Supporting Information Table of Contents

Experimental Section

Bacterial Strains and Growth Conditions. Genomic DNA from *S. hygroscopicus* subsp. *hygroscopicus* ATCC 53709 was used to PCR amplify *epnF*, *epnG*, *epnH*, and *epnI*, and genomic DNA from *G. coeruleoviolacea* ATCC 53904 was used to PCR amplify *epxE*. *E. coli* XL1-Blue was used for all cloning procedures and *E. coli* BAP1 and *E. coli* BL21 Gold (DE3) were used for protein expression. LB was used for the propagation of *E. coli* at 37ºC unless specified otherwise. Growth media was supplemented with antibiotics as required at the following concentrations: kanamycin (50 μg/mL), chloramphenicol (25 μg/mL), and spectinomycin (50 μg/mL).

Construction of Plasmids for Protein Expression in *E. coli*. Plasmids and primers used in this study are listed in Table S1 and S2, respectively. Individual genes were PCR amplified from genomic DNA and cloned into pET24b, pACYCDuet-1, or pCDFDuet-1 by restriction enzyme digestion (Thermo Scientific) and ligation with Quick T4 DNA ligase (New England Biolabs). To construct pCDFDuet-*epnG*-*epnI*, pCDFDuet-*epnG* was digested with NdeI/KpnI and *epnI* was cloned into the second MCS. For the overexpression and purification of EpnI, *epnI* was cloned into pET30 using the pET30 Xa/LIC Vector Kit (Novagen) to introduce an *N*-terminal hexahistidine tag. Plasmids were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by DNA Sequencing (UC Berkeley DNA Sequencing Facility).

Biosynthesis of Peptidyl Epoxyketones in *E. coli*. The plasmids, pACYCDuet-*epnF*, pCDFDuet-*epnG*-*epnI*, and pET24b-*epnH*, were electroporated into *E. coli* BAP1 to form strain JL6, and transformants were selected on LB agar plates supplemented with the appropriate antibiotics. Single colonies were inoculated into 3 mL of LB with antibiotics and grown overnight at 37 ºC as a seed culture, of which 0.25 mL was used to inoculate 25 mL of fresh LB with antibiotics. The cultures were then grown at 37 ºC to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.5 mM IPTG and the addition of hexanoic or octanoic acid at a final concentration of 1 mM. After induction, the temperature was dropped to 20ºC, and compound production was allowed to proceed for approximately two days. For the generation of strains JL7, 8, 9, 10, and 11, different combinations of the plasmids listed in Table S1 were electroporated into *E. coli* BAP1.

LC-HRMS Analysis of Peptidyl Epoxyketones Produced by *E. coli*. *E. coli* culture samples were pelleted by centrifugation (4000 × *g* for 10 min), and the supernatant was extracted with two volumes of ethyl acetate. The solvent was removed by rotary evaporation, and the residue was re-dissolved in 500 μL methanol for liquid chromatography-high resolution mass spectrometry (LC-HRMS) and HRMS/MS analysis (10 μL injection). LC-HRMS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument with an Agilent Eclipse Plus C18 column (4.6 x 100 mm). A linear gradient of 2-95% CH_3CN (vol/vol) over 40 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used. For HRMS/MS experiments with **3** and **4**, a linear gradient of 40-80% CH_3CN (vol/vol) over 14 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min and a collision energy of 10 V was used.

Large-scale Production, Purification, and Characterization of 4. A total of 12-L (12 ×1-L) of strain JL11 (*E. coli* BAP1 co-transformed with pET24b-*epxE*, pACYCDuet-*epnF*, and pCDFDuet-*epnG*) was cultured in LB with antibiotics. The cultures were inoculated with 10 mL of a seed culture and grown at 37 ºC to $OD_{600} \approx 0.4 – 0.6$ before induction with 0.5 mM IPTG and the addition of octanoic acid to a final concentration of 1 mM. After induction, the temperature was dropped to 20ºC, and compound production was allowed to proceed for approximately two days. Compound **4** was extracted from the cell-free supernatant using two volumes of ethyl acetate. The solvent was removed by rotary evaporation, and the combined residue was re-dissolved in 3 mL dichloromethane. The residue was chromatographed on a silica gel column and eluted with dichloromethane-methanol (95:5, vol/vol). Fractions containing **4** were determined by LC-MS (a linear gradient of 60-70% CH_3CN (vol/vol) over 10 min in H_2O with 0.1% formic acid (vol/vol) at a flow rate of 0.5 mL/min was used) and were combined for further purification by HPLC (Agilent 1260 HPLC) with an Inertsil ODS-4 column (4.6 mm i.d., 250 mm length, GL Sciences Inc.) using an isocratic program of 50% CH_3CN (vol/vol) in H_2O at a flow rate of 0.5 mL/min. The resulting 2 mg of purified **4** was dried and analyzed by LC-HRMS and NMR. NMR spectra (1D: ^{1}H, ^{13}C and 2D: HSQC, COSY, HMBC) were recorded on a Bruker Biospin 900 MHz spectrometer with a cryoprobe in chloroform-d ($CDCl_3$; Cambridge Isotope Laboratories).

Overexpression and Purification of Proteins. The plasmids, pET24b-*epnH*, pET24b-*epxE*, and pCDFDuet-*epnG*, were transformed into *E. coli* BAP1, and the plasmids, pACYCDuet-*epnF* and pET30-*epnI*, were transformed into *E. coli* BL21 Gold (DE3) for protein expression. Expression and purification for all proteins with a His_6-tag followed the same general procedure and is detailed as follows: cells were grown at 37°C in 700 mL of LB with the appropriate antibiotic to an OD_{600} of 0.5. The cells were then cooled on ice for 10 min and induced with 0.12 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 16°C. Subsequently, the cells were harvested by centrifugation (6371 × *g*, 15 min, 4°C), resuspended in 30 mL lysis buffer (25 mM HEPES, pH 8, 0.5 M NaCl, 5 mM imidazole), and lysed by homogenization on ice. Cellular debris was removed by centrifugation (27216 × *g*, 1 h, 4°C). Ni-NTA agarose resin was added to the supernatant (3 mL/L of culture), and the solution was nutated at 4°C for 1 h. The protein-resin mixture was loaded onto a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A (20 mM HEPES, pH 8.0, 1 mM DTT). Purified proteins were concentrated and buffer exchanged into Buffer A + 10% glycerol using Vivaspin Centrifugal Concentrators.

ATP-PP_i Exchange Assays. Assays were performed in 100 µL of reaction buffer (50 mM Tris-HCl/2 mM $MgCl_2$, pH 8) containing 5 mM ATP, 1 mM $Na_4[^{32}P]$-pyrophosphate (PP_i) (~3 x 10^6 cpm/mL), 1 mM TCEP, 5 mM substrate, and 5 µM enzyme. Reactions were incubated at 25°C for 2 h, then quenched by the addition of a charcoal suspension (1.6% w/v activated charcoal, 0.1 M Na_4PP_i, 3.5% $HClO_4$). Free $[^{32}P]PP_i$ was removed by centrifugation of the sample followed by washing twice with wash solution (0.1 M Na_4PP_i and 3.5% $HClO_4$). Charcoal-bound radioactivity was measured on a Beckman LS 6500 scintillation counter.

^{13}C-labeled Precursor Feeding Experiments. Strain JL11 was cultured for the production of **3** and **4** as described above, and at the time of IPTG induction and the addition of octanoic acid, either unlabeled sodium acetate, [1-^{13}C] sodium acetate, [1,2-^{13}C] sodium acetate, unlabeled L-

methionine, or [methyl-^{13}C] L-methionine was added to cultures to a final concentration of 1 g/L. Compound extraction and LC-HRMS and HRMS/MS analyses were performed as described above.

Supplementary Tables

Table S1. Plasmids used in this study.

Plasmid	**Derived from**	**Function**
pACYCDuet-*epnF*	pACYCDuet-1	Expression of EpnF in *E. coli*
pET24b-*epnH*	pET24b	Expression of EpnH in *E. coli*
pCDFDuet-*epnG*	pCDFDuet-1	Expression of EpnG in *E. coli*
pCDFDuet-*epnG-epnI*	pCDFDuet-1	Co-expression of EpnG and EpnI in *E. coli*
pCDFDuet-*epnI*	pCDFDuet-1	Expression of EpnI in *E. coli*
pET24b-*epxE*	pET24b	Expression of EpxE in *E. coli*
pET30-*epnI*	pET30-Xa/LIC	Expression of EpnI in *E. coli*

Table S2. Primers used in this study.

Primer	**Sequence (5' -> 3')**	**Description**
epnF-Duet-F	aaaCTGCAGgtgagtgacagcaaatcggt	Expression of EpnF
epnF-Duet-R	tatAAGCTTtcatcgcttccccgggtgag	
NdeI-epnH-F	aaaCATATGacgtcgaatcaacagat	Expression of EpnH
XhoI-epnH-R	aaaCTCGAGgtgattctctttctgcgata	
epnG-Duet-F	aaaCTGCAGatgaacaaggaaaaggaccg	Expression of EpnG
epnG-Duet-R	tatAAGCTTtcatgagttgcggttcctcc	
epnI-Duet-F	aaaCATATGgtgacgatcgacccgaa	Expression of EpnI
epnI-Duet-R	aaaGGTACCtcaggcatggacgttctctc	
epxE-Duet-F	aaaCATATGgttgacgagaccgccgt	Expression of EpxE
HindIII-epxE-R	tatAAGCTTctgggtaaacctccgccttt	
pET30-epnI-F	GGTATTGAGGGTCGCatggtgacgatcgacccgaa	Expression of EpnI
pET30-epnI-R	AGAGGAGAGTTAGAGCCtcaggcatggacgttctctc	

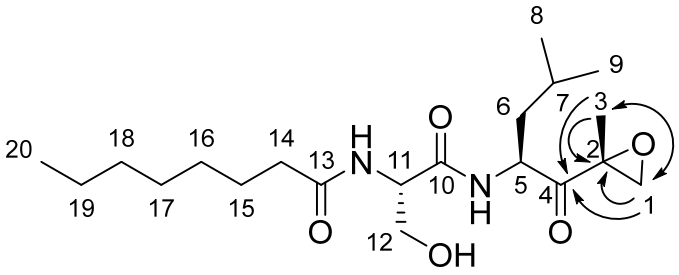

Table S3. NMR Data for **4** in $CDCl_3$, 900 MHz.

Position	δ_H (*J* in Hz)	δ_c	COSY	HMBC
1	2.91, d (4.5) 3.28, d (4.5)	52.7	1	2, 3, 4
2		59.5		
3	1.52, s	16.9		1, 2, 4
4		208.8		
5	4.50, m	51.3	5-NH, 6	4, 6, 7, 10
5-NH	6.87, d (5.4)		5	5, 6, 10
6	1.29, m 1.57, m	39.4	5, 6, 7	5, 7, 8, 9
7	1.64, m	25.4	6, 8, 9	5, 6, 9
8	0.93, d (7.2)*	21.1*	7	6, 7, 9
9	0.94, d (7.2)*	23.5*	7	6, 7, 8
10		171.6		
11	4.45, m	53.4	11-NH, 12	10, 12, 13
11-NH	6.46, d (4.5)		11	10, 11, 12, 13
12	3.54, m 4.05, m	62.8	11, 12	10, 11
13		174.0		
14	2.21, t (6.3)	36.6	15	13, 15, 16
15	1.60, m	25.8	14, 16	13, 14, 16, 17
16	1.28, m	29.1	15, 17	17, 18
17	1.25, m	29.3	16, 18	16, 19
18	1.24, m	31.8	17, 19	16, 17, 19, 20
19	1.27, m	22.7	18, 20	17, 18, 20
20	0.87, t (7.2)	14.2	19	18, 19

* May be interchanged

Supplementary Figures

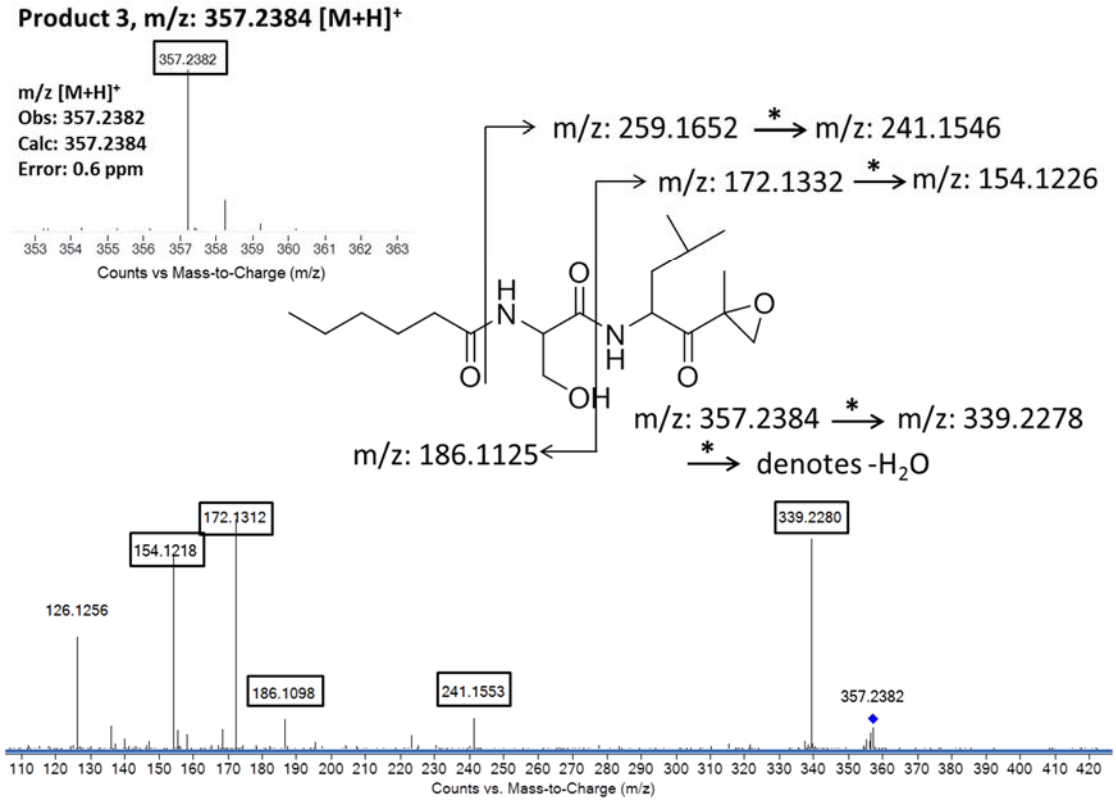

Figure S1. HRMS and HRMS/MS analysis of **3**.

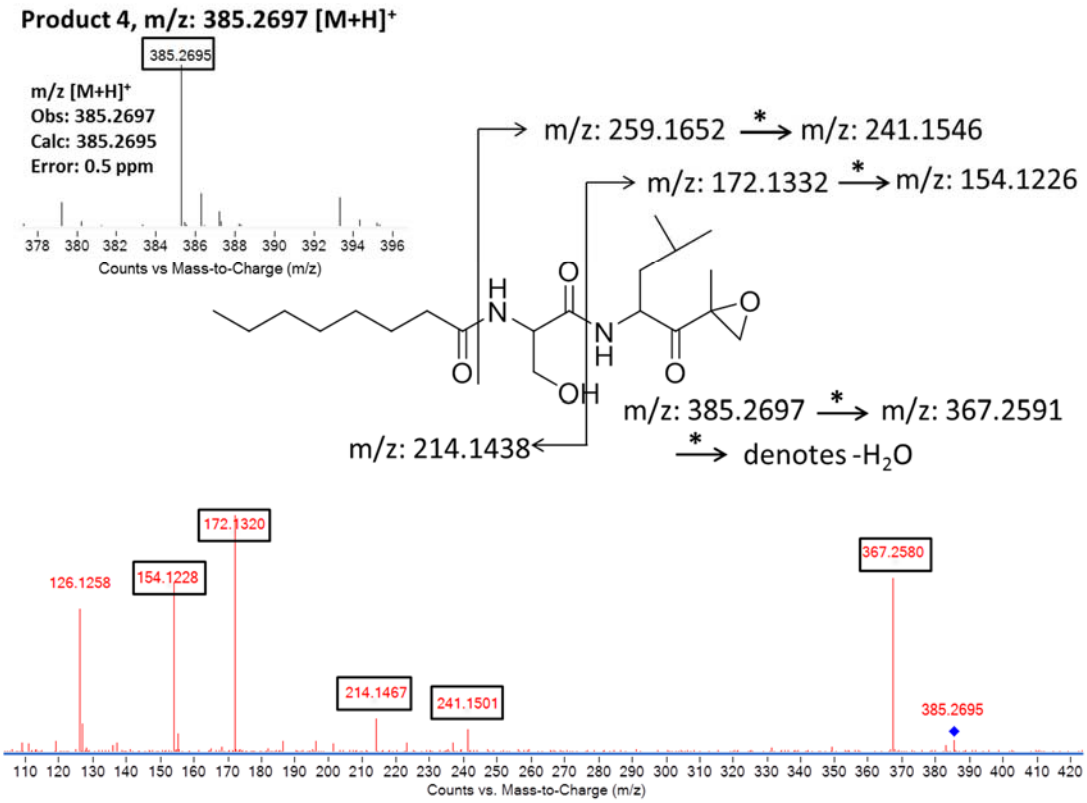

Figure S2. HRMS and HRMS/MS analysis of **4**.

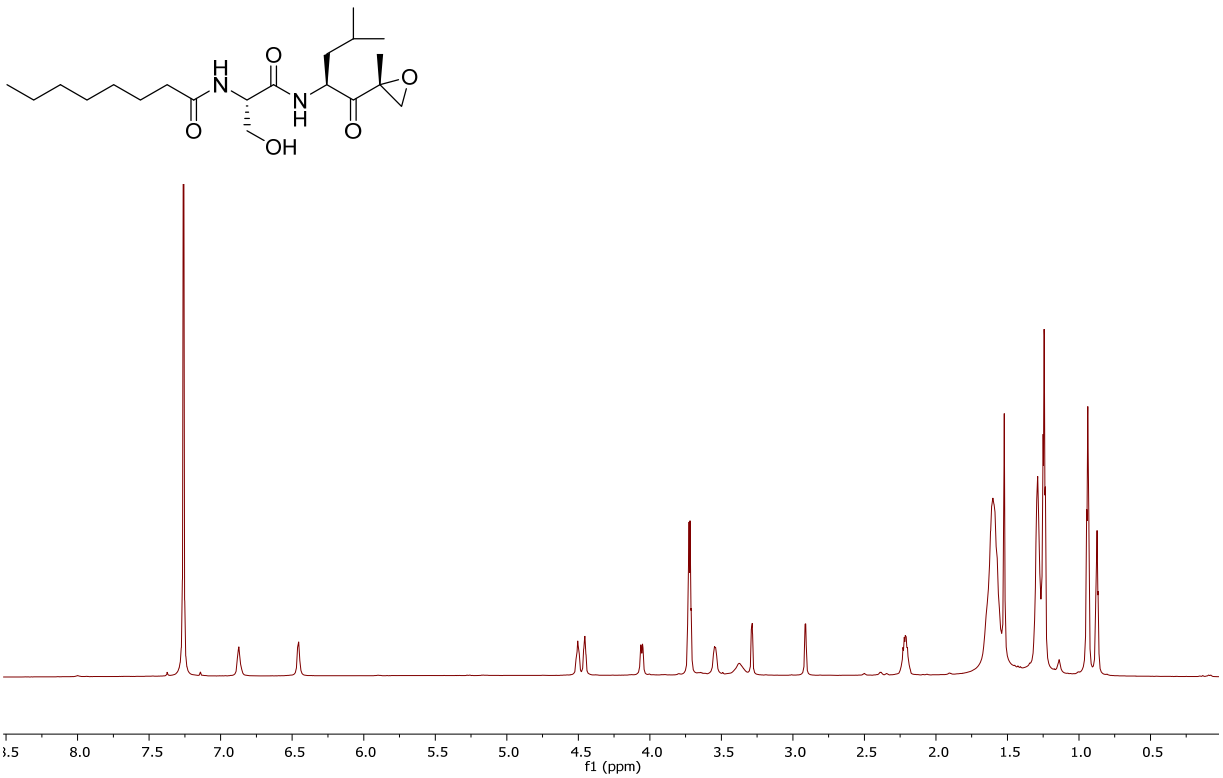

Figure S3. ^{1}H NMR spectrum of **4** ($CDCl_3$, 900 MHz).

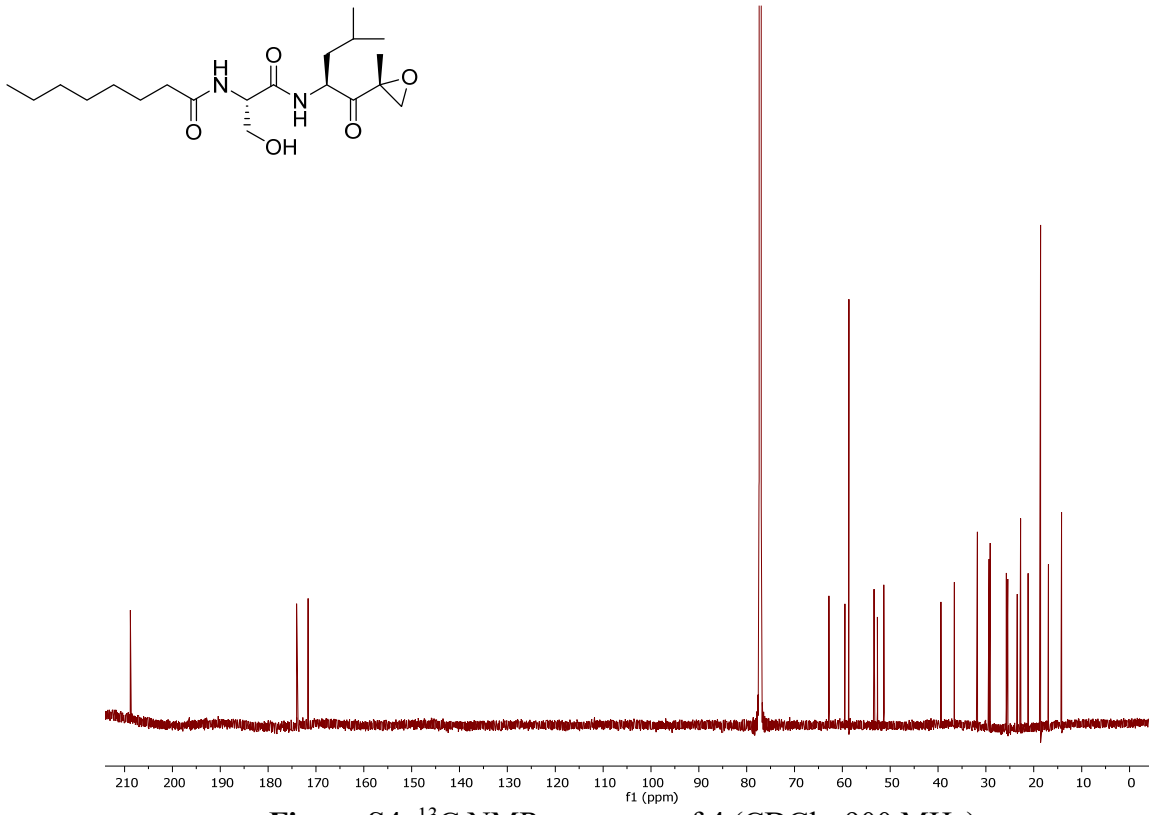

Figure S4. ^{13}C NMR spectrum of **4** ($CDCl_3$, 900 MHz).

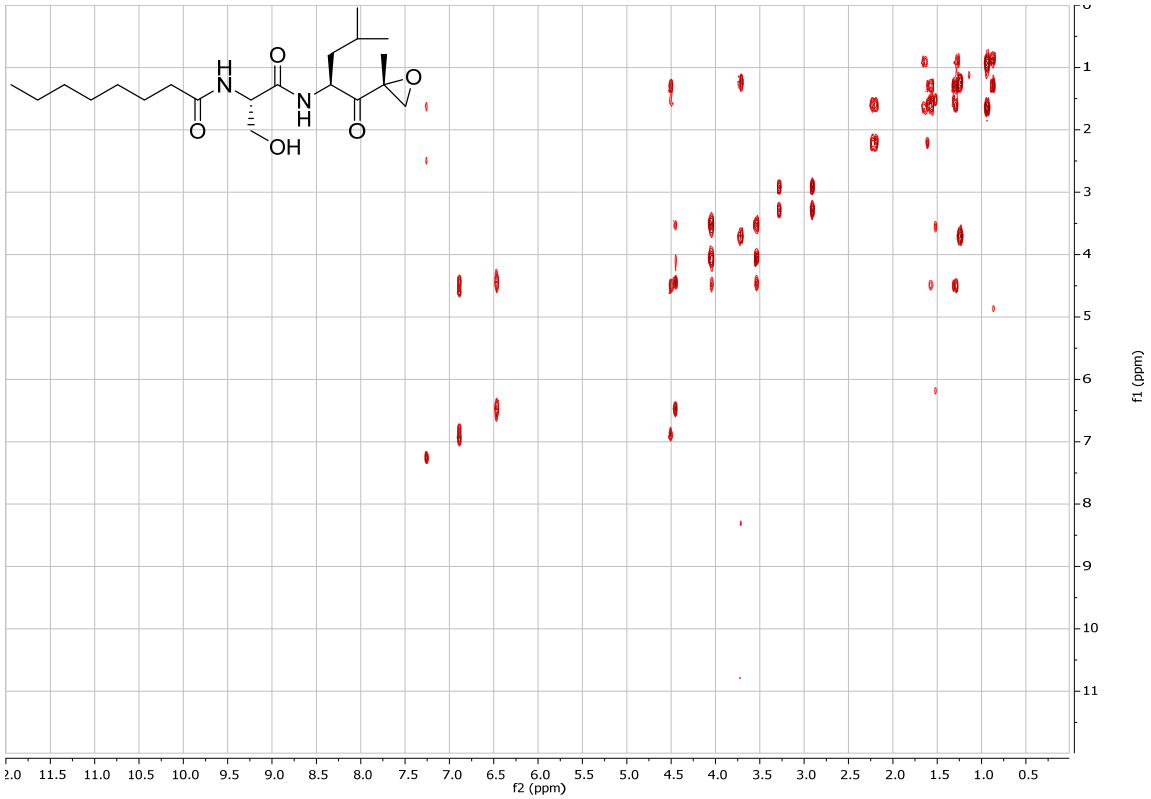

Figure S5. ^{1}H-^{1}H COSY spectrum of **4** ($CDCl_3$, 900 MHz).

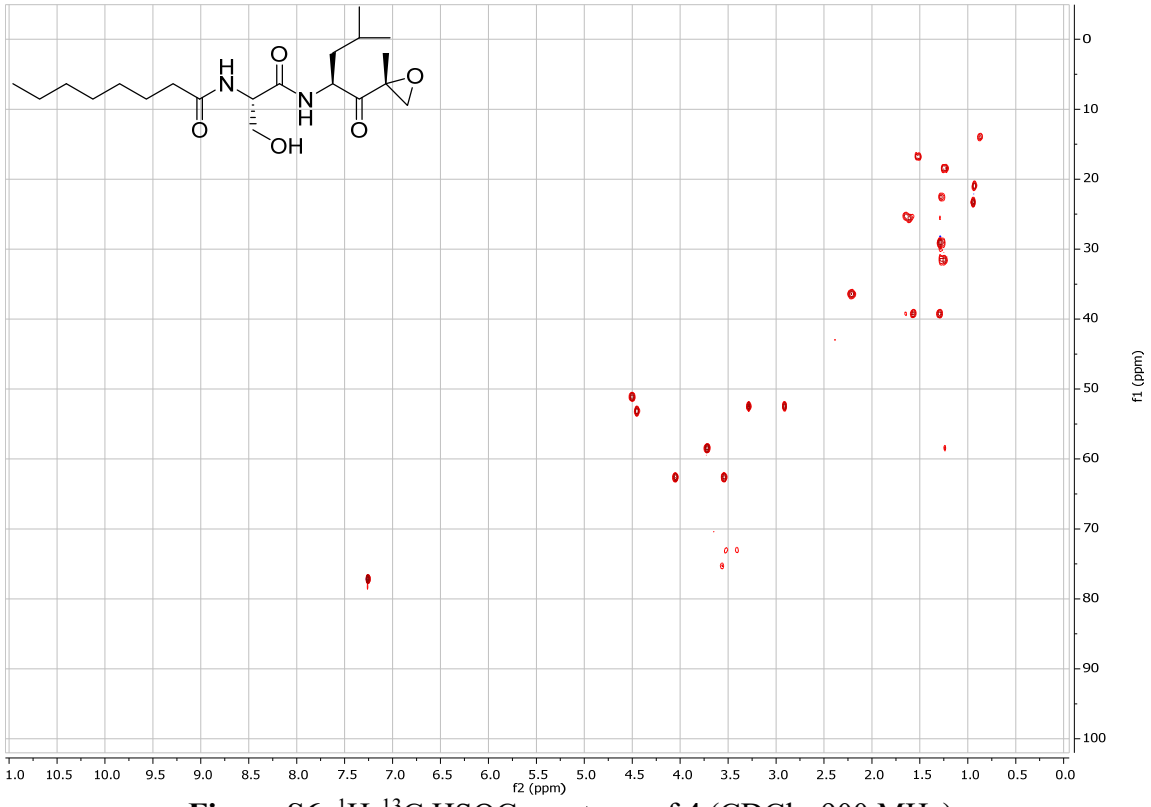

Figure S6. ^{1}H-^{13}C HSQC spectrum of **4** ($CDCl_3$, 900 MHz).

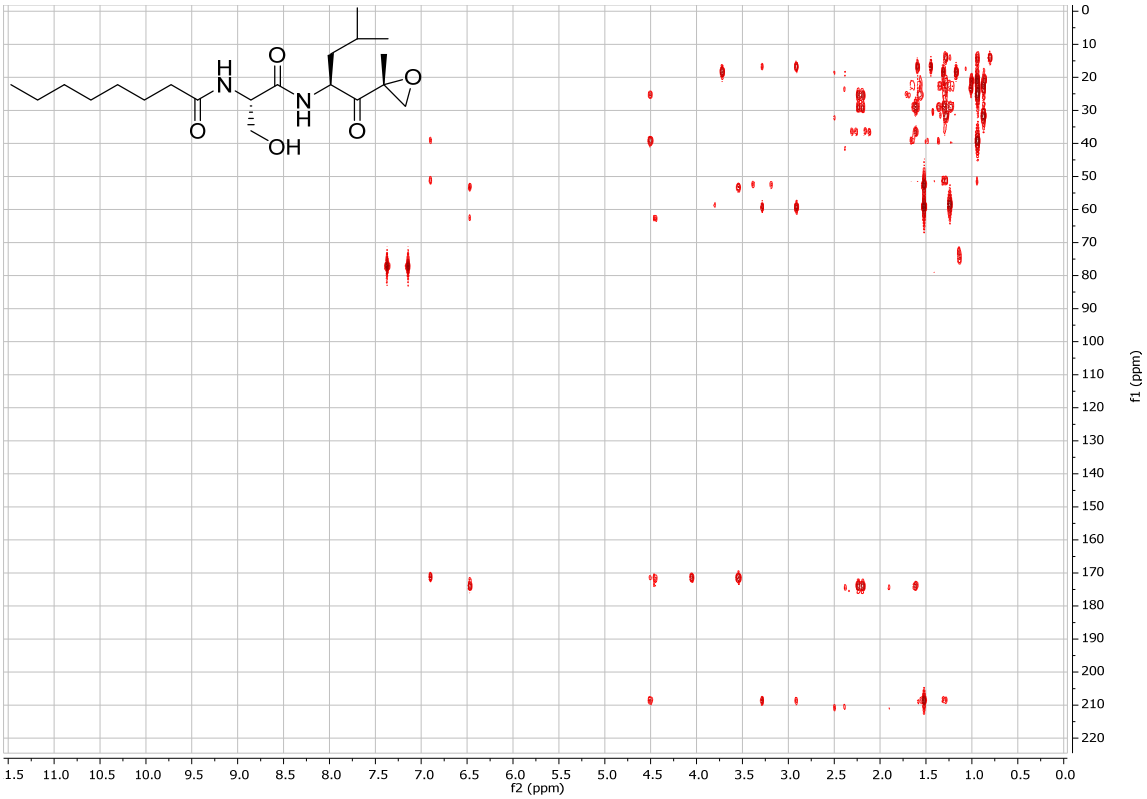

Figure S7. ^{1}H-^{13}C HMBC spectrum of **4** ($CDCl_3$, 900 MHz).

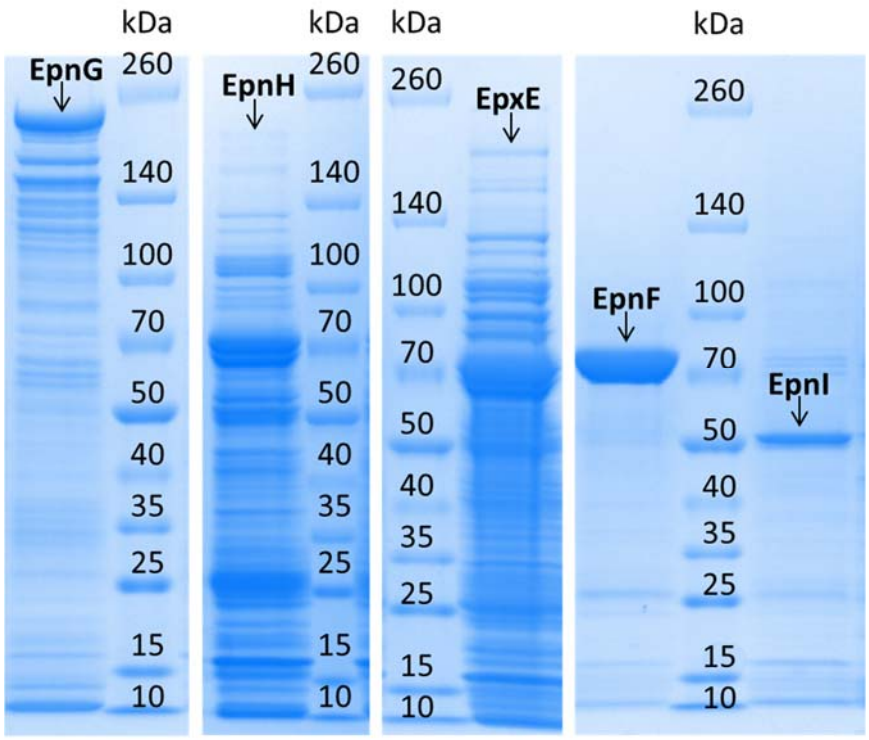

Figure S8. SDS-PAGE analysis of proteins purified from *E. coli*. EpnG, EpnF, and EpnI have *N*-terminal hexahistidine tags while EpnH and EpxE have *C*-terminal hexahistidine tags. Criterion Tris-HCl gels (4-15% precast, Biorad) were used.

Note: EpnH and EpxE are single module PKS's with a domain organization of KS-AT-cMT-ACP-TE (KS, ketosynthase; AT, acyltransferase, cMT, *C*-methyltransferase; ACP, acyl carrier protein; TE, thioesterase). Only a very small amount of these PKS megasynthases could be purified from *E. coli* BAP1 as shown here. We suspect that this PKS enzyme may have a short half-life in *E. coli* and degrades quickly upon expression, since attempts to reconstitute the production of **4** using the lysate of JL10 or JL11 and to radioactively label the PKS using a sensitive [2-^{14}C] malonyl-CoA or *S*-[methyl-^{14}C] adenosyl-L-methionine assay were both unsuccessful. Attempts to overexpress and purify these PKS's with different tags and from different *Streptomyces* hosts were also undertaken but were still unsuccessful. Nevertheless, the overexpression of EpnF and EpnI in *E. coli* both yielded soluble proteins that visually appeared to be bound to the correct cofactors, and purified EpnF was confirmed to contain FAD. However, without the ability to reconstitute the PKS activity *in vitro*, we were unable to further investigate the biochemical functions of these enzymes. Future efforts to obtain PKS enzymes that are functional *in vitro* may include the screening of additional PKS homologs such as those encoded by the clarepoxin and landepoxin biosynthetic gene clusters or the utilization of dissected EpnH/EpxE domains to overcome the problem of PKS megasynthase instability.

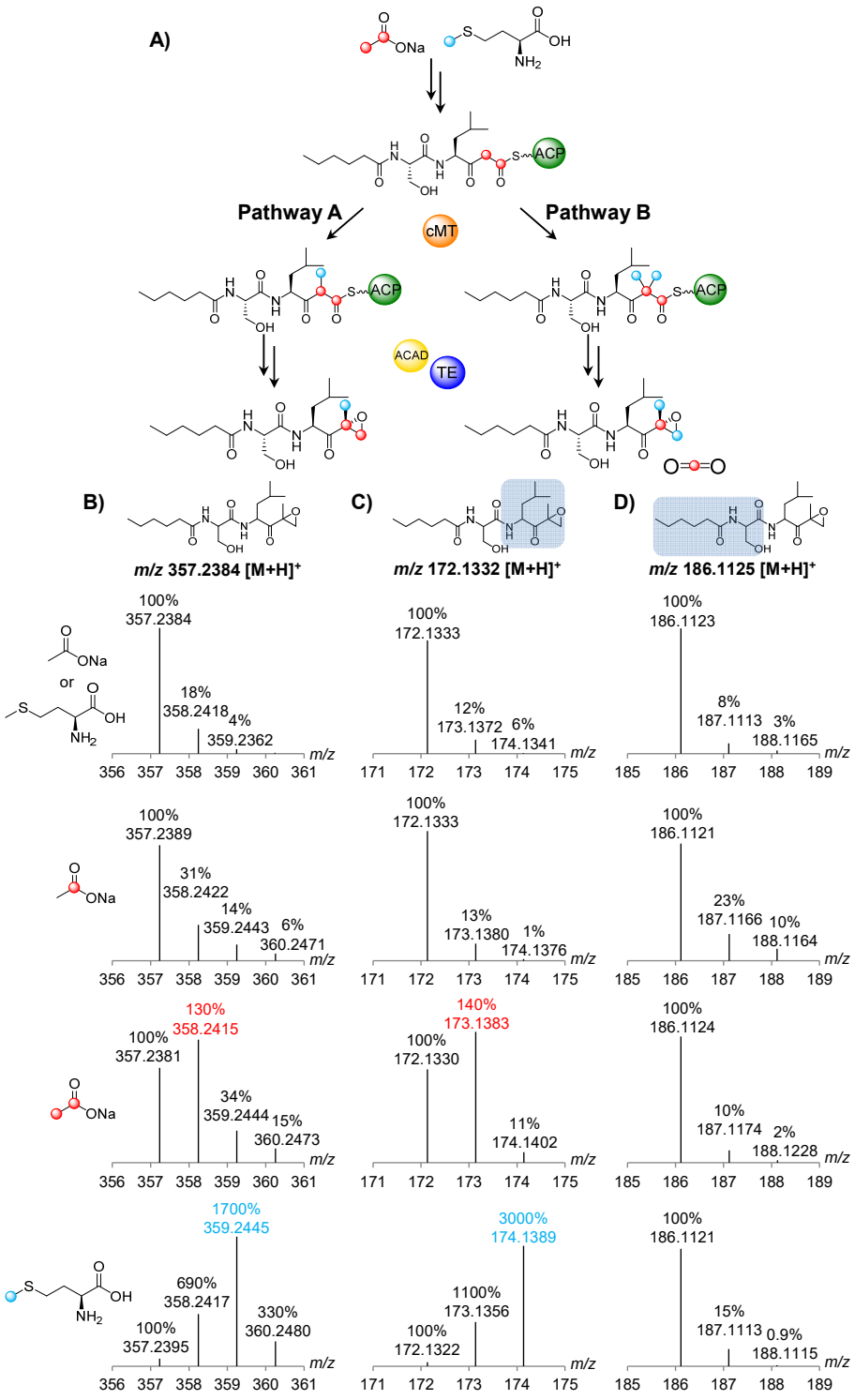

Figure S9. Isotopic peak analysis of **3** from ^{13}C-labeled precursor feeding studies. (A) Proposed routes for [1,2-^{13}C] sodium acetate and [methyl-^{13}C] L-methionine incorporation and terminal epoxyketone formation. Isotopic peak patterns of **3** (B), a fragment of **3** with the terminal epoxyketone group (C), and a fragment of **3** without the epoxyketone group (D) from cultures fed with unlabeled sodium acetate or unlabeled L-methionine (row 1), [1-^{13}C] sodium acetate (row 2), [1,2-^{13}C] sodium acetate (row 3), and [methyl-^{13}C] L-methionine (row 4) are also shown.

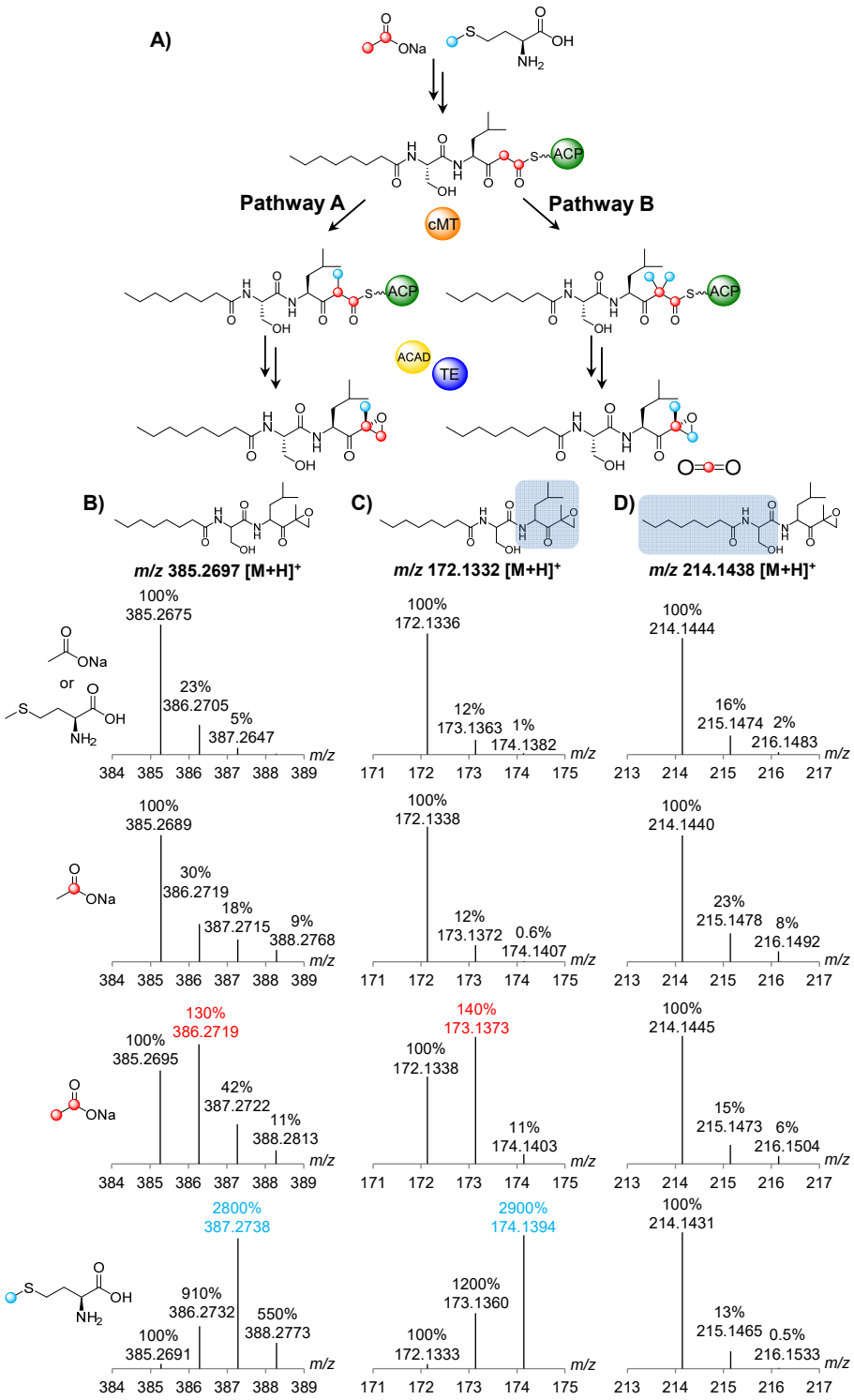

Figure S10. Isotopic peak analysis of **4** from ^{13}C-labeled precursor feeding studies. (A) Proposed routes for [1,2-^{13}C] sodium acetate and [methyl-^{13}C] L-methionine incorporation and terminal epoxyketone formation. Isotopic peak patterns of **4** (B), a fragment of **4** with the terminal epoxyketone group (C), and a fragment of **4** without the epoxyketone group (D) from cultures fed with unlabeled sodium acetate or unlabeled L-methionine (row 1), [1-^{13}C] sodium acetate (row 2), [1,2-^{13}C] sodium acetate (row 3), and [methyl-^{13}C] L-methionine (row 4) are also shown.