Supporting Information for:

Characterization of an orphan diterpenoid biosynthetic operon from Salinispora arenicola

Meimei Xu†, Matthew L. Hillwig†,‡, Amy L. Lane§,£, Mollie S. Tiernan‡, Bradley S. Moore§, Reuben J. Peters*†

† Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, IA 50011 USA
§ Scripps Institution of Oceanography and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093 USA

Content:
Figure S1. Amino acid sequence alignment of biochemically characterized bacterial class II diterpene cyclases.
Figure S2. Amino acid sequence alignment of bacterial class I diterpene synthases.
Figure S3. GC-MS extracted ion (m/z = 257) chromatograms demonstrating that SaCPS produces normal CPP (2) from GGPP (1).
Figure S4. Numbering and observed HMBC correlations for compounds 3 and 4.
Table S1. NMR Data (500 MHz, CDCl₃) for isopimara-8,15-diene (3).
Table S2. NMR Data (500 MHz, CDCl₃) for isopimara-8,15-dien-19-ol (4).
Figure S5. ¹H NMR spectrum for 3 (CDCl₃, 500 MHz).
Figure S6. ¹³C NMR spectrum for 3 (CDCl₃, 125 MHz).
Figure S7. GC-MS chromatograms and associated mass spectra demonstrating equivalent activity of CYP1051A1 and CYP99A3 with 3.
Figure S8. ¹H NMR spectrum for 4 (CDCl₃, 500 MHz).
Figure S9. ¹³C NMR spectrum for 4 (CDCl₃, 125 MHz).
Detailed Materials and Methods
References
Figure S1. Amino acid sequence alignment of biochemically characterized bacterial class II diterpene cyclases (DxDD motif is underlined). Shown are CPS from *Streptomyces* sp. strain KO-3988 (SsCPS, BAD826797),1 *Bradyrhizobium japonicum* (BjCPS, NP_768789),2 *Streptomyces platensis* (SsCPS, ACO31276),3 *Mesorhizobium loti* Sinorhizobium fredii (SfCPS, NP_443949),4 and *Xanthomonas oryzae* pv. oryzicola (XocCPS, YP_005626503),5 as well as that from *Salinispora arenicola* CNS-205 (SaCPS, YP_001536182) characterized here. Also shown is the terpentidienol diphosphate synthase from *Kitasatospora griseola* (KgTPS, Q9AJE4),6 and the halimadienyl diphosphate synthase from *Mycobacterium tuberculosis* (MtHPS, NP_217894).7
Figure S2. Amino acid sequence alignment of bacterial class I diterpene synthases (DDxxD and NDxx(S/T/G)xxx(D/E) motifs are underlined). Shown are the pimaradiene synthase from *Streptomyces* sp. strain KO-3988 (SsPS, BAD826798), 8 kaurene synthases from *Streptomyces platensis* (SsKS, ACO31279), 3 *Bradyrhizobium japonicum* (BjKS, NP_768790), 2 *Sinorhizobium fredii* (SIKS, NP_443948), 3 and *Xanthomonas oryzae* pv. oryzicola (XoKS, YP_005626502), 5 and the terpentediene synthase from *Kitasatospora griseola* (KgTS, Q9AJE3), 6 as well as the pimaradiene synthase from *Salinispora arenicola* CNS-205 characterized here (SaDTS, YP_001536181).

Figure S3. GC-MS extracted ion (m/z = 257) chromatograms demonstrating that SaCPS produces normal CPP (2) from GGPP (1). When co-expressed with the normal CPP specific AgAS:D404, which is incapable of producing but readily reacts with 2,9 the expected mixture of products are observed (compare to wild-type AgAS – 1, palustradiene; 2, levopimaradiene; 3, abietadiene; 4, neoabietadiene; 5 & 6, C13 epimers of abita-8(14)-en-13-ol, 10, 11 while upon co-expression with the ent-CPP specific AtKS, no production of ent-kaurene is observed.
Figure S4. Numbering and observed HMBC correlations (double headed arrows indicate reciprocal correlations) for 3 (A and B) and 4 (C and D).

Table S1: NMR Data (500 MHz, CDCl$_3$) for isopimara-8,15-diene (3)

| position | $\delta_C$, type | $\delta_H$, mult. (J in Hz) |
|----------|------------------|-----------------------------|
| 1 a      | 36.00, CH$_2$    | 0.997, m                    |
|          | b                | 1.712, m                    |
| 2 a      | 19.26, CH$_2$    | 1.573, m                    |
|          | b                | 1.656, m                    |
| 3 a      | 42.06, CH$_2$    | 1.143, m                    |
|          | b                | 1.385, m                    |
| 4        | 33.48, C         |                             |
| 5        | 52.07, CH        | 1.121, m                    |
| 6 a      | 19.16, CH$_2$    | 1.434, m                    |
|          | b                | 1.573, m                    |
| 7 a      | 32.76, CH$_2$    | 1.910, m                    |
|          | b                | 1.910, m                    |
| 8        | 124.45, C        |                             |
| 9        | 137.35, C        |                             |
| 10       | 37.78, C         |                             |
| 11 a     | 21.41, CH$_2$    | 1.895, br s                 |
|          | b                | 1.895, br s                 |
| 12 a     | 35.31, CH$_2$    | 1.309, m                    |
|          | b                | 1.488, m                    |
| 13       | 35.26, C         |                             |
| 14 a     | 42.29, CH$_2$    | 1.762, m                    |
|          | b                | 1.762, m                    |
| 15       | 146.60, CH       | 5.756, q (10.8)             |
| 16 a     | 110.85, CH$_2$   | 4.816, d (10.8)             |
|          | b                | 4.862, d (17.6)             |
| 17       | 28.08, CH$_3$    | 0.967, s                    |
| 18       | 33.55, CH$_3$    | 0.879, s                    |
| 19       | 21.95, CH$_3$    | 0.841, s                    |
| 20       | 19.69, CH$_3$    | 0.958, s                    |

Table S2: NMR Data (500 MHz, CDCl$_3$) for isopimara-8,15-dien-19-ol (4)

| position | $\delta_C$, type | $\delta_H$, mult. (J in Hz) |
|----------|------------------|-----------------------------|
| 1 a      | 35.60, CH$_2$    | 0.958, m                    |
|          | b                | 1.786, m                    |
| 2 a      | 18.90, CH$_2$    | 1.454, m                    |
|          | b                | 1.751, m                    |
| 3 a      | 36.82, CH$_2$    | 1.032, m                    |
|          | b                | 1.748, m                    |
| 4        | 33.09, CH$_2$    | 1.884, m                    |
|          | b                | 1.884, m                    |
| 5        | 52.83, CH        | 1.284, m                    |
| 6 a      | 19.39, CH$_2$    | 1.437, m                    |
|          | b                | 1.751, m                    |
| 7 a      | 35.19, CH$_2$    | 1.305, m                    |
|          | b                | 1.489, m                    |
| 8        | 21.46, CH$_2$    | 1.869, br s                 |
|          | b                | 1.869, br s                 |
| 9        | 35.19, CH$_2$    | 1.305, m                    |
|          | b                | 1.489, m                    |
| 10       | 35.26, C         |                             |
| 11 a     | 42.16, CH$_2$    | 1.767, m                    |
|          | b                | 1.767, m                    |
| 12 a     | 110.94, CH$_2$   | 4.827, d (10.8)             |
|          | b                | 4.887, d (17.6)             |
| 13       | 28.23, CH$_3$    | 0.968, s                    |
| 14 a     | 26.89, CH$_3$    | 0.989, s                    |
|          | b                | 3.782, d (11.0)             |
| 15       | 65.83, CH$_2$    | 3.477, d (11.0)             |
|          | b                | 3.782, d (11.0)             |

$^a$Configuration at C13 assigned by comparison of $^{13}$C shift values (particularly for C15 & 16) to those previously reported for the epimers. There are numerous minor differences (< 1 ppm) that may be attributable to the use of carbon tetrachloride as solvent in ref. 12. While there are significant differences in values for C6, 9 & 11, those reported here are supported by HMBC correlations (similar values also are found in 4).

$^b$Assignment supported by comparison of $^{13}$C shift values to those previously reported, although there are numerous minor differences (< 1 ppm), and several more significant differences (i.e., for C3, 4, 6 & 11). Again, the values reported here are supported by HMBC correlations.
Figure S5. $^1$H NMR spectrum for 3 (CDCl$_3$, 500 MHz).

Figure S6. $^{13}$C NMR spectrum for 3 (CDCl$_3$, 125 MHz).
**Figure S7.** GC-MS chromatograms and associated mass spectra demonstrating equivalent activity of CYP1051A1 and CYP99A3 with 3. Extracted (m/z = 257) chromatograms of extracts from strains engineered to produce 3 and co-expressing either A) CYP1051A1 (along with SaFdx and SaFdR), or B) CYP99A3 (along with a plant CYP reductase). Mass spectra of 4 from either C) CYP1051A1 or D) CYP99A3. Total ion chromatograms from separate injection of E) CYP1051A1 product (RT = 17.04 min.), F) co-injection (mixture of equal amounts; RT = 17.04 min.), and G) CYP99A3 product (RT = 17.02 min.).

**Figure S8.** $^1$H NMR spectrum for 4 (CDCl$_3$, 500 MHz).
Detailed Materials and Methods

The genes encoding the enzymes whose activity is described here were cloned from fosmids utilized in the *S. arenicola* CNS-205 genome sequencing effort (the *terp1* operon is on fosmid BPPW4701, Fdx on BPPW8023 and FdR on BPPW8886). Gene-specific primers were employed for PCR amplification, with the resulting gene fragments cloned into the Gateway vector system (Invitrogen) via directional topoisomerization into pENTR-SD-dTOPO, with the resulting constructs verified by complete gene sequence. For incorporation into the metabolic engineering system, SaCPS was sub-cloned into the pGG-DEST expression plasmid that carries a plant GGPS, \(^{14}\) and SaDTS into pDEST15 (Invitrogen), via directional recombination. For incorporation of the CYP (assigned as CYP1051A1 by Prof. David Nelson, Univ. Tenn), the SaFdx was first sub-cloned into the second multiple cloning site of a previously reported pCDFduet-DEST vector, using PCR amplification to introduce 5′ NdeI and 3′ XhoI restriction sites which were used for this purpose. CYP1051A1 was then sub-cloned via directional recombination into the resulting pCDFduet-DEST/SaFdx vector. SaFdR was similarly sub-cloned via directional recombination into a previously described pCOLAduet-DEST vector.

The recombinant strains were typically grown with vigorous shaking (200 rpm) in 50 mL TB media in 250 mL Erlenmeyer flasks at 37 °C to an optical density at 600 nm of ~0.6, the temperature lowered to 16 °C, the culture induced with 0.5 mM IPTG after an hour, and then further fermented for 72 hrs prior to extraction with an equal volume of hexanes. This organic extract was separated out and dried under a gentle stream of N\(_2\), and the residue resuspended in 200 µL fresh hexanes for GC-MS analysis, carried out using a Varian 3900 GC with Saturn 2100 ion trap mass spectrometer in electron ionization (70 eV) mode. Samples (1 µL) were injected in splitless mode at 50 °C and, after holding at 50 °C for 3 min., the oven temperature was raised at a rate of 14 °C/min. to 300 °C, where it was held for 3 min. MS data was collected starting after 12 min. until the end of the run. The observed enzymatic products were identified either by comparison to authentic samples from previous work, or by subsequent structural characterization by NMR.

**Figure S9.** \(^{13}\)C NMR spectrum for 4 (CDCl\(_3\), 125 MHz).
Analysis of SaCPS and SaDTS activity was carried out using recombinant expression in the C41 OverExpression strain of *E. coli* (Lucigen). SaCPS was investigated using expression pGG-DEST::SaCPS, first alone to verify activity, then in combination with stereospecific plant class I diterpene synthases, much as previously described. Briefly, SaCPS was co-expressed with either the *ent*-CPP specific kaurene synthase from *Arabidopsis thaliana* (AtKS), or normal CPP specific D404A mutant of the abietadiene synthase from *Abies grandis* (AgAS:D404A), with only the production of abietadienes (and not *ent*-kaurene) observed (Figure S3). SaDTS activity was investigated by co-transformation of pDEST15::SaDTS with pGG-DEST::SaCPS. For production of sufficient amounts of the resulting product for NMR analysis, pDEST15::SaDTS was co-transformed with pGGnC, which produces larger amounts of CPP than pGG-DEST::SaCPS, in a 1 L culture in 2.8 L fernbach flask, otherwise grown as described above. This culture (media and cells) was extracted with an equal volume of hexanes and the resulting organic extract dried in a rotary evaporator, the residue resuspended in 10 mL fresh hexanes, and purified by passage over a silica gel column.

Analysis of CYP1051A1 activity was carried out using recombinant expression in the C43 OverExpress strain of *E. coli* (Lucigen). First, pCDFduet-DEST::CYP1051A1/SaFdx was co-transformed with pCOLAduet-DEST::SaFdr, as well as pGGnC and pDEST15::SaDTS, and the resulting recombinant strain found to produce an oxygenated derivative of 3 (i.e., 4). While efficient flux to 4 was observed with the original small-scale cultures (Figure 2C), this was compromised in larger cultures (i.e., 1 L media in 2.8 L fernbach flasks), as demonstrated by attempted purification from 5 L of culture (as described below), which yielded only ~0.1 mg. Fortuitously, we discovered that the plant microsomal CYP99A3 efficiently catalyzes the same reaction as CYP1051A1 (Figure S7). Specifically, using co-expression of CYP99A3 with the requisite CYP reductase (CPR) from the previously described pCDFduet-DEST::CYP99A3/AtCPR1. A recombinant strain co-transformed with this vector, pGGnC and pDEST15::SaDTS could be grown in larger scale cultures with retention of reasonable flux to 4, which was then extracted from 3 x 1 L cultures (grown as described above) with an equal volume of hexanes. The pooled organic extract was dried in a rotary evaporator and resuspended in 10 mL of hexanes and passed over a silica gel column (10 mL), which was then washed with 50 mL of hexanes, with 4 eluted by 20 mL of 5% ethyl acetate (in hexanes) and 20 mL 10% ethyl acetate (in hexanes). These fractions were dried separated under a gentle stream of N₂, and each resuspended in 1 mL of methanol. 4 was then purified from each of these fractions by HPLC carried out with an Agilent 1100 series instrument equipped with fraction collector and diode array detector using a ZORBAX Eclipse XDB-C8 column (4.6 x 150 mm, 5 µm) at a flow rate of 0.5 mL/min. After loading the column was washed with 20% acetonitrile/dH₂O (0-2 min), eluted with a gradient of 20-100% acetonitrile (2-7 min), followed by a 100% acetonitrile wash (7-27 min). Fractions containing 4, as determined by GC-MS analysis, were pooled for NMR analysis.

The two purified products were each dried under a gentle stream of N₂ and resuspended in 0.5 mL deuterated chloroform, with this evaporation-resuspension process repeated twice more to completely remove the protonated hexanes solvent. Each resulting sample (~2 mg of 3 or ~1 mg of 4) was then analyzed by NMR at 25 °C using a Bruker Avance 500 spectrometer equipped with cryogenic probe for ¹H and ¹³C. Structural analysis was carried out using standard proton, triple quantum filtered-COSY,
were evaluated for metabolites matching the retention time and fragmentation pattern of extracts were evaluated for terpenes using a Thermo Traceplus GC-MS system. Each culture was extracted twice via liquid partitioning using an equal volume of hexanes or ethyl acetate, and samples were concentrated in vacuo. Chemical extracts were evaluated for terpenes using a Thermo Traceplus GC-MS system. Extracts were evaluated for metabolites matching the retention time and fragmentation pattern of enzymatically prepared 3 and 4. In addition, GC-MS metabolite profiles were compared.
between chemical extracts from wild type and mutant *S. arenicola*. Chemical profiles of mutant and wild type chemical extracts were also compared by low resolution LC-MS using an Agilent 1100 series LC-MS system with diode array detector.

References:
(1) Kawasaki, T.; Kuzuyama, T.; Kuwamori, Y.; Matsuura, N.; Itoh, N.; Furihata, K.; Seto, H.; Dairi, T., *J. Antibioto*. 2004, 57, 739-747.
(2) Morrone, D.; Chambers, J.; Lowry, L.; Kim, G.; Anterola, A.; Bender, K.; Peters, R. J., *FEBS Lett.* 2009, 583, 475-480.
(3) Smanski, M. J.; Yu, Z.; Casper, J.; Lin, S.; Peterson, R. M.; Chen, Y.; Wendt-Pienkowski, E.; Rajsik, S. R.; Shen, B., *Proc Natl Acad Sci U S A* 2011, 108, 13498-503.
(4) Hershey, D. M.; Lu, X.; Zi, J.; Peters, R. J., *J. Bact.* 2014, 196, 100-106.
(5) Lu, X.; Hershey, D. M.; Wang, L.; Bogdanove, A. J.; Peters, R. J., *New Phytol* submitted.
(6) Dairi, T.; Hamano, Y.; Kuzuyama, Y.; Itoh, N.; Furihata, K.; Seto, H., *J. Bact.* 2001, 183, 6085-6092.
(7) Nakano, C.; Okamura, T.; Sato, T.; Dairi, T.; Hoshino, T., *Chem. Commun.* 2005, 2005, 1016-1018.
(8) Kawasaki, T.; Hayashi, Y.; Kuzuyama, T.; Furihata, K.; Itoh, N.; Seto, H.; Dairi, T., *J. Bacteriol.* 2006, 188, 1236-1244.
(9) Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B., *J. Am. Chem. Soc.* 2001, 123, 8974-8978.
(10) Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B., *Biochemistry* 2000, 39, 15592-15602.
(11) Keeling, C. I.; Madilao, L. L.; Zerbe, P.; Dullat, H. K.; Bohlmann, J., *J. Biol. Chem.* 2011, 286, 21145-53.
(12) Wenkert, E.; Buckwalter, B. L., *J. Am. Chem. Soc.* 1972, 94, 4367-4369.
(13) Ceccherelli, P.; Curini, M.; Marcotullio, M. C., *J. Chem. Soc., Perkin Trans. I* 1985, 2173-2175.
(14) Morrone, D.; Lowry, L.; Determan, M. K.; Hershey, D. M.; Xu, M.; Peters, R. J., *Appl. Microbiol. Biotechnol.* 2010, 85, 1893-1906.
(15) Wu, Y.; Zhou, K.; Toyomasu, T.; Sugawara, C.; Oku, M.; Abe, S.; Usui, M.; Mitsuhashi, W.; Chono, M.; Chandler, P. M.; Peters, R. J., *Phytochemistry* 2012, 84, 40-46.
(16) Cyr, A.; Wilderman, P. R.; Determan, M.; Peters, R. J., *J. Am. Chem. Soc.* 2007, 129, 6684-6685.
(17) Wang, Q.; Hillwig, M. L.; Peters, R. J., *Plant J.* 2011, 65, 87-95.
(18) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F., *Proc Natl Acad Sci U S A* 2003, 100, 1541-6.
(19) Eustaquio, A. S.; Pojer, F.; Noel, J. P.; Moore, B. S., *Nat Chem Biol* 2008, 4, 69-74.