Production of Octaketide Polyenes by the Calicheamicin Polyketide Synthase CalE8: Implications for the Biosynthesis of Enediyne Core Structures

Katherine Belecki, Jason M. Crawford[†], and Craig A. Townsend*

Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218

email: ctownsend@jhu.edu

†   present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, 02115

Supporting Information

Materials.   PCR primers were ordered from Sigma-Genosys (St. Louis, MO). Subcloning vector pCR-Blunt was obtained from Invitrogen (Carlsbad, CA), and pET expression vectors from Novagen (Madison, WI). Malonyl-CoA was purchased from Sigma-Aldrich (St. Louis, MO). NADPH was purchased from Roche (Mannheim, Germany).

Cloning and Mutagenesis.

General DNA manipulation. Routine cloning was carried out in *Escherichia coli* DH5α. All primers used for amplification of *calE8* and *calE7* (GenBank accession number AF497482) and for construction of the CalE8 mutant are listed in Table S1 (engineered restriction sites are underlined). Plasmids and linear DNA fragments were purified using commercially available kits (Qiagen, Valencia, CA), and all other DNA manipulations were performed according to standard procedures. Amplified DNA was sequence-verified (Synthesis and Sequencing Facility at Johns Hopkins Medical School, Baltimore, MD), and polymerase errors resulting in nonsynonymous mutations were repaired.

CalE8. Repeated attempts to amplify the 6-kb *calE8* in one piece were not successful, despite variations in annealing temperatures, extension times, PCR techniques, and/or primer lengths, likely owing to its length and GC-rich nature. The native sequence was therefore cloned from the gDNA of *Micromonospora echinospora* ssp. *calichensis* (NRRL 15839) in two pieces using primer pairs CalE8-5L/CalE8-Dra3 and CalE8-Dra5/CalE8-3L. The resulting PCR products were ligated into subcloning vector pCR-Blunt giving pCCalE85 and pCCalE83 respectively. The full-length *calE8* was constructed in pET-24a(+) making use of a unique *Dra*III restriction site close to the middle of the gene and the engineered *Nde*I and *Hind*III sites to give expression vector pECalE8.

The stop codon of *calE8* was removed by amplification of a fragment from subcloning vector pCCalE83 using primers E8Pst-5 and CalE8-3His. This sequence was subcloned into pCR-Blunt and subsequently used to replace the analogous *Pst*I-*Hind*III fragment of pECalE8, putting

CalE8 in frame with the C-terminal 6xHis tag coding sequence in the new construct, pECalE8His.

CalE8-C211A. The cysteine 211 mutation of CalE8 was introduced by overlap extension PCR using external primers CalE8-5 and CalE8-Dra3, internal primers E8-C211A-5 and E8-C211A-3 (altered base pairs highlighted in bold), and subcloning vector pCCalE85 as a template. The resultant PCR product was ligated into pCR-Blunt, giving subcloning vector pCCalE85-C211A. The *Nde*I-*Dra*III fragment of pECalE8His containing coding sequence for the first half of *calE8* was excised and replaced with the analogous fragment from pCCalE85-C211A to give pECalE8H-C211A, a bacterial expression vector encoding the KS domain mutant CalE8-C211A with a C-terminal 6xHis tag.

CalE7. The amplification of *calE7* from the gDNA template using primers CalE7-5 and CalE7-3 was straightforward. The PCR product was subcloned into pCR-Blunt prior to cloning in pET-28b(+) using the engineered *Nde*I and *Xho*I restriction sites. The resulting vector, pECalE7His, encodes *calE7* with an N-terminal 6xHis tag.

Table S1. Primers used in this study

Name	Sequence
CalE8-5	5-TCTAGACATATGAGCAGGATCGCCGTCGTC-3
CalE8-5L	5-TCTAGACATATGAGCAGGATCGCCGTCGTCGGCCTGGCCTGCCGCTTC-3
CalE8-3L	5-AAGCTTCTAGATCACCTCCCCGCCCCGCTGAGCACGGCGACCACGGCCGG-3
CalE8-Dra3	5'-CGTTCCGCCACGAGGTGCCGCACCAGCTCGAG-3'
CalE8-Dra5	5'-CTCGAGCTGGTGCGGCACCTCGTGGCGGAACG-3'
CalE83His	5-AAGCTTCCTCCCCGCCCCGCTGAGC-3
E8Pst-5	5-GGTCTGGTGCGCGGAGGAGTG-3
E8-C211A-5	5-G GTG GAC GGC GCC **GC**C GCC TCC TCC CTG C-3
E8-C211A-3	5-G CAG GGA GGA GGC G**GC** GGC GCC GTC CAC C-3
CalE7-5	5-GGATCCATATGAGCATGCCGCGCTACTACG-3
CalE7-3	5-GGATCCTCGAGTCACCCCTGCCCCACCGTGG-3

Protein Expression and Purification. *E. coli* Rosetta2(DE3) cells harboring pECalE8His, pECalE8H-C211A, or pECalE7His were grown in LB media supplemented with kanamycin (25 μg/mL), chloramphenicol (25 μg/mL) and glycerol (10%) at 37 °C to OD_{600} = 0.6 and cooled for 10 min at 4 °C. Protein expression was induced with IPTG at a final concentration of 1 mM. After 18 h at 19 °C, cells were pelleted (4000 x g) and resuspended in lysis buffer (300 mM sodium chloride, 50 mM potassium phosphate pH 7.6, 10 mM imidazole, 10% glycerol) with 1 mg/mL lysozyme. Following sonication, the lysate was clarified by centrifugation (25,000 x g) and protein was purified using Ni-NTA resin (Qiagen) according to the manufacturer's instructions. Fractions containing the desired protein were dialyzed against 50 mM potassium phosphate pH 7.0 with 10% glycerol and 2 mM DTT. CalE8 and CalE8-C211A solutions were concentrated by Amicon Ultra (Millipore, Billerica, MA), and enzyme purity for CalE8 (203.9 kD with 6xHis tag) and CalE7 (19.4 kD with 6xHis tag) was assessed by SDS-PAGE (Figure

S1). Absorption spectra were recorded on a Cary 50 UV/Vis Spectrophotometer (Varian, Palo Alto, CA). Protein concentrations were determined in triplicate by Bradford assay (Bio-Rad, Hercules, CA) using bovine serum albumin as a standard. Enzymes were stored at –80 °C and used without further purification or modification.
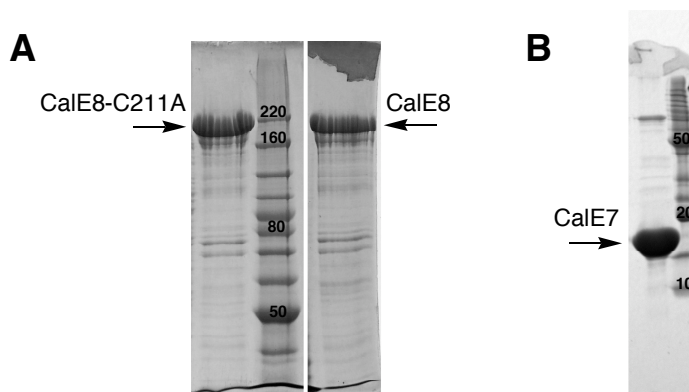

Figure S1. (A) 7% SDS-PAGE gel. (B) 15% SDS-PAGE gel.

***In vitro* reactions of CalE8 and CalE7.** CalE8 (1 mg/mL), CalE7 (0.7 mg/mL), malonyl-CoA (0.5 mM) and NADPH (0.5 mM) were incubated at 24 °C in 50 mM potassium phosphate buffer (pH 7.0) containing 10% glycerol. Control reactions contained CalE8-C211A (1 mg/mL) in place of CalE8. Reactions were monitored by UV/Vis spectroscopy and were typically allowed to proceed for 2 h. Unless noted, reactions were extracted directly with ethyl acetate to isolate polyketide products. Extracts were concentrated by rotary evaporator and resuspended in HPLC starting buffer or methanol for analysis.

In order to monitor the decarboxylation of β-keto acid **3** to methyl ketone **2**, ethyl acetate extracts from a typical reaction were separated into three equal portions. One aliquot was concentrated and analyzed by HPLC immediately. The other two aliquots were stored as ethyl acetate solutions at 0 °C for 1 or 5 d prior to concentration and analysis by HPLC.

Analysis of polyketide products.

Instrumentation and general considerations. Routine HPLC separation was carried out on either an Agilent 1100 series or 1200 series system (Santa Clara, CA) and analyzed using a diode array detector. Unless noted, samples were injected onto a 4.6 mm x 250 mm Prodigy ODS3 100 Å, 5 μm column (Phenomenex, Torrence, CA) and separated using a gradient of acetonitrile with 0.1% (v/v) formic acid (solvent A) in water with 0.1% (v/v) formic acid (solvent B) and a flow rate of 1 mL/min. The octaketide polyenes **1**-**3** were monitored at 375 nm and truncation products **8**-**11** were monitored at 280 nm and 325 nm.

Nominal masses were obtained using a Bruker Daltonics (Billerica, MA) Esquire 3000 LC-Quadrupole Ion Trap at the Mass Spectrometry Facility located in the Structural

Biochemistry Center at the University of Maryland, Baltimore County. High-resolution mass spectrometry was performed on a JEOL (Tokyo, Japan) AccuTOF-CS at the Department of Chemistry and Biochemistry Mass Spectrometry Facility at the University of Maryland, College Park using chalcone (calculated m/z = 209.06640 for MH$^+$) as a standard. Both LC/MS instruments used in this study interface with an Agilent 1100 series HPLC system and can be equipped with either an electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) source. All mass spectra were collected in positive mode; reported masses correspond to MH$^+$ ions for each compound.

Each mass was correlated to a corresponding UV/Vis peak based on its extracted ion chromatogram (EIC). All reported compounds were verified to be absent from CalE8-C211A control reactions based on HPLC chromatogram overlays, EIC overlays, and direct comparison of mass spectra.

HPLC and LC/MS analysis of octaketide polyenes. A gradient of 40% to 90% solvent A over 0 to 20 min was used to monitor the conversion of **3** to **2** shown in Figure 2 of the main text. For routine monitoring of all three octaketides (**1**-**3**) a more efficient gradient was developed (55% to 90% solvent A over 0 to 18 min, 90% solvent A over 18 to 28 min) as the standard program.

Nominal mass data for **2** (m/z = 215.2) was obtained with LC-ESI-MS using a slight modification of the standard program (55% to 80% solvent A over 0 to 18 min). The EIC for this mass mirrors the peak for **2** in the HPLC chromatogram, and the UV/Vis spectrum from the diode array detector is consistent with that reported by Liang and coworkers (see reference 16 from main text). High resolution mass spectrometry was obtained for methyl ketone **2** using both LC-APCI-MS and LC-ESI-MS giving a mass of m/z = 215.1436 (calculated 215.14359 for $C_{15}H_{18}OH^+$). A T-joint was installed between the HPLC detector and ion source in order to accommodate both the low flow rate required for the AccuTOF instrument and the 1 mL/min flow rate of the separation program.

The same nominal mass LC-ESI-MS program used for **2** was employed for **3**, correlating a mass of m/z = 259.2 with the broad peak observed in the HPLC chromatogram. Successful high resolution LC/MS of **3** required several technical optimizations. It was necessary to minimize the amount of time elapsed between *in vitro* reaction and HRMS analysis owing to the transient nature of this compound. Also, in order to avoid partitioning the sample with a T-joint, the HPLC column was replaced with a 2.1 mm x 150 mm Agilent Zorbax Eclipse XDB-C18 column (5 μm) and analyzed using an isocratic method (100% methanol) with a flow rate of 0.2 mL/min. Using LC-APCI-MS, the exact mass of **3** was determined to be m/z = 259.1373, establishing its molecular formula as $C_{16}H_{18}O_3$ (calculated 259.13342 for MH$^+$).

The nominal mass of **1** was found to be m/z = 199.2 using the LC-APCI-MS program reported by Shen and coworkers (see reference 15 from main text) but with an Agilent Zorbax Eclipse XDB-C8 column (4.6 mm x 150 mm, 5 μm). EIC analysis verified the association of this mass with elution of the heptaene chromophore. High-resolution mass data, obtained using the same parameters as for β-keto acid **3** above, confirmed a molecular formula assignment of $C_{15}H_{18}$ for heptaene **1** (found m/z = 199.1479, calculated 199.14868 for MH$^+$).

HPLC and LC/MS analysis of truncation products. Reactions intended for analyzing truncation products **8**-**11** were acidified with hydrochloric acid prior to ethyl acetate extraction, and more polar HPLC methods were developed for their observation. Separation of these four compounds can be achieved using a gradient of 10 to 15% solvent A over 0 to 5 min, 15 to 40% solvent A over 5 to 7 min, and 40 to 90% solvent A over 7 to 35 min (Figure S2). The chromatogram at 325 nm is shown for simplicity, however **10** and **11** are best visualized at other wavelengths.
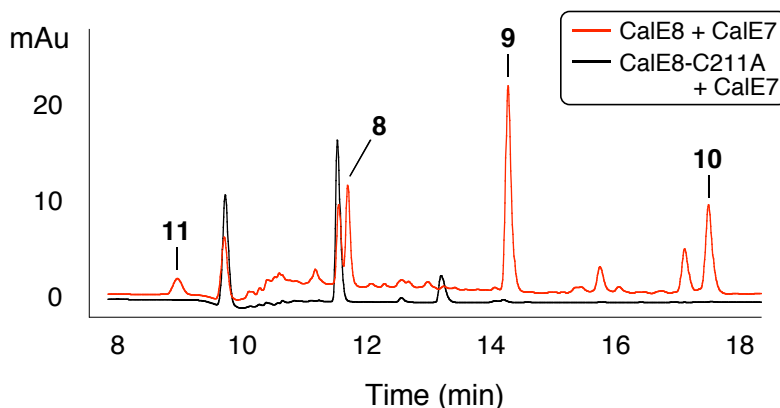

Figure S2. HPLC analysis of truncation products from an *in vitro* reaction of CalE8 and CalE7. Chromatogram was recorded at 325 nm.

Nominal and exact masses were obtained using LC-ESI-MS, revealing a general formula of $C_xH_xO_3$ for this set of compounds (Table S2). Fragmentation data is consistent with the 2-pyrone structures. The identity of triketide **11** has been confirmed by co-migration with an authentic sample of triactetic acid lactone (4-hydroxy-6-methyl-2-pyrone, purchased from Sigma-Aldrich).

Table S2. Exact mass data for truncation products

Compound	Formula	Mass Found	Calculated MH$^+$
11	$C_6H_6O_3$	127.0361	127.03951
8	$C_8H_8O_3$	153.0525	153.05517
9	$C_{10}H_{10}O_3$	179.0703	179.07082
10	$C_{12}H_{12}O_3$	205.0860	205.08647