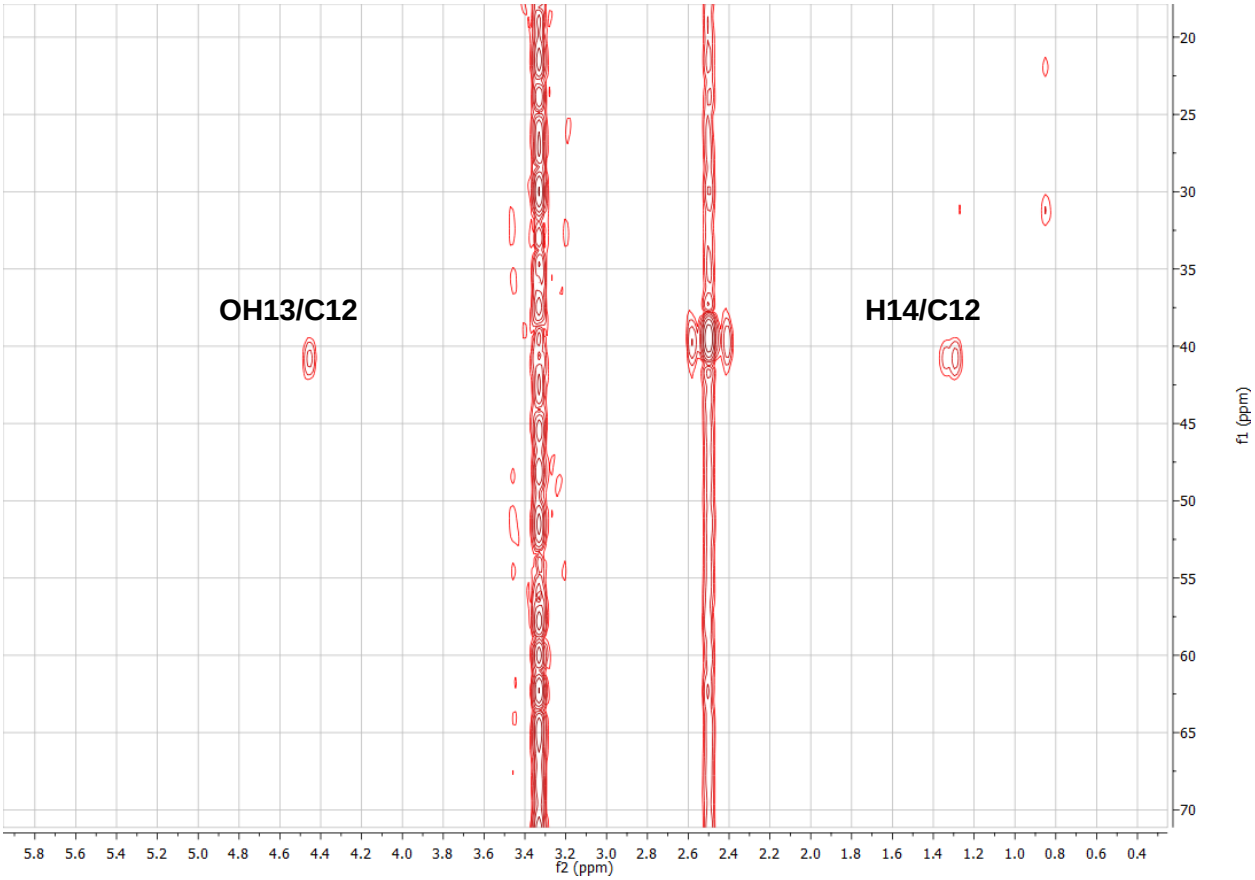



**Figure S16.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, HMBC NMR (800 MHz, *d*$_6$-DMSO).

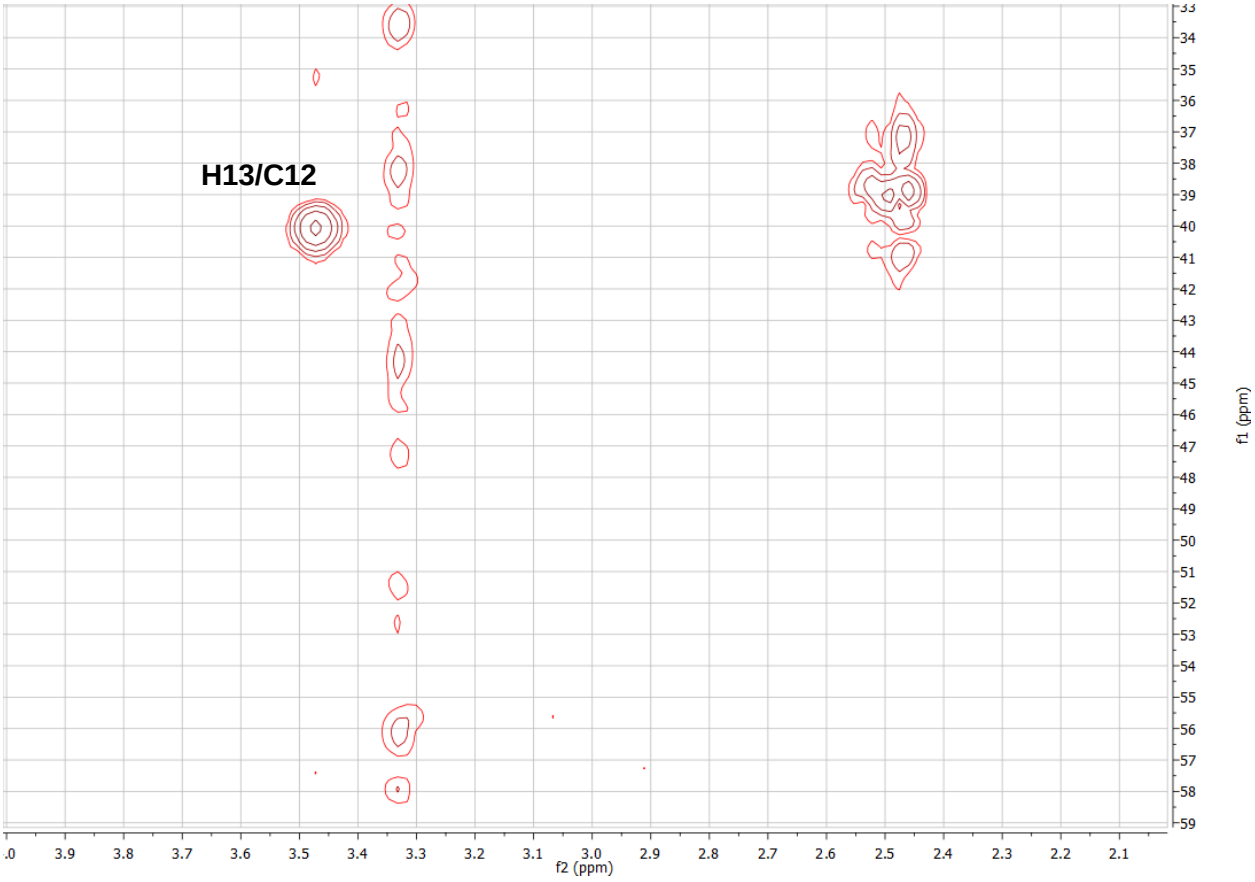



**Figure S17.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, H2BC NMR (800 MHz, *d*$_6$-DMSO).

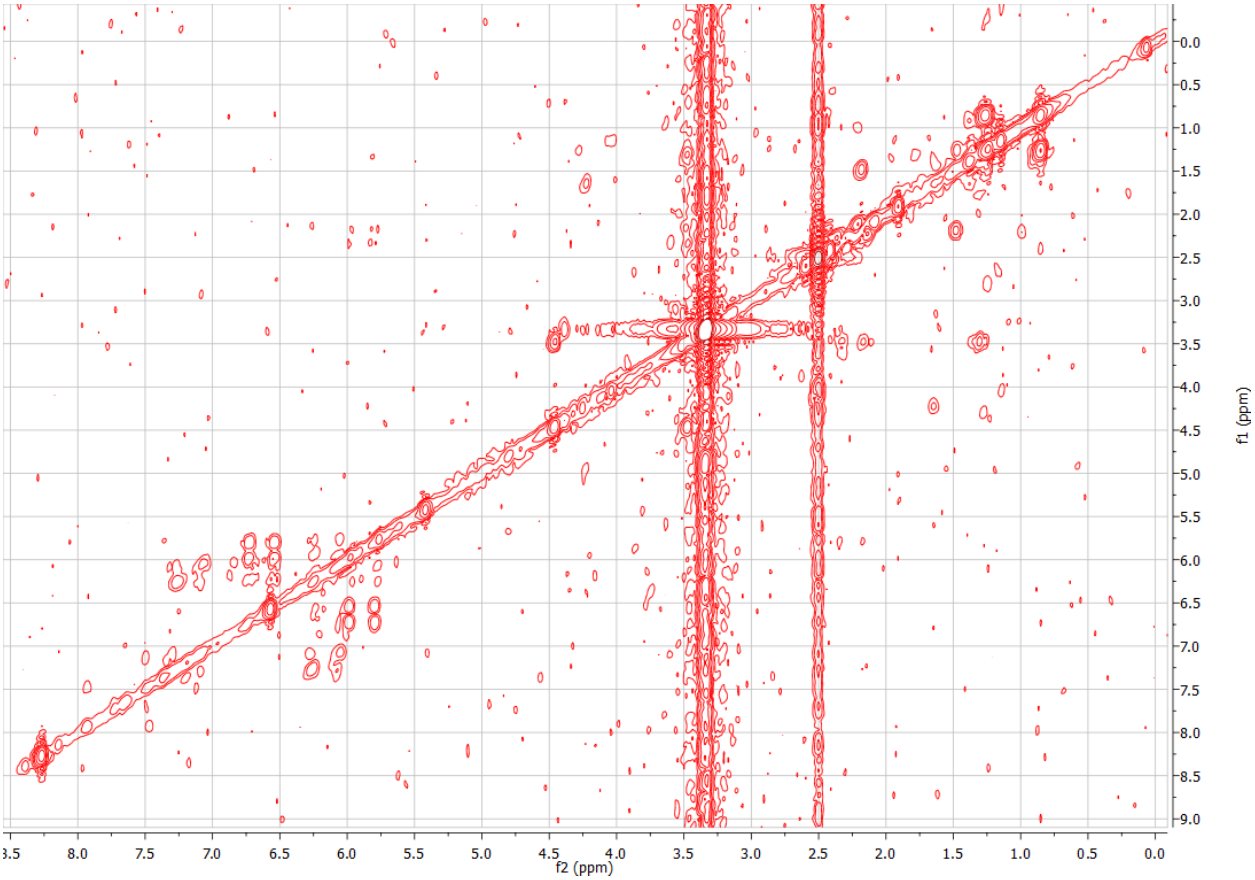



**Figure S18.** Desmethyl heptaketide **1**, uniformly [$^{13}$C]-labeled, $^{1}$H-$^{1}$H COSY NMR (800 MHz, $d_6$-DMSO).

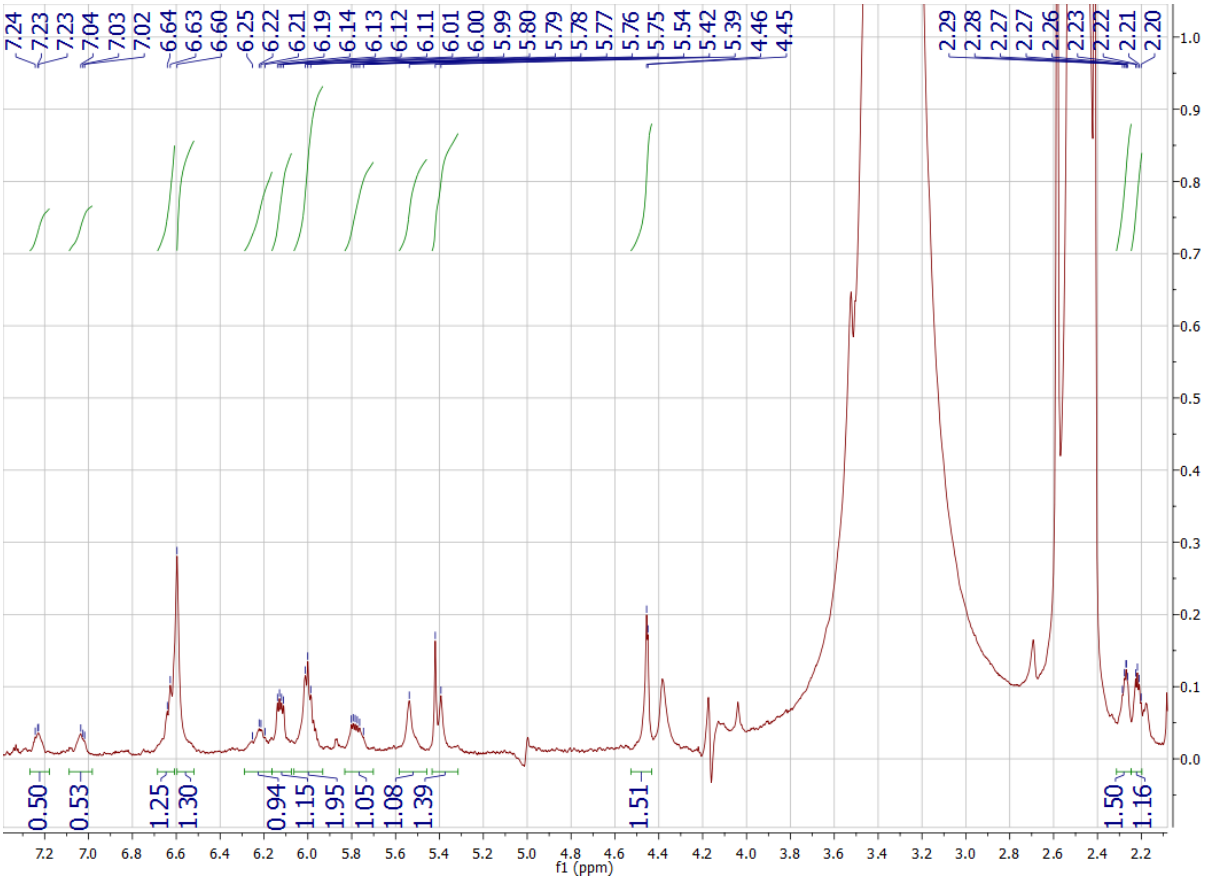



**Figure S19.** Desmethyl heptaketide **1**, [$^{13}C_2$]-1,3-malonate labeled, $^1$H NMR (800 MHz, $d_6$-DMSO).

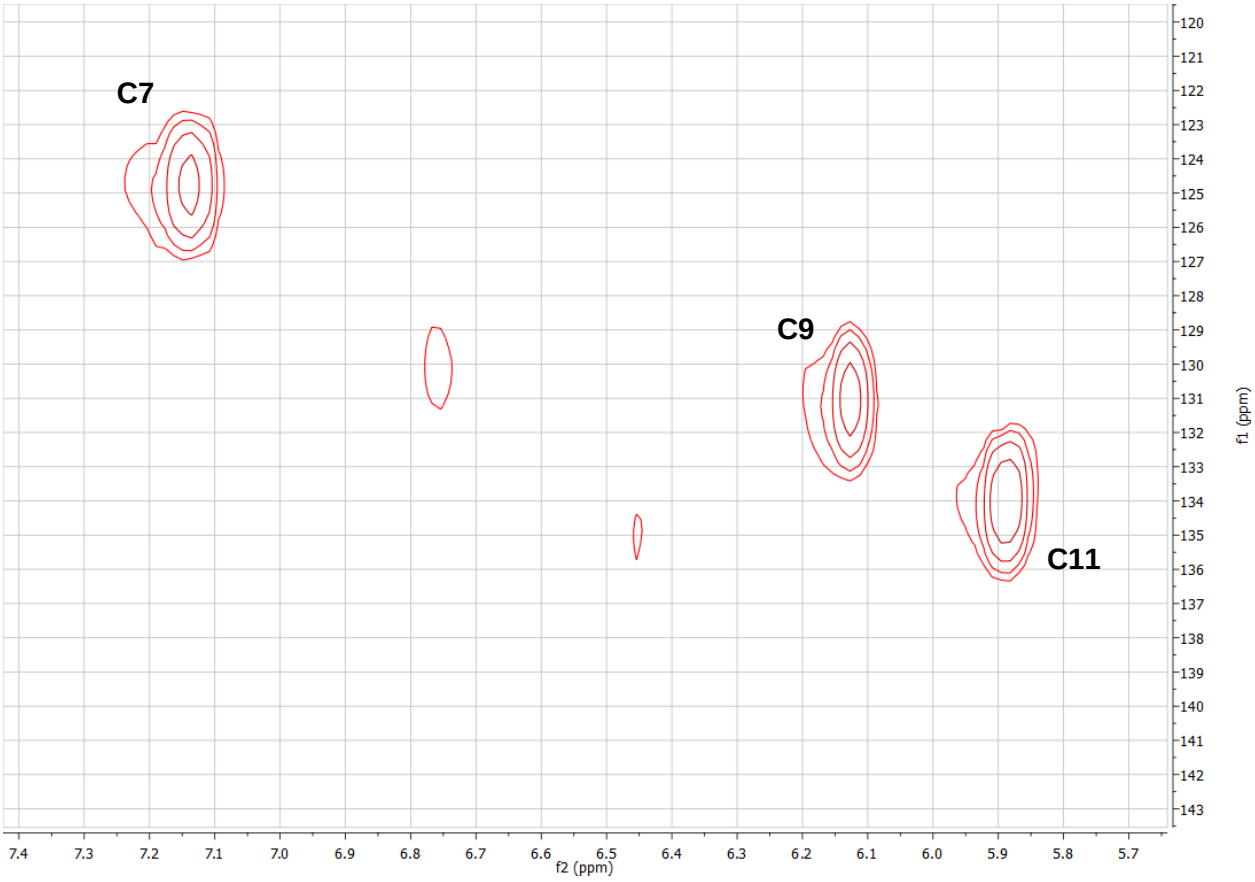



**Figure S20.** Desmethyl heptaketide **1**, [1,3-$^{13}C_2$]-malonate labeled, HMQC NMR (800 MHz, *d*$_6$-DMSO).

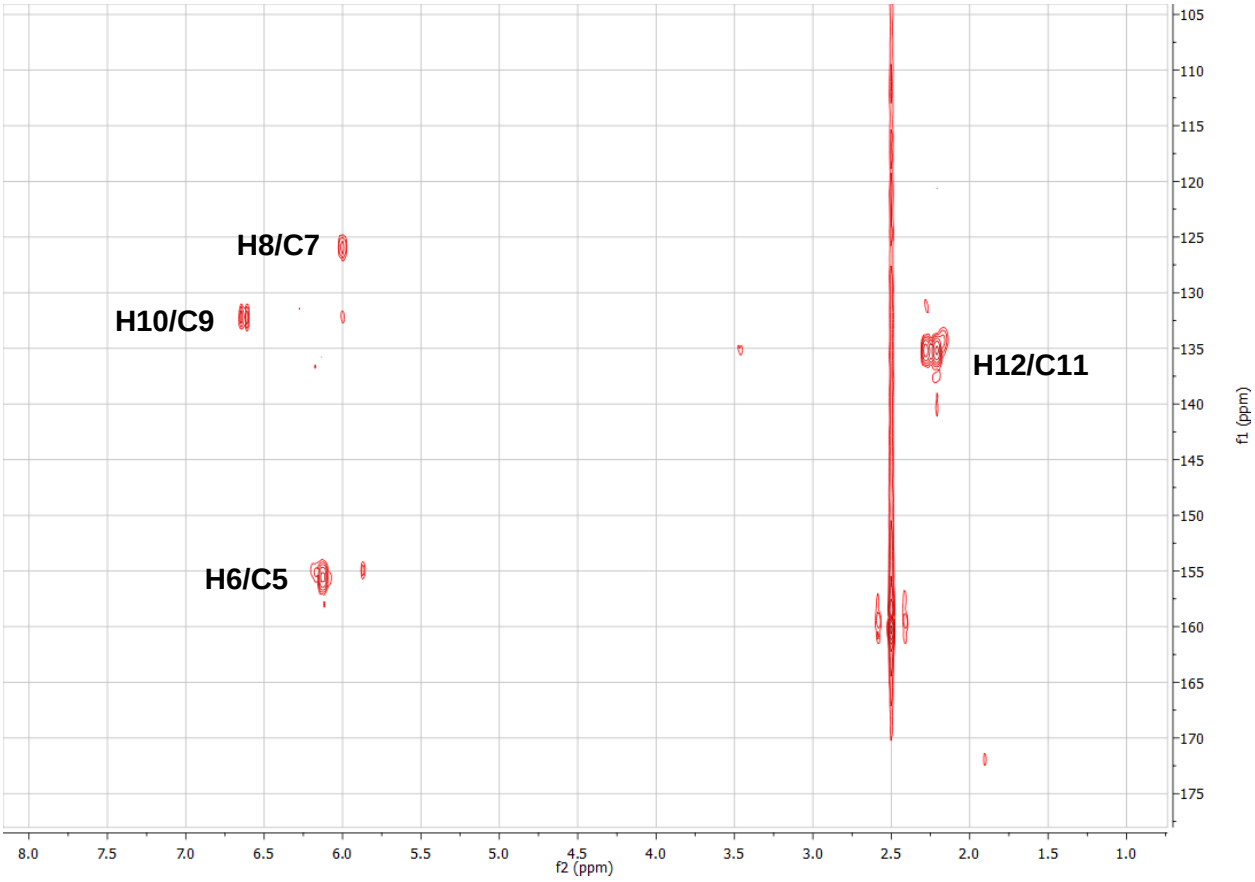



**Figure S21.** Desmethyl heptaketide **1**, [1,3-$^{13}C_2$]-malonate labeled, HMBC NMR (800 MHz, *d*$_6$-DMSO).

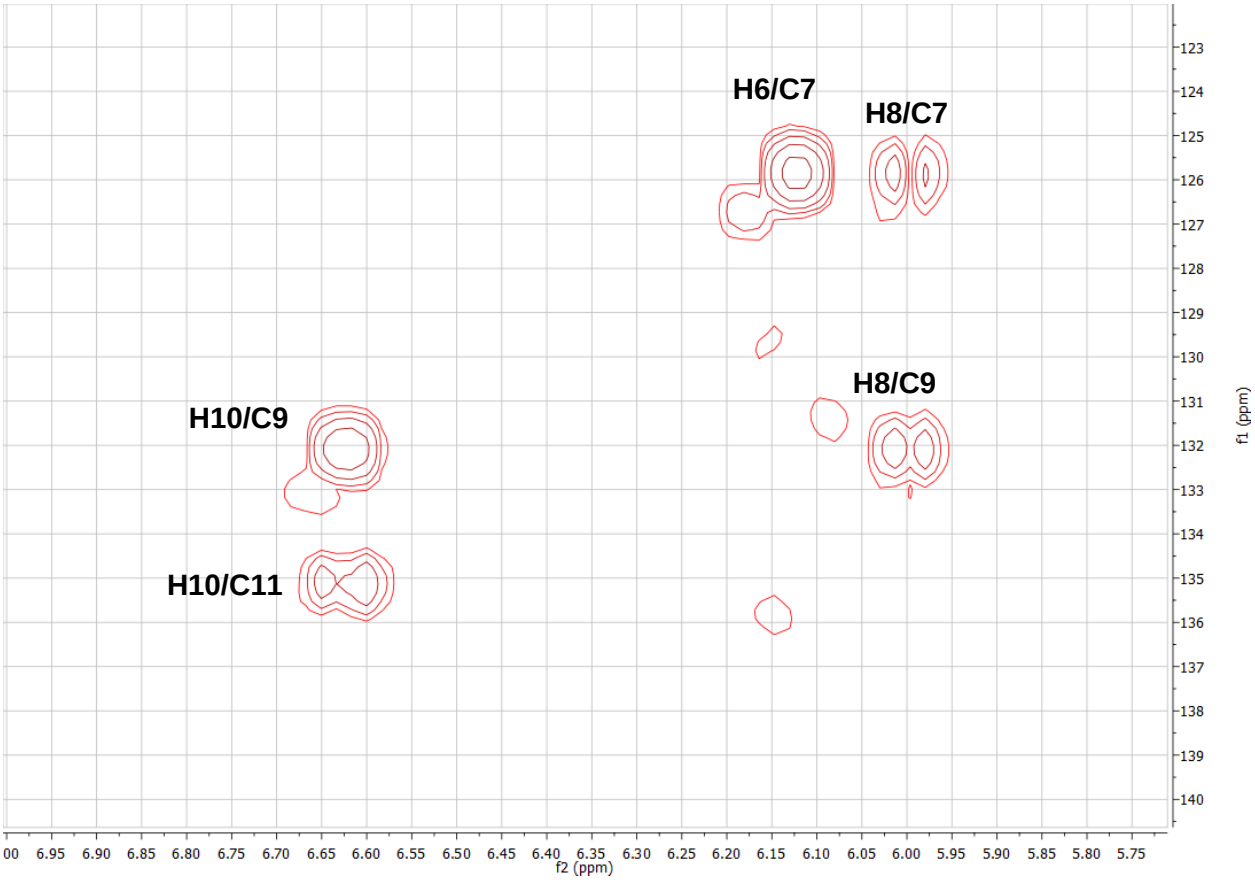



**Figure S22.** Desmethyl heptaketide **1**, [1,3-$^{13}C_2$]-malonate labeled, H2BC NMR (800 MHz, *d*$_6$-DMSO).

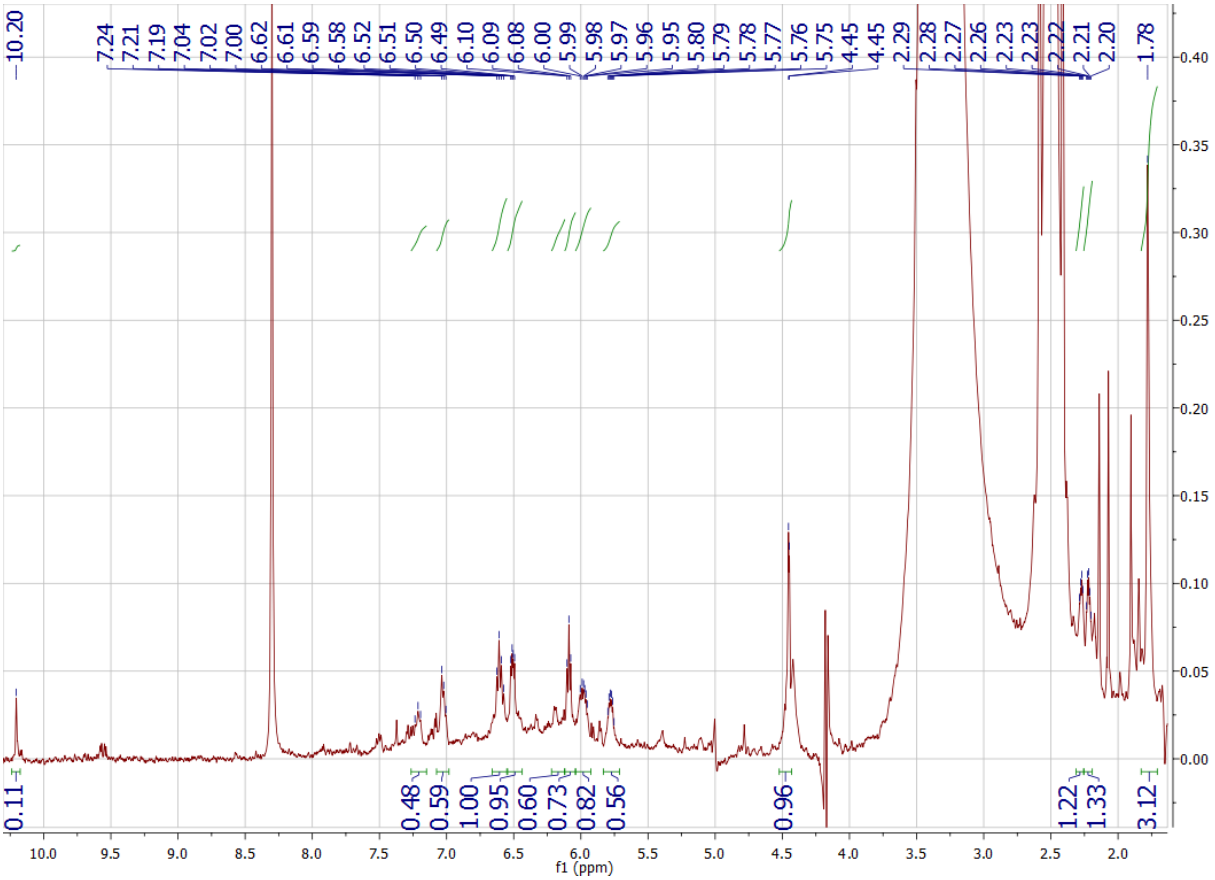



**Figure S23.** Methyl heptaketide **2**, [1,3-$^{13}C_2$]-malonate labeled, $^1$H NMR (800 MHz, *d*$_6$-DMSO).

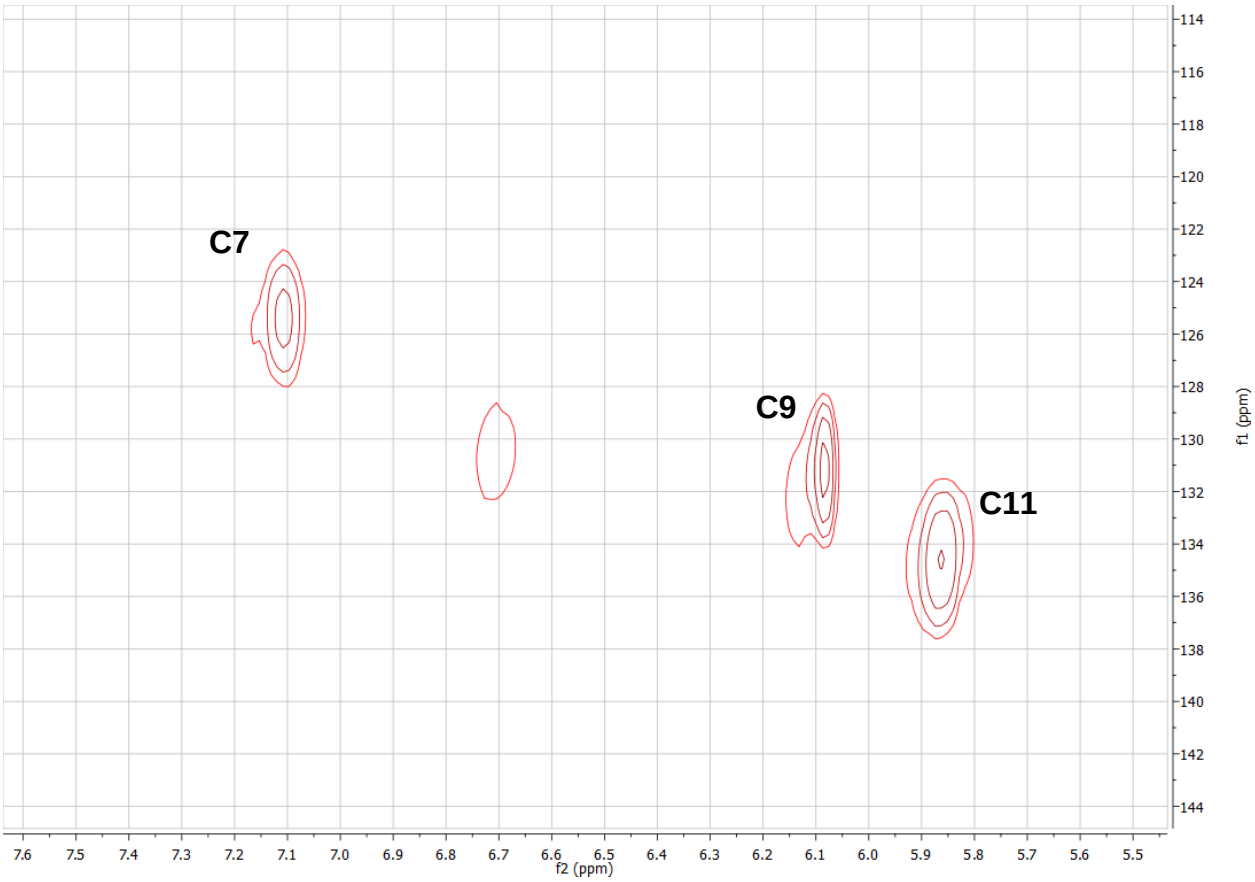



**Figure S24.** Methyl heptaketide **2**, [1,3-$^{13}C_2$]-malonate labeled, HMQC NMR (800 MHz, $d_6$-DMSO).

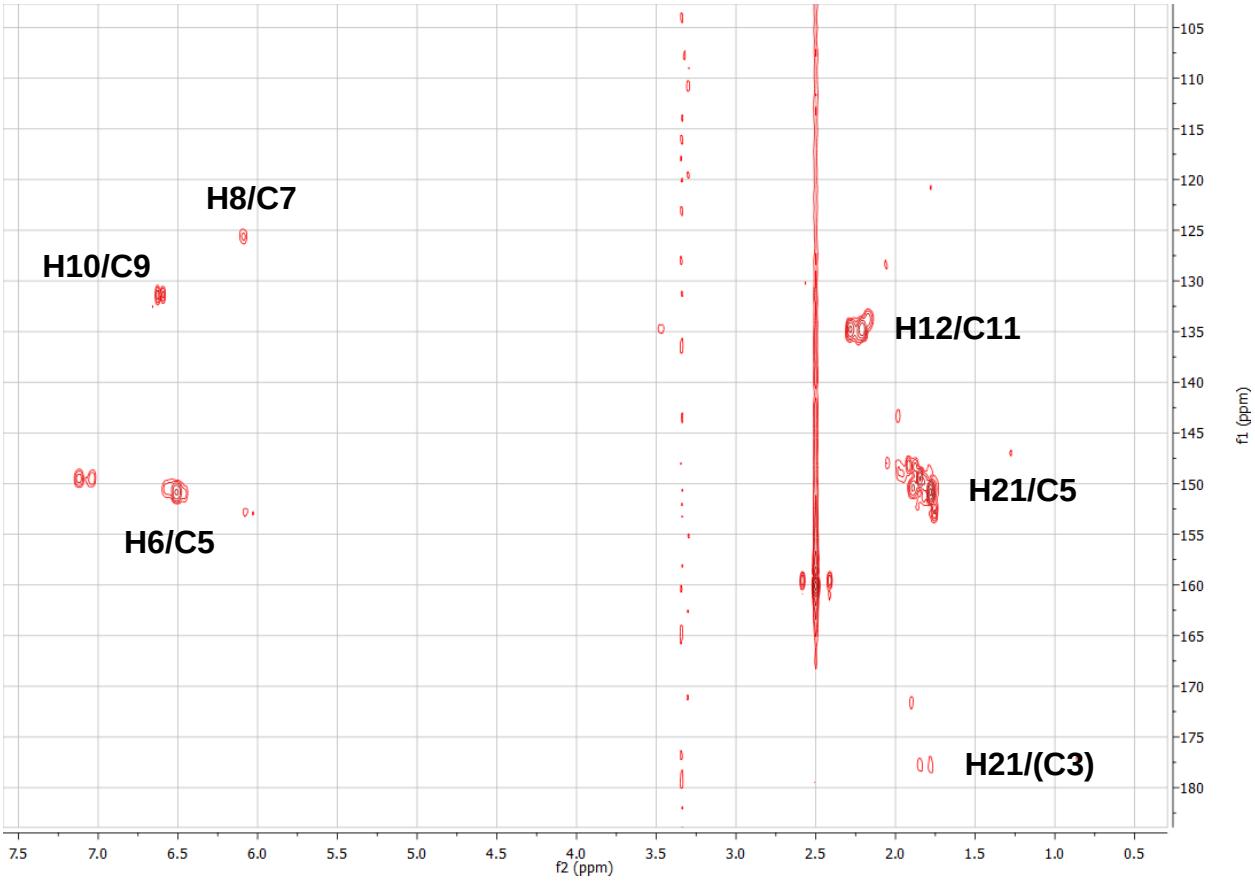



**Figure S25.** Methyl heptaketide **2**, [1,3-$^{13}C_2$]-malonate labeled, HMBC NMR (800 MHz, $d_6$-DMSO).

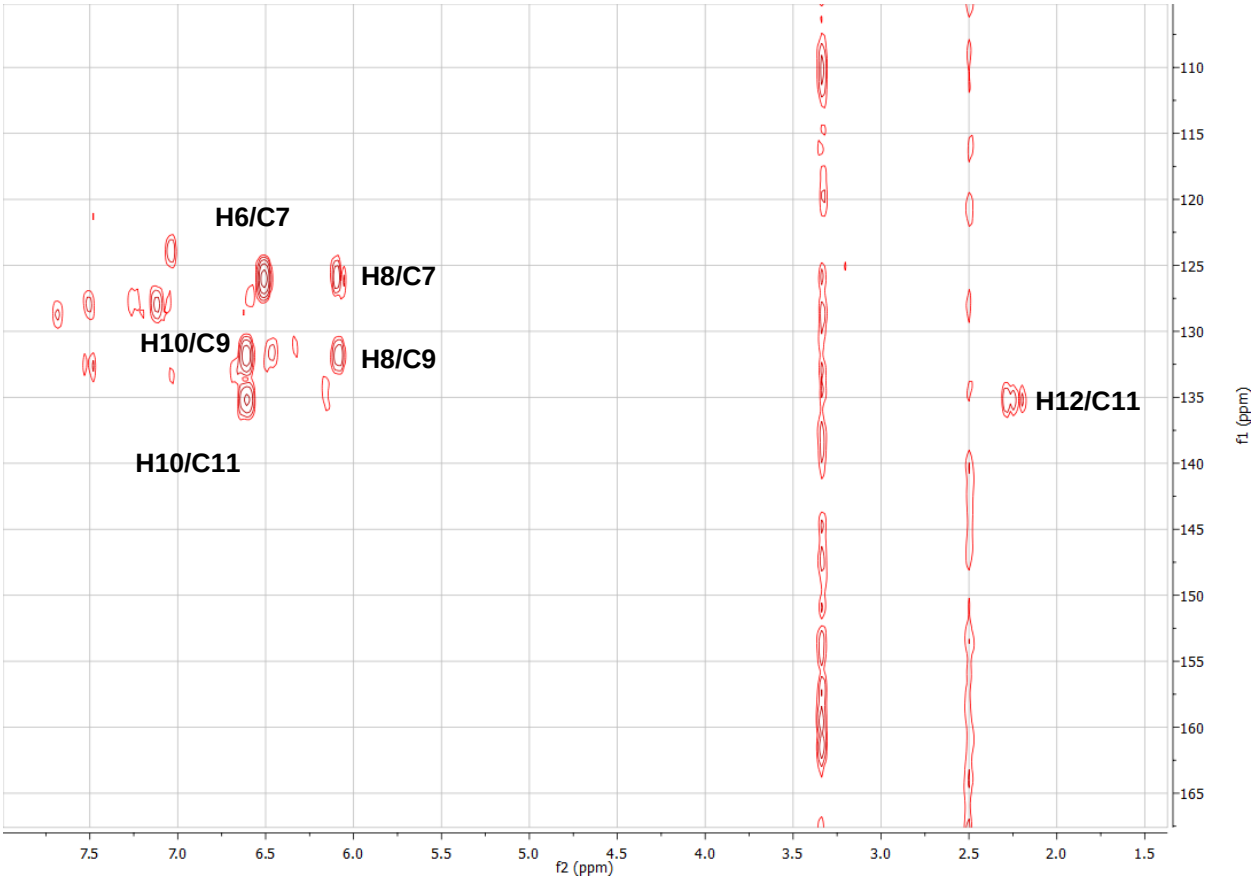



**Figure S26.** Methyl heptaketide **2**, [1,3-$^{13}C_2$]-malonate labeled, HMQC NMR (800 MHz, $d_6$-DMSO).

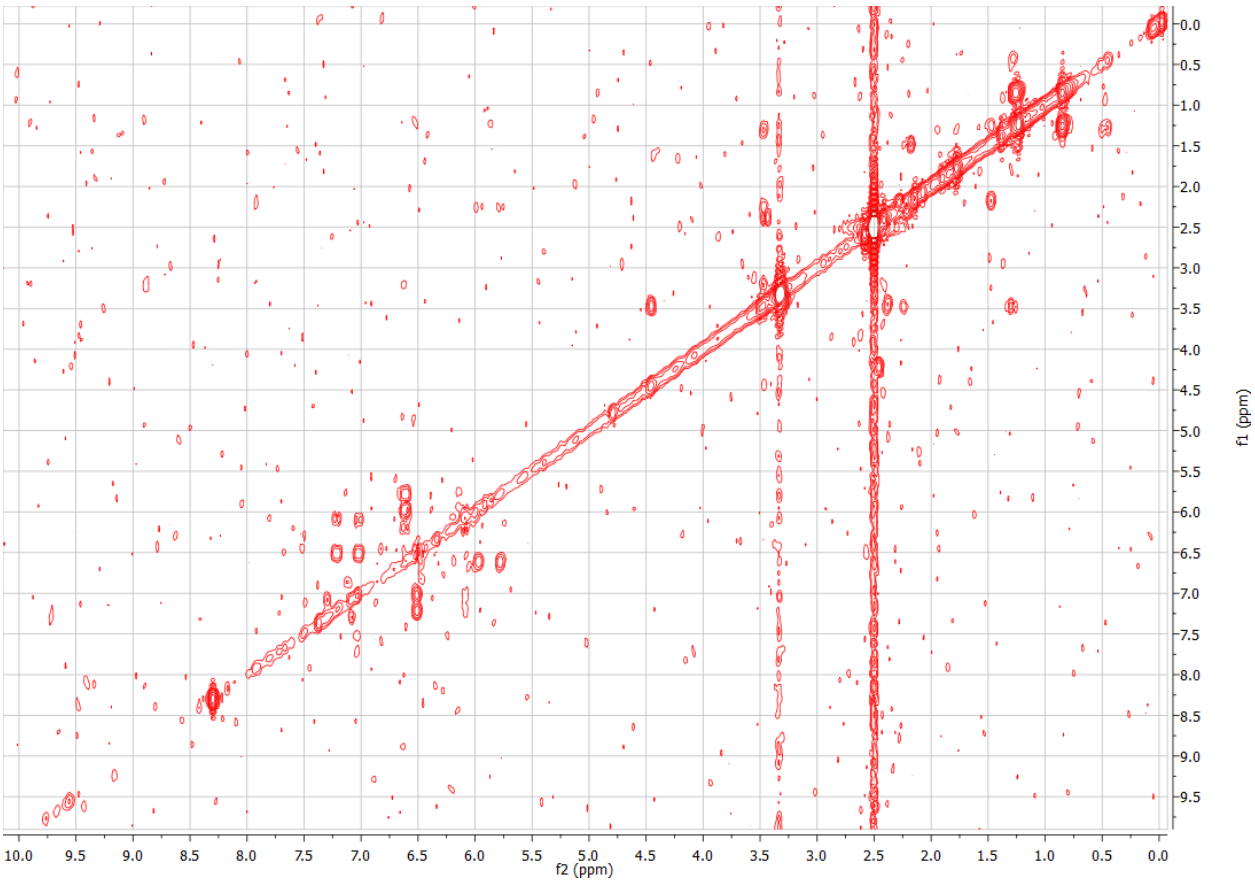



**Figure S27.** Methyl heptaketide **2**, [1,3-$^{13}C_2$]-malonate labeled, $^{1}$H-$^{1}$H COSY NMR (800 MHz, $d_6$-DMSO).

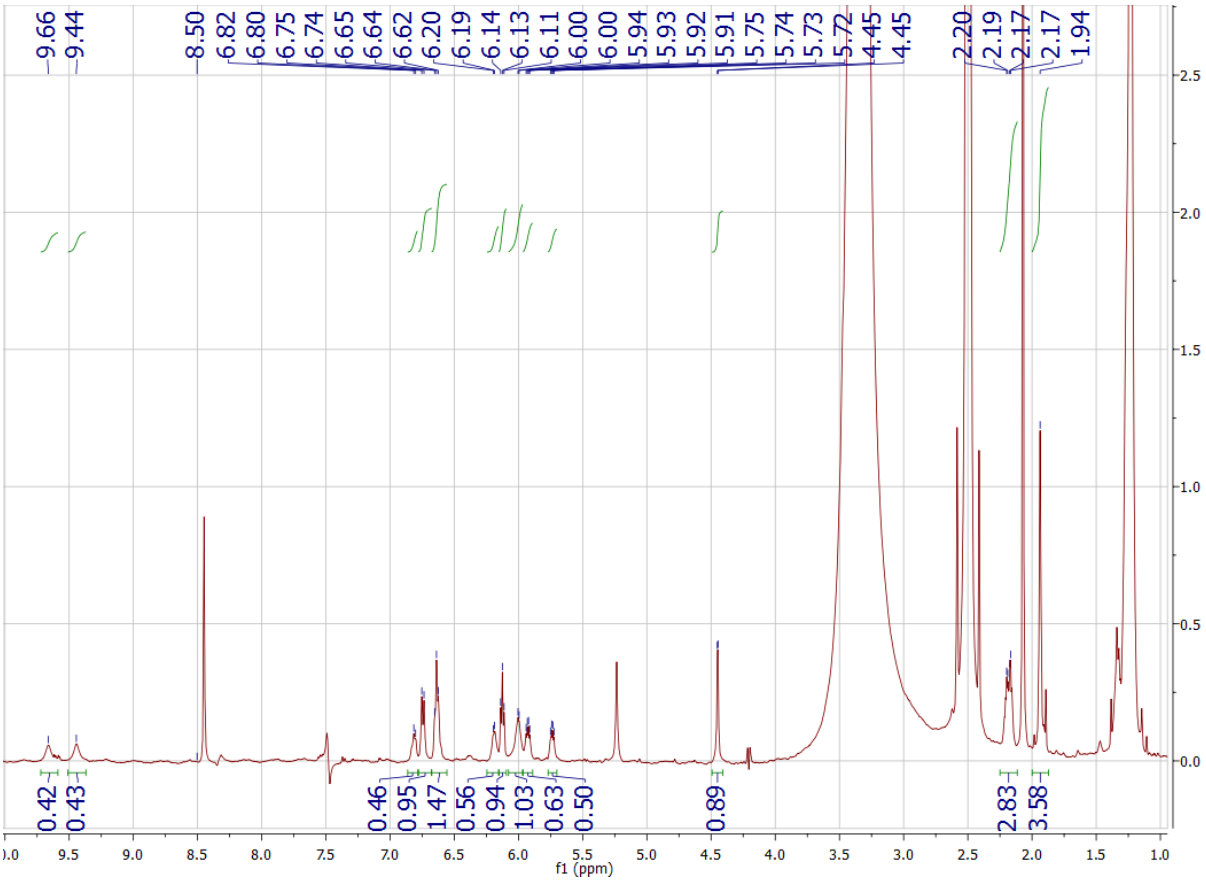



**Figure S28.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, $^1$H NMR (800 MHz, *d*$_6$-DMSO).

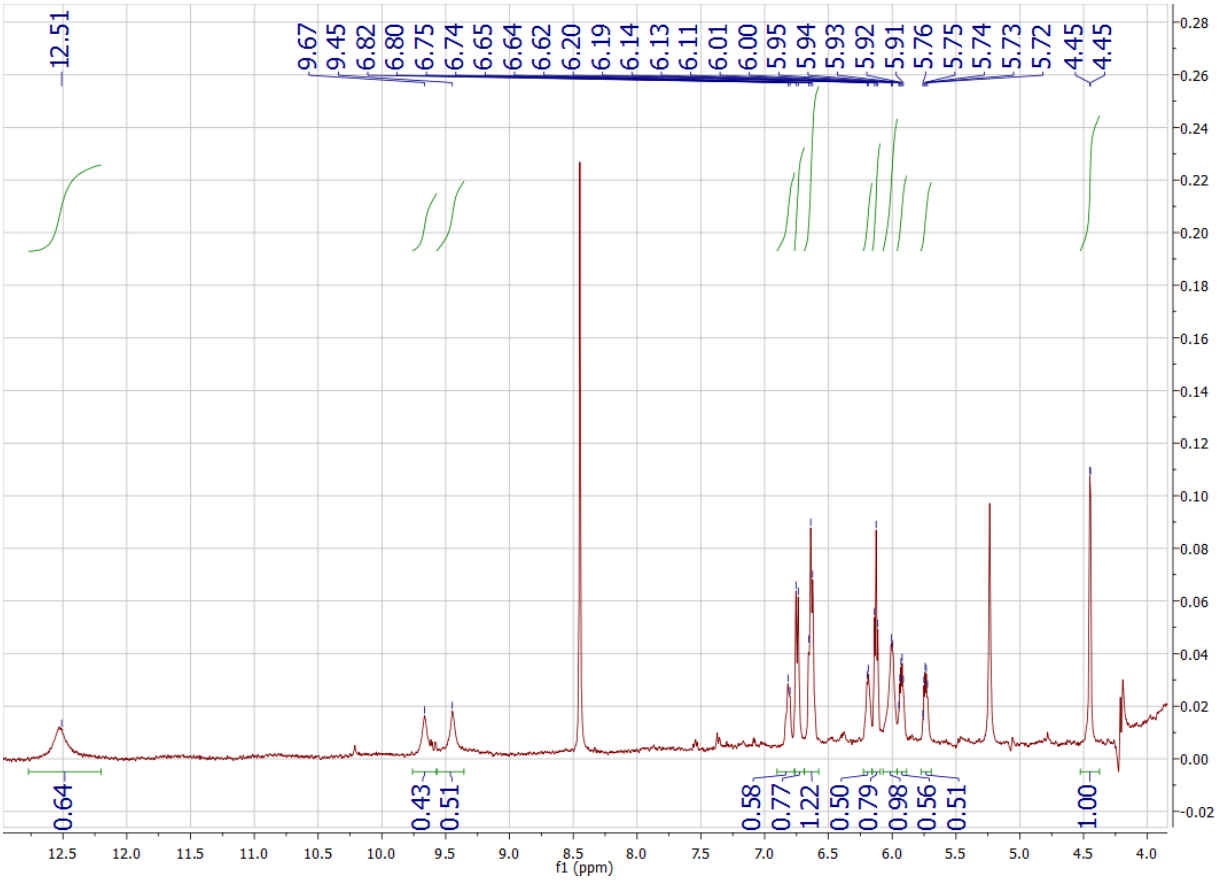



**Figure S29.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, $^{1}$H NMR wide spectral width (800 MHz, *d*$_6$-DMSO).

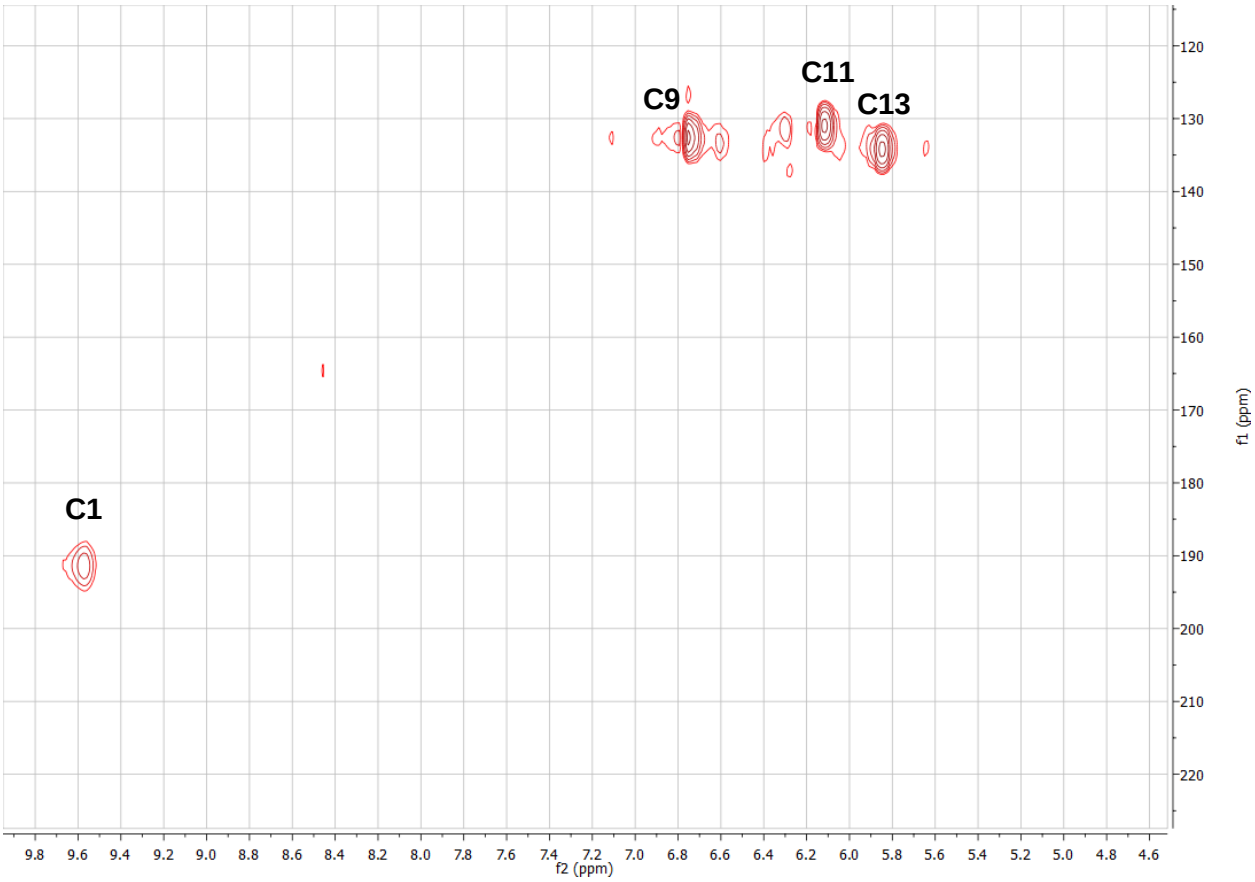



**Figure S30.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, HMQC NMR (800 MHz, $d_6$-DMSO).

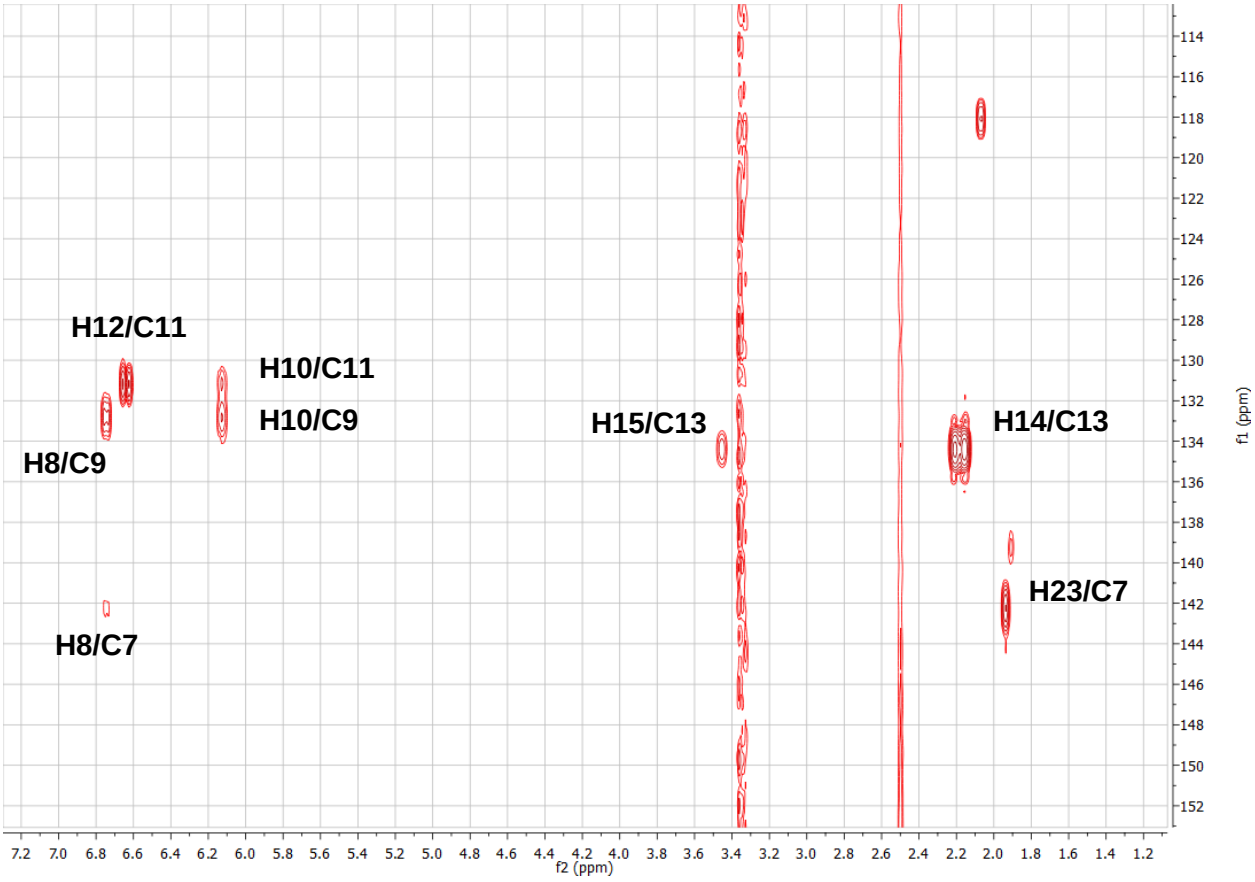



**Figure S31.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, HMBC NMR (800 MHz, *d*$_6$-DMSO).

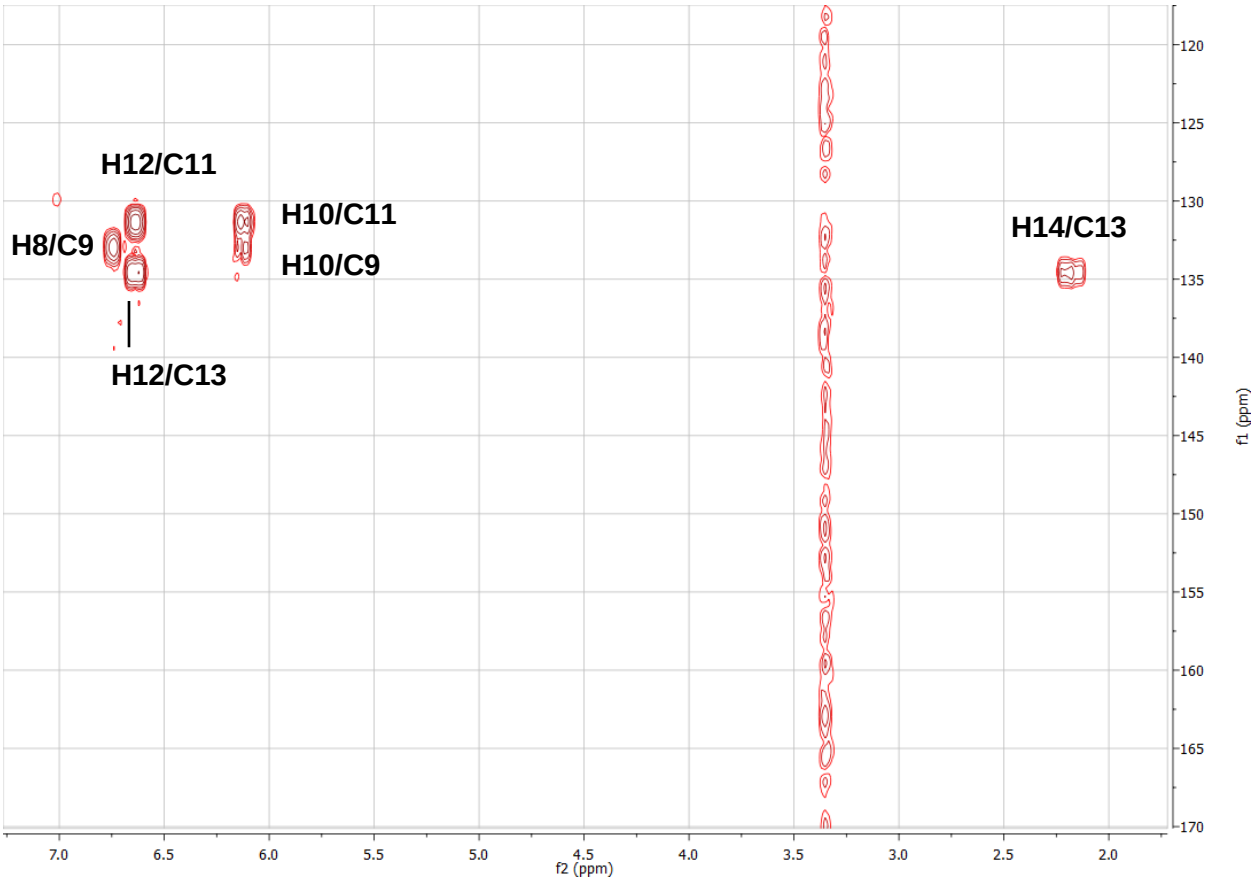



**Figure S32.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, H2BC NMR (800 MHz, $d_6$-DMSO).

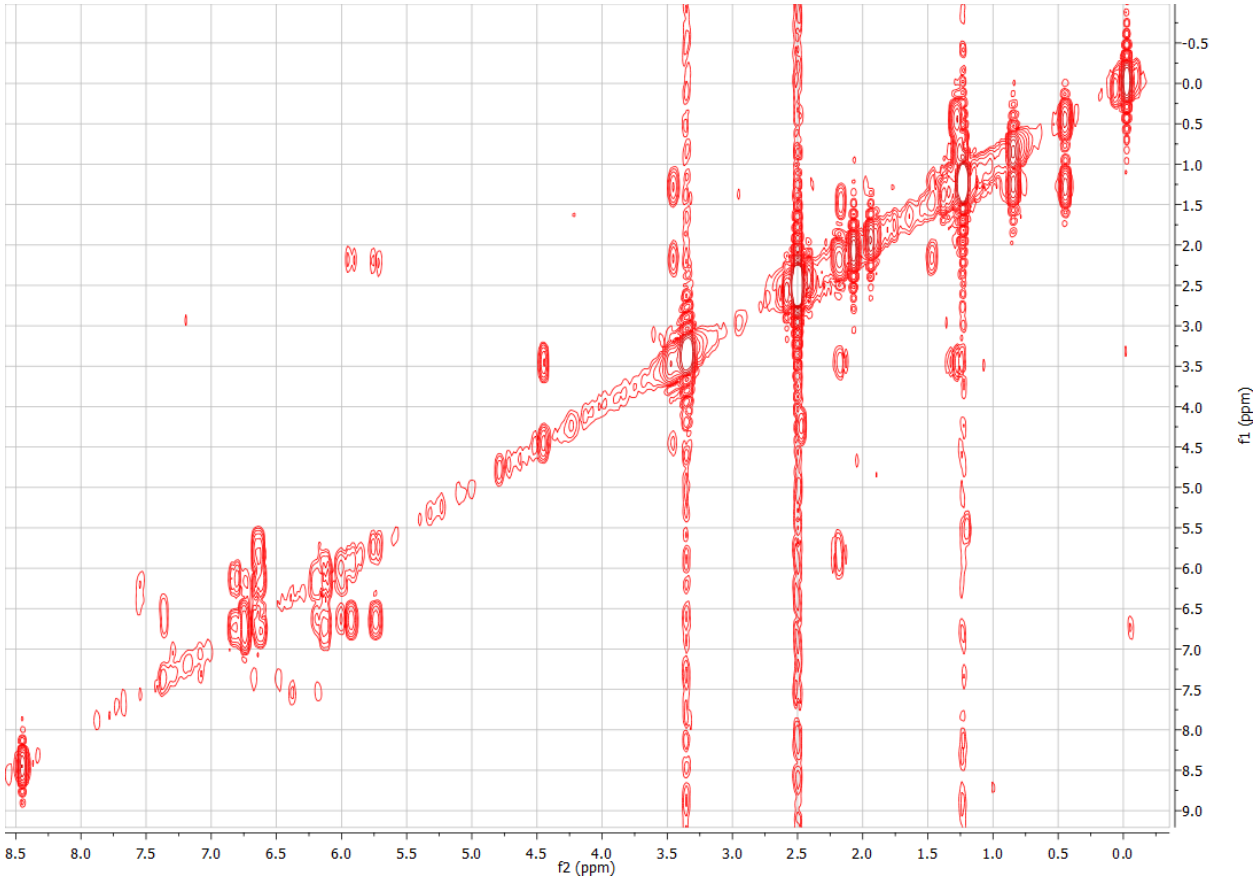



**Figure S33.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, $^{1}$H-$^{1}$H COSY NMR (800 MHz, $d_6$-DMSO).

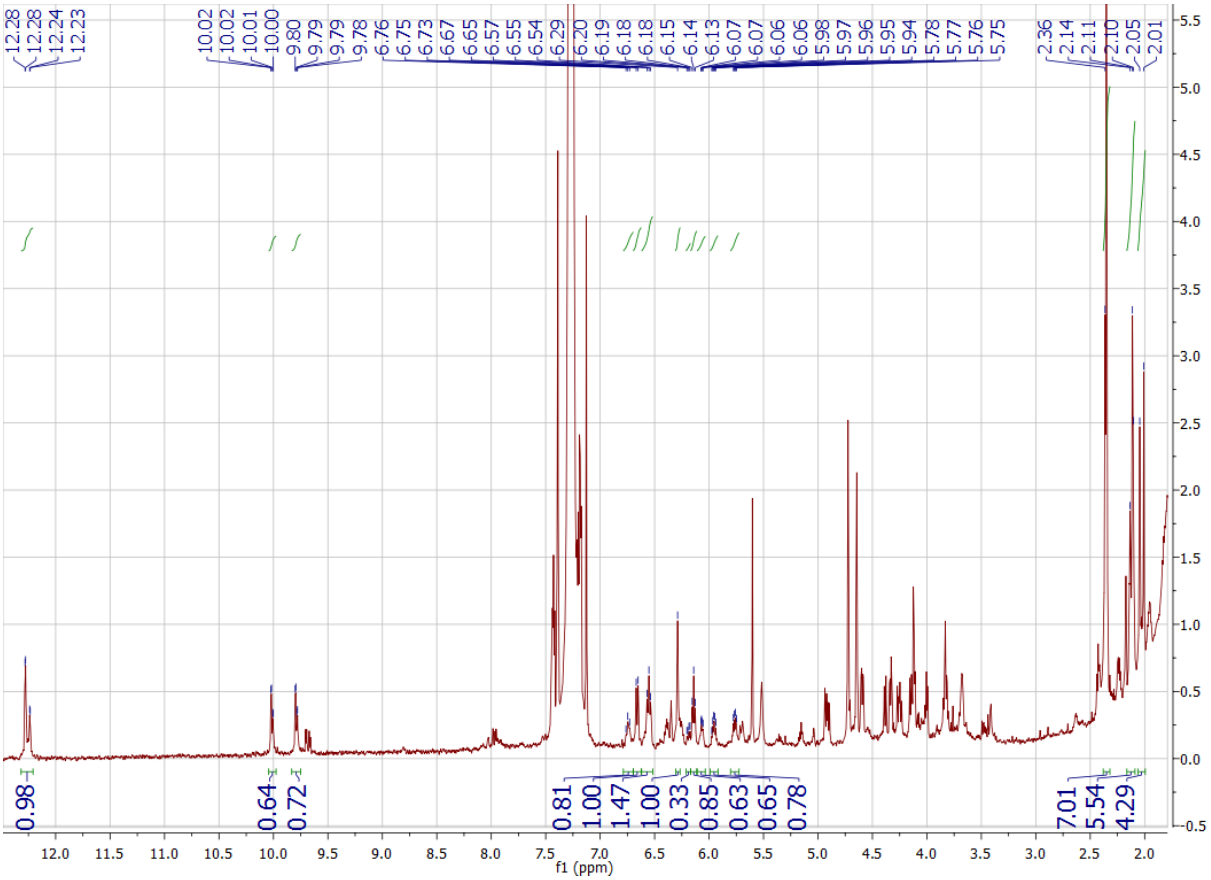



**Figure S34.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, $^{1}$H NMR (800 MHz, $CDCl_3$). Chloroform-dissolved plastic-related contaminants were abundant in the 1D, but did not have comparable signal in 2D experiments because of the $^{13}$C-labeling of the compound.


**Figure S35.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, HMQC NMR (800 MHz, $CDCl_3$).

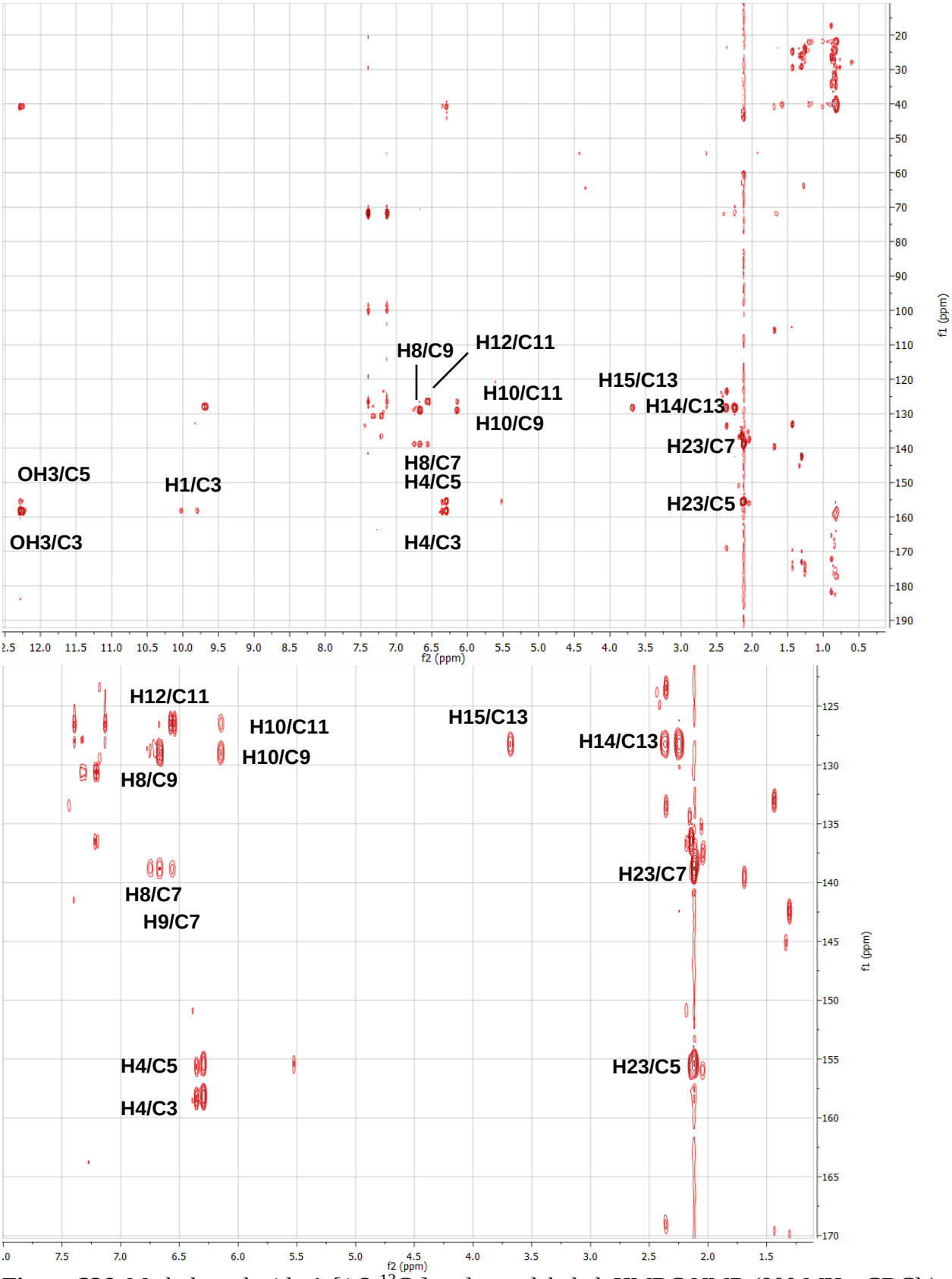



**Figure S36.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, HMBC NMR (800 MHz, $CDCl_3$).

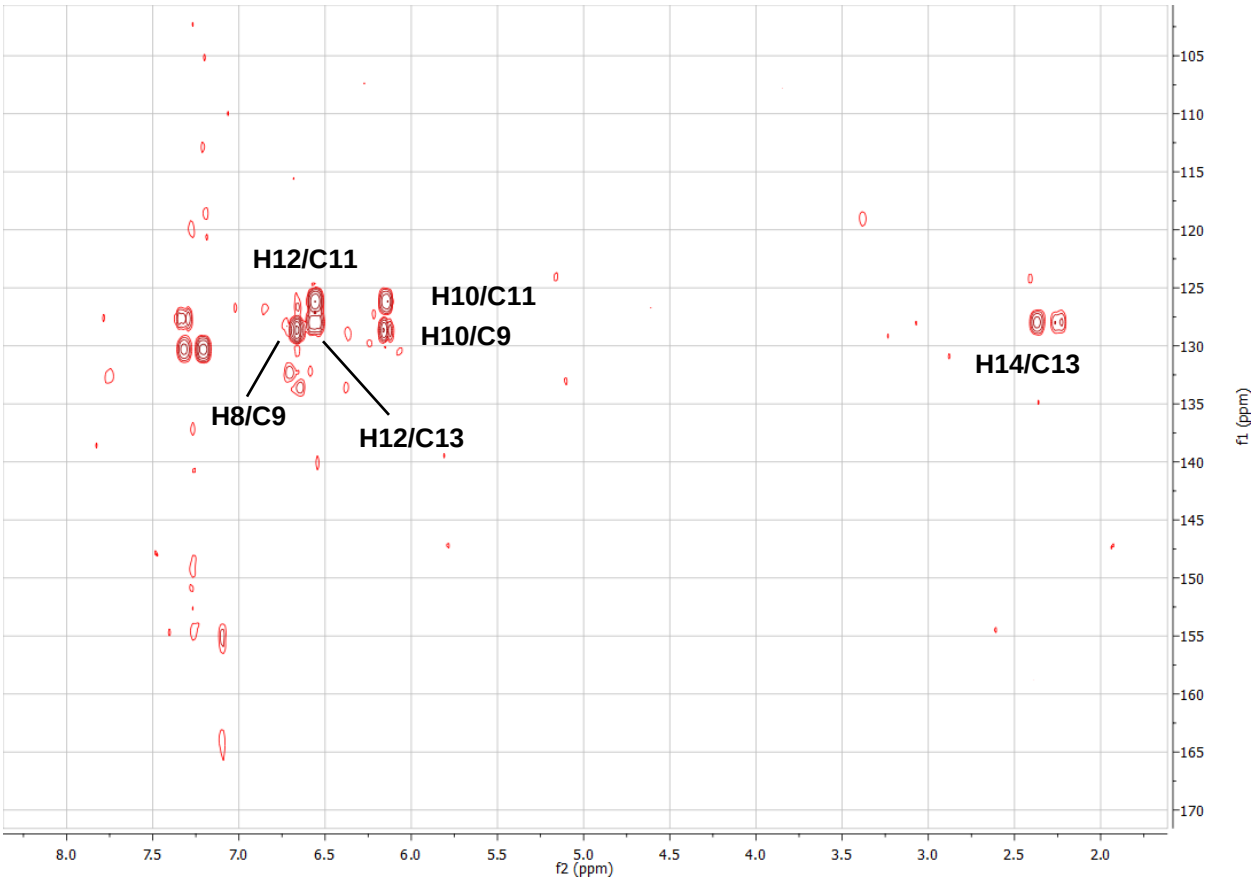



**Figure S37.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, H2BC NMR (800 MHz, $CDCl_3$).

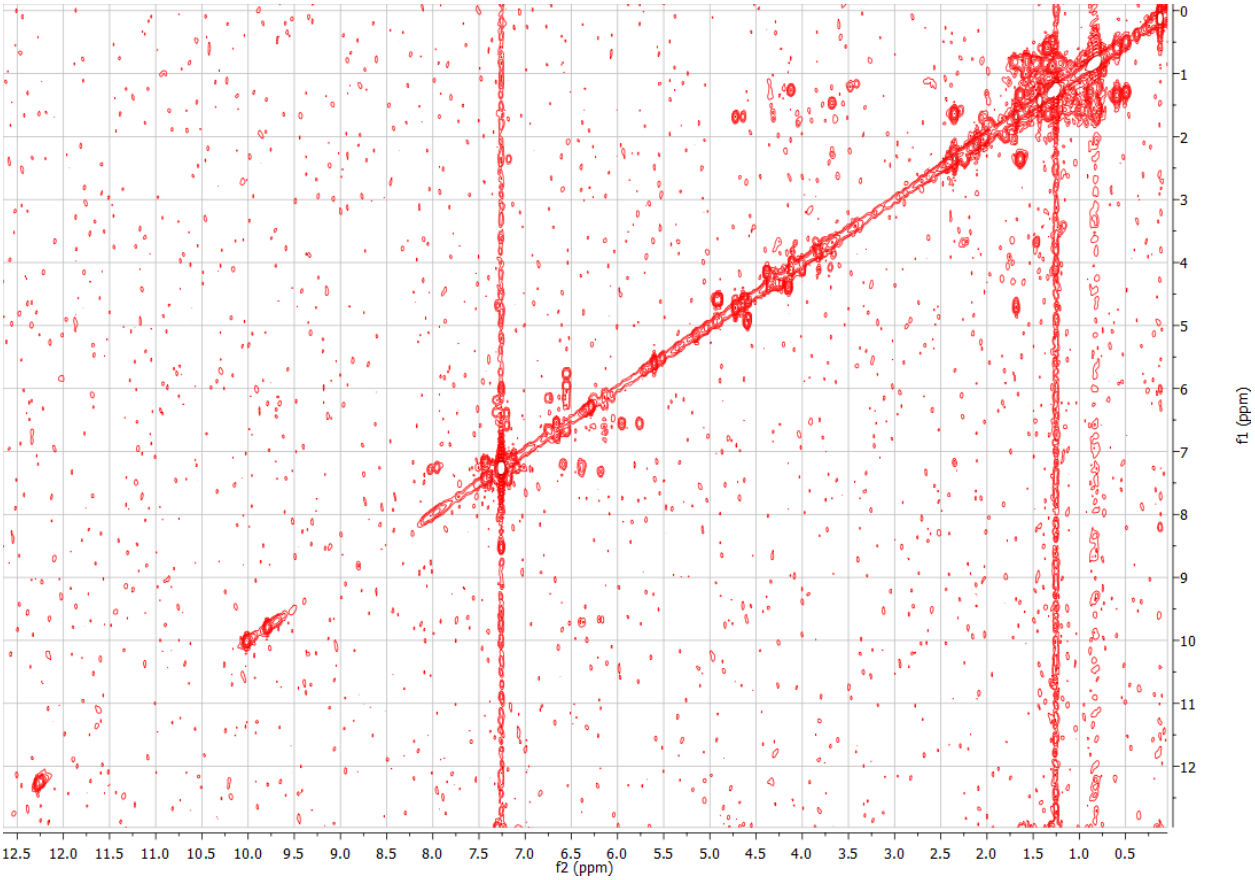



**Figure S38.** Methyl octaketide **4**, [1,3-$^{13}C_2$]-malonate labeled, $^{1}H$-$^{1}H$ COSY NMR (800 MHz, $CDCl_3$).

## MS/MS Spectra and Analysis

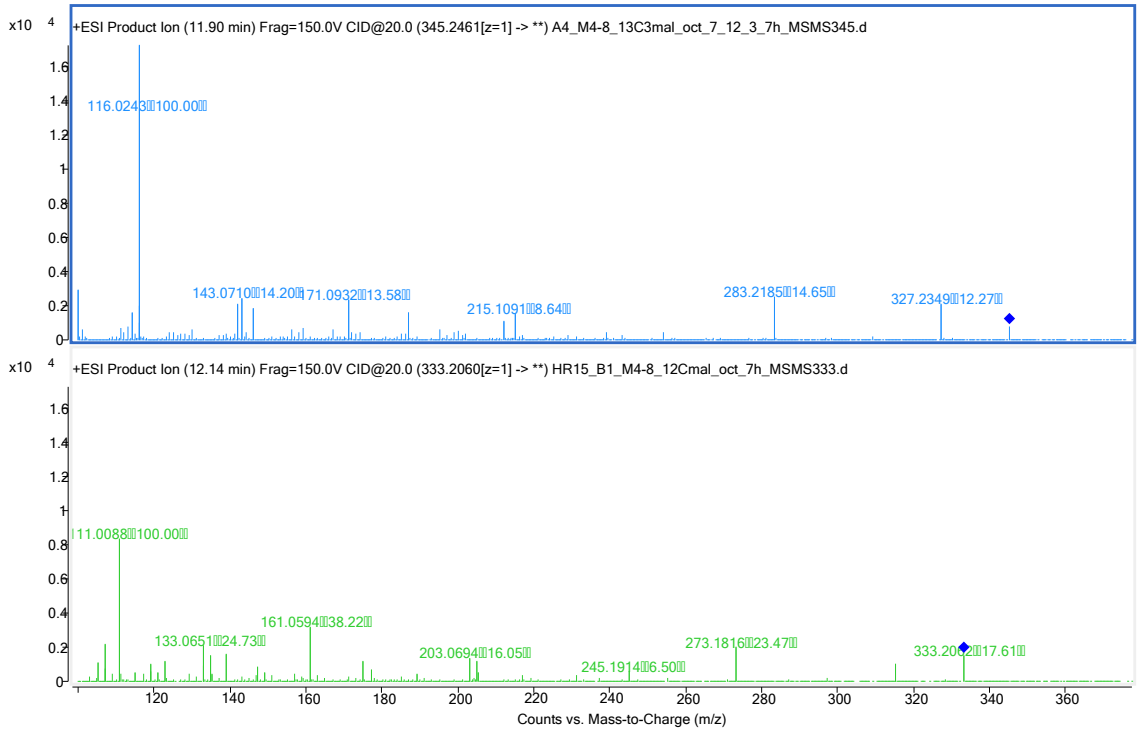



**Figure S39. MS/MS fragmentation analysis of desmethyl heptaketide 1**. Spectra (20V fragmentation) were obtained with uniformly $^{13}$C-labeled (top) and unlabeled (bottom) samples.

**A**

| m/z (obsd.) | $^{12}$C m/z (calcd.) | # of [$^{13}$C] atoms | $^{13}$C m/z (calcd.) | Formula | Δm/z (ppm) | Putative structure |
|---|---|---|---|---|---|---|
| 333.2062 | 333.2060 | 12 | 345.2463 | C20H29O4+ | 0.6 | a |
| 315.1946 | 315.1955 | 12 | 327.2358 | C20H27O3+ | -2.9 | b |
| 273.1850 | 273.1849 | 12 | 285.2252 | C18H25O2+ | 0.4 | c |
| 245.1907 | 245.1900 | 9 | 254.2202 | C17H25O+ | 2.9 | d |
| 205.1946 | 205.1951 | 7 | 212.2186 | C15H25+ | -2.4 | e |
| 203.0691 | 203.0703 | 12 | 215.1106 | C12H11O3+ | -5.9 | f |
| 177.0539 | 177.0546 | 10 | 187.0882 | C10H9O3+ | -4.0 | g |
| 161.0592 | 161.0597 | 10 | 171.0933 | C10H9O2+ | -3.1 | h |
| 139.0400 | 139.0390 | 7 | 146.0625 | C7H7O3+ | 7.2 | i |
| 135.0460 | 135.0441 | 8 | 143.0709 | C8H7O2+ | 14.1 | j |
| 111.0076 | 111.0077 | 5 | 116.0245 | C5H3O3+ | -0.9 | k |

**B**

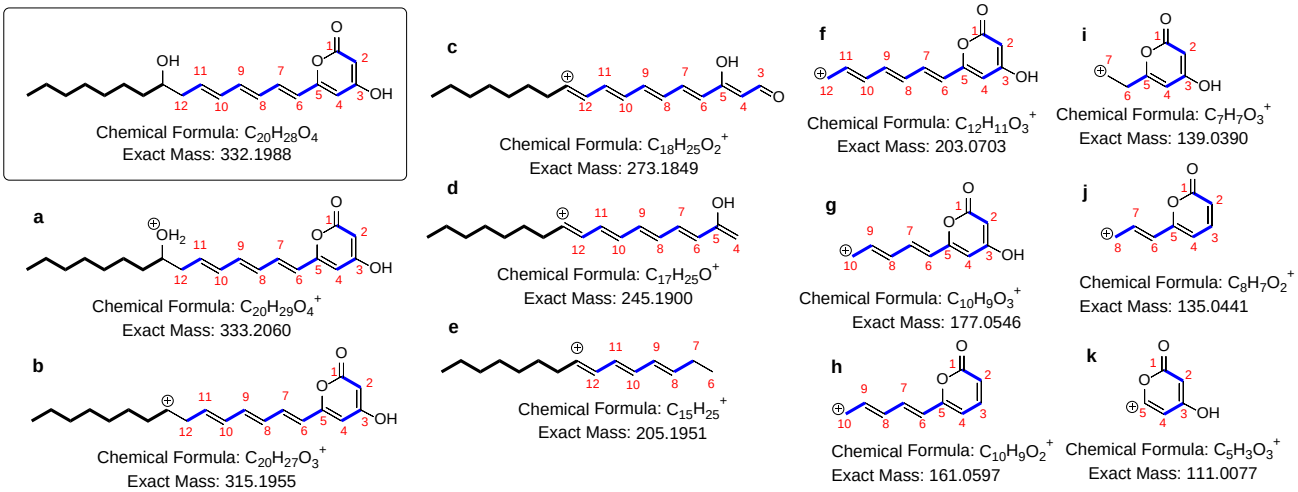



**Figure S40. Deduced structures of MS/MS fragments of desmethyl heptaketide 1, as observed in Figure S39.** (A) Compiled mass data for fragments observed with unlabeled and uniformly $^{13}$C-labeled samples, and (B) putative structures of each ion species.

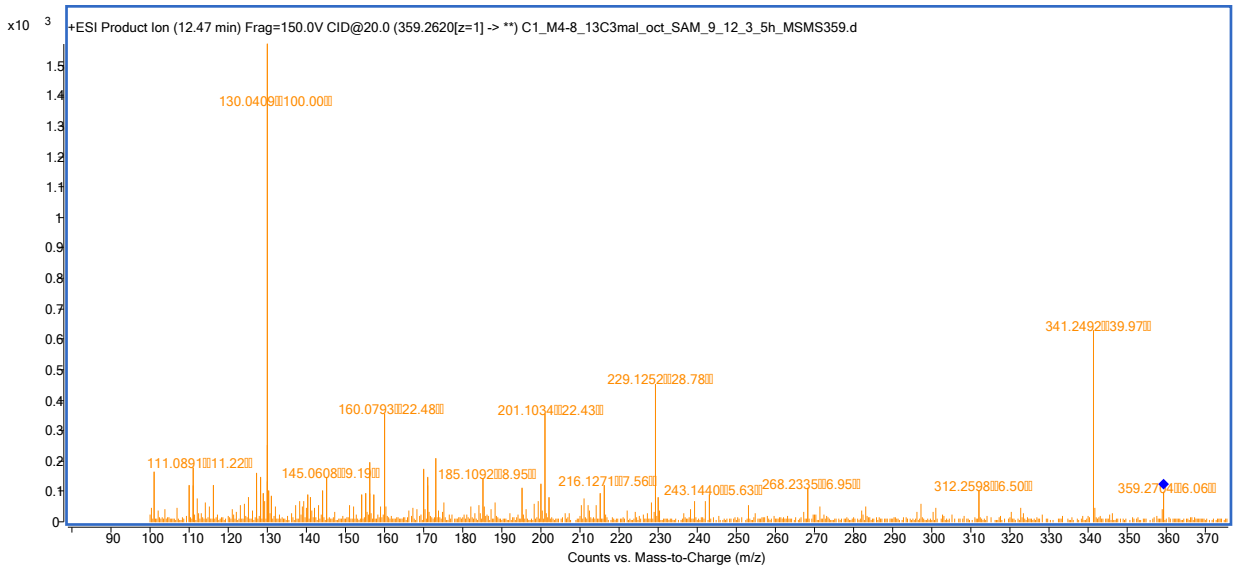



**Figure S41. MS/MS fragmentation analysis of methyl heptaketide 2**. Spectra (20V fragmentation) were obtained with uniformly $^{13}$C-labeled malonate.

**A**

| m/z (obsd.) | $^{12}$C m/z (calcd.) | # of [$^{13}$C] atoms | $^{13}$C m/z (calcd.) | Formula | Δm/z (ppm) | Putative structure |
|---|---|---|---|---|---|---|
| 359.2704 | 347.2217 | 12 | 359.262 | C21H31O4+ | 23.5 | a |
| 341.2492 | 329.2111 | 12 | 341.2514 | C21H29O3+ | -6.3 | b |
| 312.2598 | 301.2162 | 11 | 312.2531 | C20H29O2+ | 21.4 | c |
| 268.2335 | 259.2056 | 9 | 268.2358 | C18H27O+ | -8.6 | d |
| 229.1252 | 217.0859 | 12 | 229.1262 | C13H13O3+ | -4.2 | f |
| 201.1034 | 191.0703 | 10 | 201.1039 | C11H11O3+ | -2.2 | g |
| 185.1092 | 175.0754 | 10 | 185.109 | C11H11O2+ | 1.4 | h |
| 160.0793 | 153.0546 | 7 | 160.0781 | C8H9O3+ | 7.6 | i |
| 145.0608 | 139.0390 | 6 | 145.0591 | C7H7O3+ | 11.5 | j |
| 130.0409 | 125.0233 | 5 | 130.0401 | C6H5O3+ | 6.3 | k |

**B**

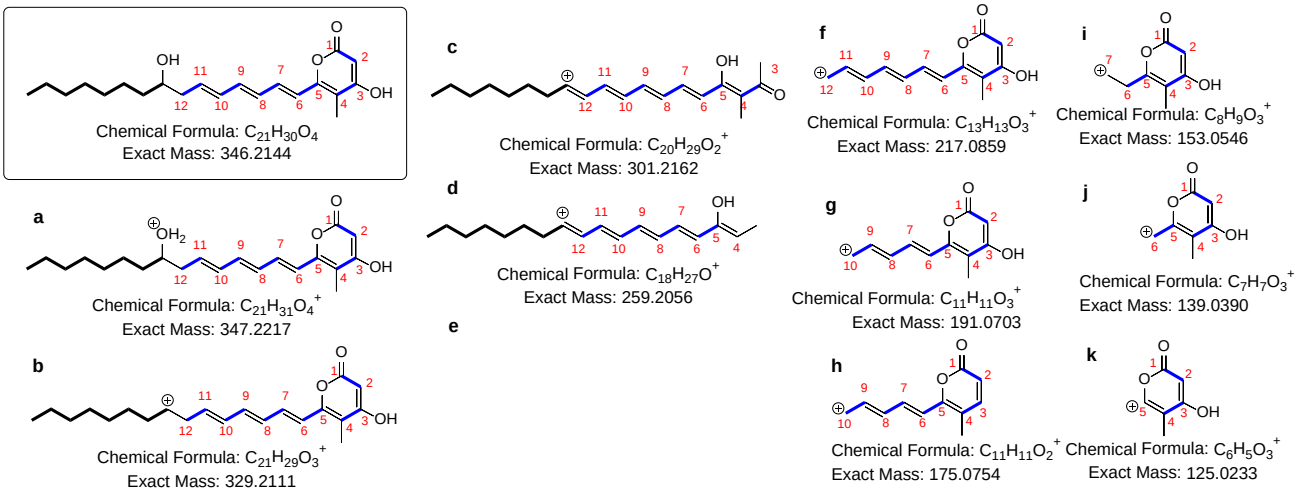



**Figure S42. Deduced structures of MS/MS fragments of methyl heptaketide 2, as observed in Figure S41.** (A) Compiled mass data for fragments observed, and (B) putative structures of each ion species.

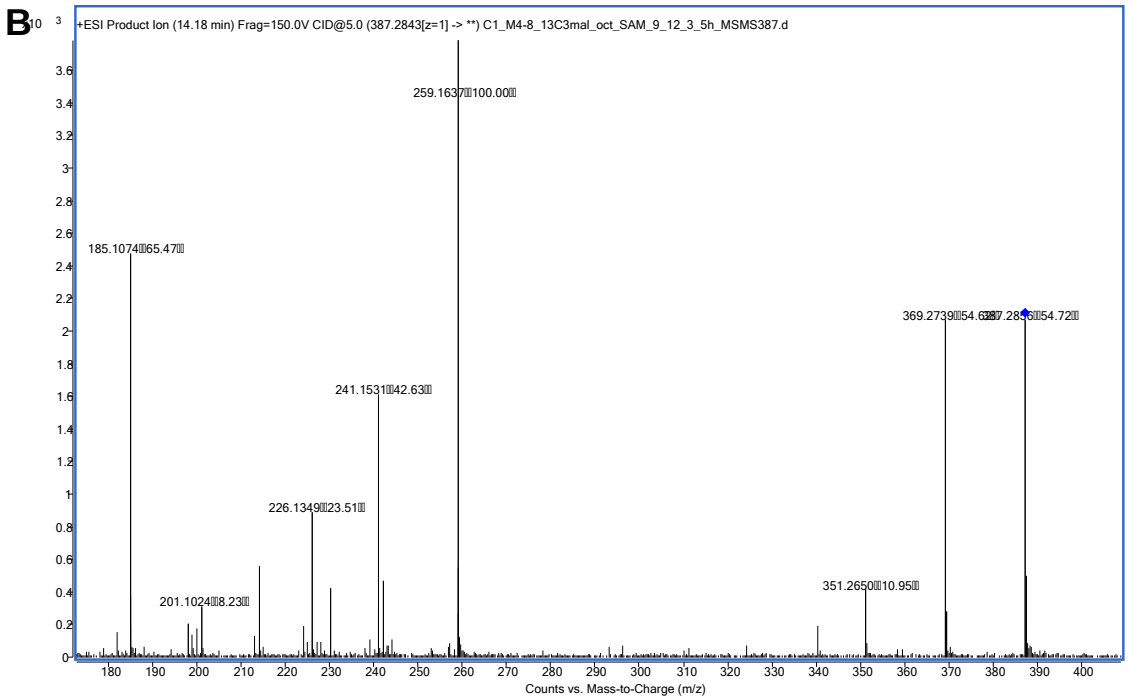
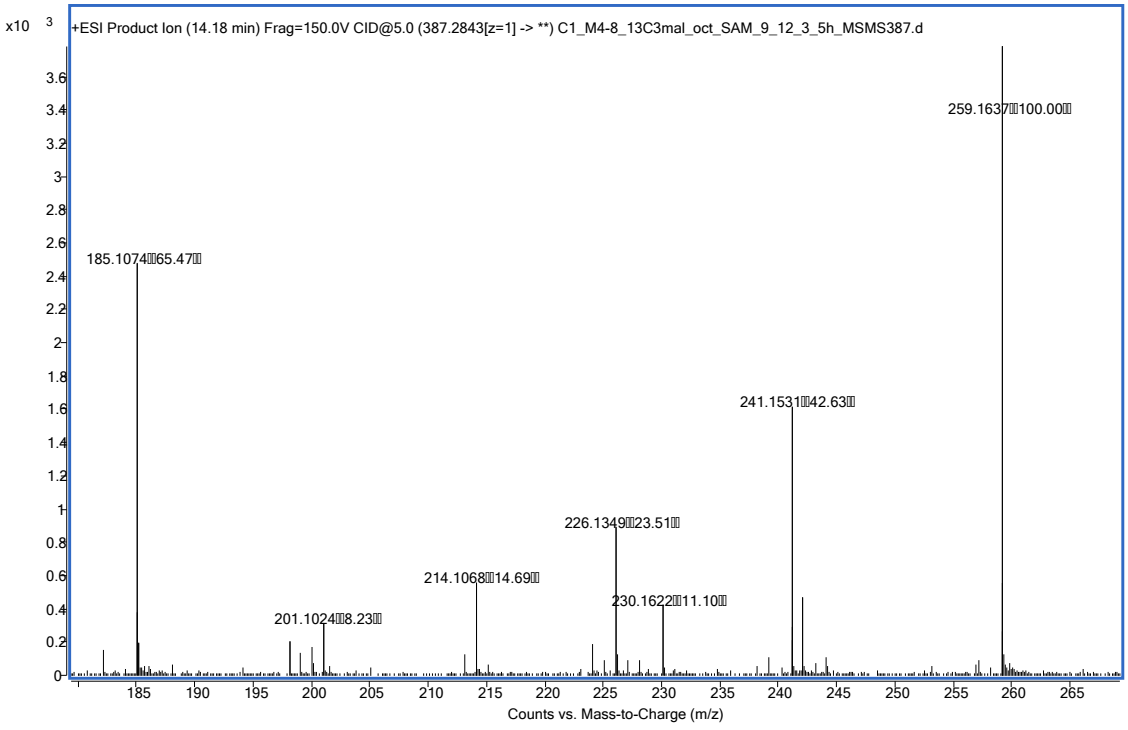



**Figure S43. MS/MS fragmentation analysis of methyl octaketide 4**. Spectra (5V




fragmentation) were obtained with uniformly $^{13}$C-labeled malonate. (A) Full view, (B) zoom-in view.

**A**

| m/z (obsd.) | $^{12}$C m/z (calcd.) | # of [$^{13}$C] atoms | $^{13}$C m/z (calcd.) | Formula | Δm/z (ppm) | Putative structure |
|---|---|---|---|---|---|---|
| 387.2856 | 373.2373 | 14 | 387.2843 | C23H33O4+ | 3.4 | a |
| 369.2674 | 355.2268 | 14 | 369.2738 | C23H31O3+ | -17.3 | b |
| 351.2650 | 337.2162 | 14 | 351.2632 | C23H29O2+ | 5.2 | c |
| 259.1637 | 245.1172 | 14 | 259.1642 | C15H17O3+ | -1.8 | d |
| 241.1531 | 227.1067 | 14 | 241.1537 | C15H15O2+ | -2.4 | e |
| 226.1349 | 213.0910 | 13 | 226.1346 | C14H13O2+ | 1.3 | f |
| 214.1068 | 203.0703 | 11 | 214.1072 | C12H11O3+ | -1.9 | g |
| 201.1024 | 191.0703 | 10 | 201.1039 | C11H11O3+ | -7.2 | h |
| 185.1074 | 175.0754 | 10 | 185.1090 | C11H11O2+ | -8.4 | i |
| 172.1052 | 163.0754 | 9 | 172.1056 | C10H11O2+ | -2.3 | j |

**B**

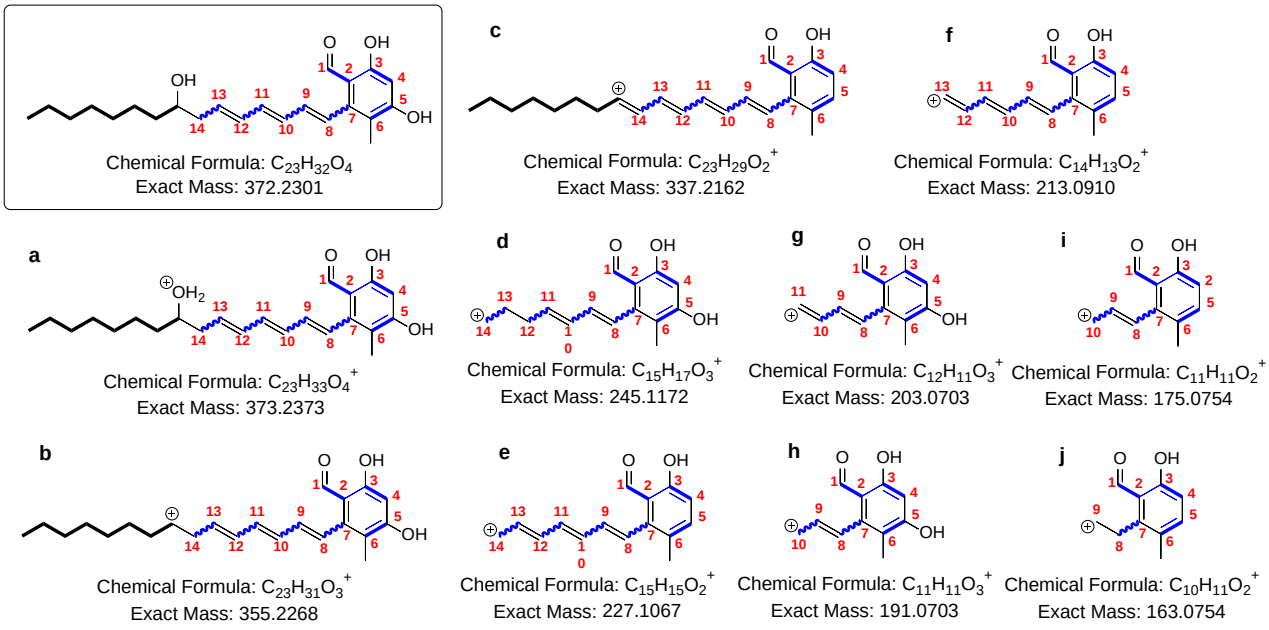
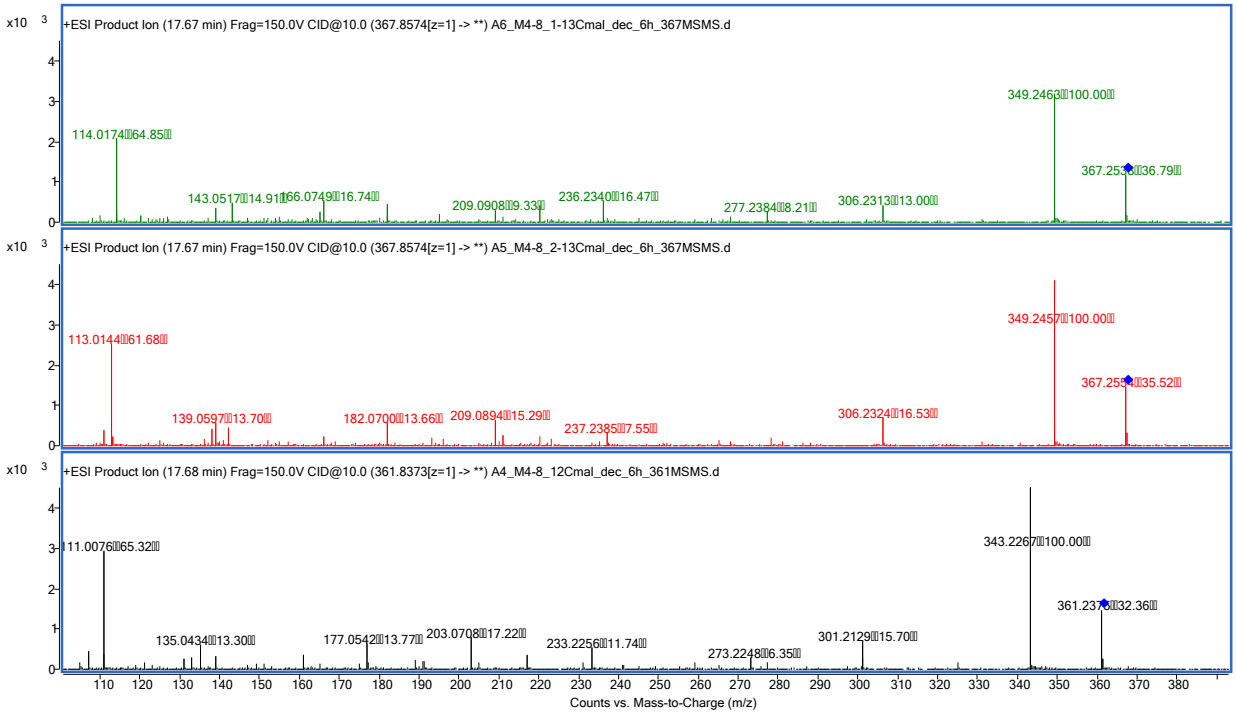



**Figure S44. Deduced structures of MS/MS fragments of methyl octaketide 4, as observed in Figure S43.** (A) Compiled mass data for fragments observed, and (B) putative structures of each ion species.

**A**




**B**

| m/z (obsd.) | $^{12}$C m/z (calcd.) | Formula | Δm/z (ppm) | Putative structure |
|---|---|---|---|---|
| 361.2375 | 361.2373 | C22H33O4+ | -0.6 | a |
| 343.2267 | 343.2268 | C22H31O3+ | 0.3 | b |
| 301.2129 | 301.2162 | C20H29O2+ | 11.0 | c |
| 273.2248 | 273.2213 | C19H29O+ | -12.8 | d |
| 233.2256 | 233.2264 | C17H29+ | 3.4 | e |
| 203.0708 | 203.0703 | C12H11O3+ | -2.5 | f |
| 177.0542 | 177.0546 | C10H9O3+ | 2.3 | g |
| 161.0606 | 161.0597 | C10H9O2+ | -5.6 | h |
| 139.0372 | 139.039 | C7H7O3+ | 12.9 | i |
| 135.0434 | 135.0441 | C8H7O2+ | 5.2 | j |
| 111.0076 | 111.0077 | C5H3O3+ | 0.9 | k |

**Figure S45. MS/MS fragmentation analysis of the decanoyl-CoA derived analog of heptaketide 1 (1b)**. (A) Spectra (10V fragmentation) were obtained with compound produced in the presence of [1,3-$^{13}C_2$]-labeled malonate (top), [2-$^{13}$C]-labeled malonate (middle), and unlabeled malonate (bottom). (B) Compiled mass data for fragments observed; the putative structures are analogous to those shown in Figure S40B.

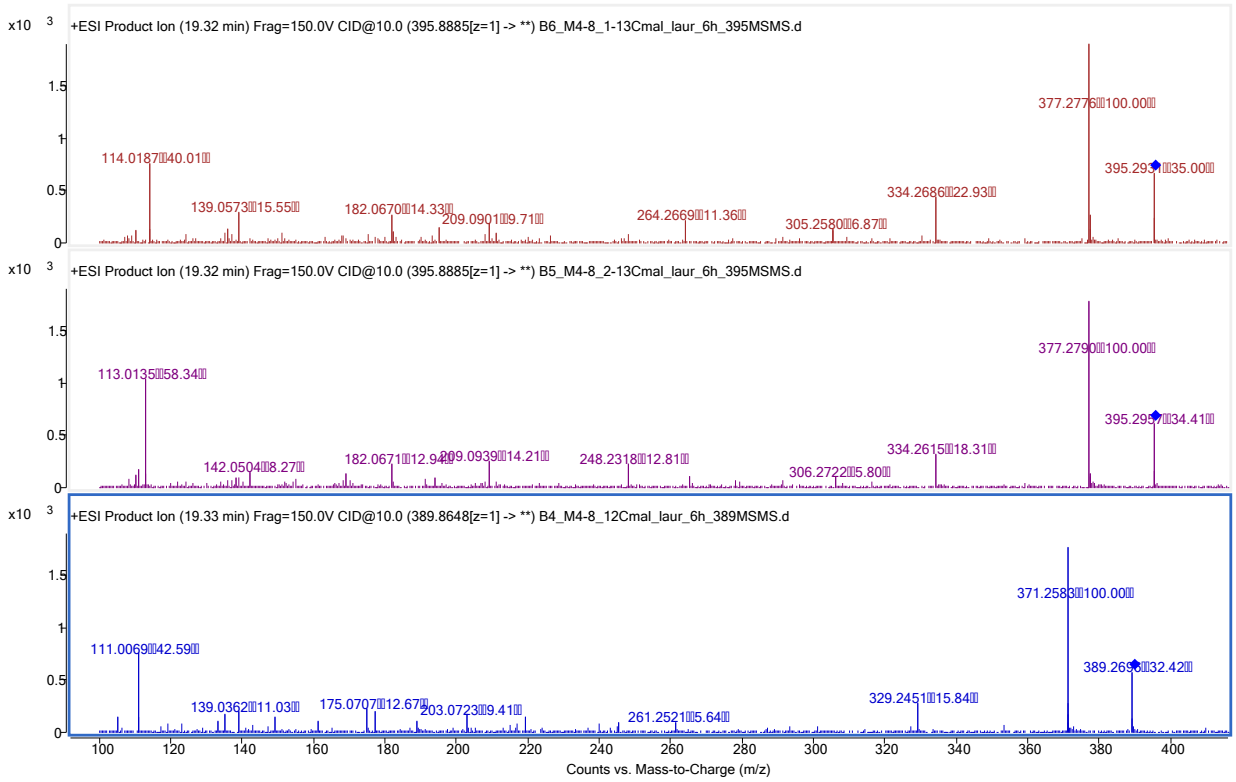



**Figure S46. MS/MS fragmentation analysis of the lauroyl-CoA derived analog of desmethyl heptaketide 1 (1c)**. Spectra (10V fragmentation) were obtained with compound produced in the presence of [1,3-$^{13}C_2$]-labeled malonate (top), [2-$^{13}$C]-labeled malonate (middle), and unlabeled malonate (bottom). The putative structures of the observed fragments are analogous to those shown in Figure S40B.

## References


(1) O'Brien, R. V.; Davis, R. W.; Khosla, C.; Hillenmeyer, M. E. Computational identification and analysis of orphan assembly-line polyketide synthases. *J Antibiot* **2014**, *67*, 89.

(2) Blin, K.; Medema, M. H.; Kazempour, D.; Fischbach, M. A.; Breitling, R.; Takano, E.; Weber, T. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* **2013**, *41*, W204.

(3) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009**, *6*, 343.

(4) Kuo, J.; Khosla, C. The initiation ketosynthase (FabH) is the sole rate-limiting enzyme of the fatty acid synthase of Synechococcus sp PCC 7002. *Metab Eng* **2014**, *22*, 53.

(5) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of E-coli. *Science* **2001**, *291*, 1790.

(6) Hughes, A. J.; Keatinge-Clay, A. Enzymatic extender unit generation for in vitro polyketide synthase reactions: structural and functional showcasing of Streptomyces coelicolor MatB. *Chem Biol* **2011**, *18*, 165.

(7) Dunn, B. J.; Cane, D. E.; Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **2013**, *52*, 1839.

(8) Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **2008**, *9*, 504.

(9) Kageyama, A.; Yazawa, K.; Mukai, A.; Kohara, T.; Nishimura, K.; Kroppenstedt, R. M.; Mikami, Y. Nocardia araoensis sp. nov. and Nocardia pneumoniae sp. nov., isolated from patients in Japan. *Int J Syst Evol Microbiol* **2004**, *54*, 2025.

(10) Iida, S.; Kageyama, A.; Yazawa, K.; Uchiyama, N.; Toyohara, T.; Chohnabayashi, N.; Suzuki, S.; Nomura, F.; Kroppenstedt, R. M.; Mikami, Y. Nocardia exalbida sp. nov., isolated from Japanese patients with nocardiosis. *Int J Syst Evol Microbiol* **2006**, *56*, 1193.

(11) Baio, P. V.; Ramos, J. N.; dos Santos, L. S.; Soriano, M. F.; Ladeira, E. M.; Souza, M. C.; Camello, T. C.; Ribeiro, M. G.; Hirata Junior, R.; Vieira, V. V.; Mattos-Guaraldi, A. L. Molecular identification of nocardia isolates from clinical samples and an overview of human nocardiosis in Brazil. *PLoS Negl Trop Dis* **2013**, *7*, e2573.

(12) Kageyama, A.; Poonwan, N.; Yazawa, K.; Mikami, Y.; Nishimura, K. Nocardia asiatica sp. nov., isolated from patients with nocardiosis in Japan and clinical specimens from Thailand. *Int J Syst Evol Microbiol* **2004**, *54*, 125.

(13) Kageyama, A.; Torikoe, K.; Iwamoto, M.; Masuyama, J.; Shibuya, Y.; Okazaki, H.; Yazawa, K.; Minota, S.; Kroppenstedt, R. M.; Mikami, Y. Nocardia arthritidis sp. nov., a new pathogen isolated from a patient with rheumatoid arthritis in Japan. *J Clin Microbiol* **2004**, *42*, 2366.

(14) le Roes, M.; Meyers, P. R. Nocardia gamkensis sp. nov. *Antonie Van Leeuwenhoek* **2006**, *90*, 291.

(15) Moser, B. D.; Klenk, H. P.; Schumann, P.; Potter, G.; Lasker, B. A.; Steigerwalt, A. G.; Hinrikson, H. P.; Brown, J. M. Nocardia niwae sp. nov., isolated from human pulmonary sources. *Int J Syst Evol Microbiol* **2011**, *61*, 438.

(16) Yassin, A. F.; Rainey, F. A.; Burghardt, J.; Brzezinka, H.; Mauch, M.; Schaal, K. P. Nocardia paucivorans sp. nov. *Int J Syst Evol Microbiol* **2000**, *50 Pt 2*, 803.
(17) Yassin, A. F.; Rainey, F. A.; Mendrock, U.; Brzezinka, H.; Schaal, K. P. Nocardia abscessus sp. nov. *Int J Syst Evol Microbiol* **2000**, *50 Pt 4*, 1487.
(18) Thoden, J. B.; Hegeman, A. D.; Wesenberg, G.; Chapeau, M. C.; Frey, P. A.; Holden, H. M. Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry* **1997**, *36*, 6294.
(19) Gurcha, S. S.; Baulard, A. R.; Kremer, L.; Locht, C.; Moody, D. B.; Muhlecker, W.; Costello, C. E.; Crick, D. C.; Brennan, P. J.; Besra, G. S. Ppm1, a novel polyprenol monophosphomannose synthase from Mycobacterium tuberculosis. *Biochem J* **2002**, *365*, 441.
(20) Perna, N. T.; Plunkett, G.; Burland, V.; Mau, B.; Glasner, J. D.; Rose, D. J.; Mayhew, G. F.; Evans, P. S.; Gregor, J.; Kirkpatrick, H. A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E. J.; Davis, N. W.; Limk, A.; Dimalanta, E. T.; Potamousis, K. D.; Apodaca, J.; Anantharaman, T. S.; Lin, J. Y.; Yen, G.; Schwartz, D. C.; Welch, R. A.; Blattner, F. R. Genome sequence of enterohaemorrhagic Escherichia coli O157 : H7. *Nature* **2001**, *409*, 529.
(21) Steffensky, M.; Muhlenweg, A.; Wang, Z. X.; Li, S. M.; Heide, L. Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891. *Antimicrob Agents Ch* **2000**, *44*, 1214.
(22) Bate, N.; Butler, A. R.; Smith, I. P.; Cundliffe, E. The mycarose-biosynthetic genes of Streptomyces fradiae, producer of tylosin. *Microbiology* **2000**, *146 ( Pt 1)*, 139.
(23) Bonin, C. P.; Freshour, G.; Hahn, M. G.; Vanzin, G. F.; Reiter, W. D. The GMD1 and GMD2 genes of Arabidopsis encode Isoforms of GDP-D-mannose 4,6-dehydratase with cell type-specific expression patterns. *Plant Physiol* **2003**, *132*, 883.
(24) Fouces, R.; Mellado, E.; Diez, B.; Barredo, J. L. The tylosin biosynthetic cluster from Streptomyces fradiae: genetic organization of the left region. *Microbiol-Uk* **1999**, *145*, 855.
(25) Blanco, G.; Patallo, E. P.; Brana, A. F.; Trefzer, A.; Bechthold, A.; Rohr, J.; Mendez, C.; Salas, J. A. Identification of a sugar flexible glycosyltransferase from Streptomyces olivaceus, the producer of the antitumor polyketide elloramycin. *Chemistry & Biology* **2001**, *8*, 253.
(26) Patallo, E. P.; Blanco, G.; Fischer, C.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. Deoxysugar methylation during biosynthesis of the antitumor polyketide elloramycin by Streptomyces olivaceus - Characterization of three methyltransferase genes. *J Biol Chem* **2001**, *276*, 18765.
(27) Morita, Y. S.; Sena, C. B. C.; Waller, R. F.; Kurokawa, K.; Sernee, M. F.; Nakatani, F.; Haites, R. E.; Billman-Jacobe, H.; McConville, M. J.; Maeda, Y.; Kinoshita, T. PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria. *J Biol Chem* **2006**, *281*, 25143.
(28) Karoly, E. D.; Rose, R. L. Sequencing, expression, and characterization of cDNA expressed flavin-containing monooxygenase 2 from mouse. *J Biochem Mol Toxic* **2001**, *15*, 300.

(29) Zhao, B.; Lin, X.; Lei, L.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R.; Cane, D. E. Biosynthesis of the sesquiterpene antibiotic albaflavenone in Streptomyces coelicolor A3(2). *J Biol Chem* **2008**, *283*, 8183.
(30) Guan, S.; Bastin, D. A.; Verma, N. K. Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiol-Uk* **1999**, *145*, 1263.
(31) Allen, E. E.; Bartlett, D. H. Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9. *Microbiol-Sgm* **2002**, *148*, 1903.
(32) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **2014**, *22*, 444.
(33) Dunn, B. J.; Watts, K. R.; Robbins, T.; Cane, D. E.; Khosla, C. Comparative Analysis of the Substrate Specificity of trans- versus cis-Acyltransferases of Assembly Line Polyketide Synthases. *Biochemistry* **2014**, *53*, 3796.