Supporting Information

Genome Mining in *Streptomyces coelicolor*. Molecular Cloning and Characterization of a New Sesquiterpene Synthase

*Xin Lin, Russell Hopson, and David E. Cane**

[a]Department of Chemistry, Brown University, Box H, Providence, RI 02912-9108, USA

e-mail: David_Cane@brown.edu

Sequence alignment. SCO5222 vs Pentalenene Synthase. SC7E4.19, possible lyase, len: 361 aa; similar to SW:PTLS_STRS3 (EMBL:Q55012) Streptomyces sp. pentalenene synthase (EC 4.6.1.5), 336 aa; fasta scores: opt: 254 z-score: 317.5 E(): 3.3e-10; 23.8% identity in 311 aa overlap. (http://www.sanger.ac.uk/Projects/S_coelicolor/SCO_html/SCO5222.html)

```
tr|Q9K499|Q9K499_STRCO      MHAFPHGTTATPTAIAVPPSLRLPVIEAAFPRQLHPYWPKLQETTRTWLL 50
sp|Q55012|PTLS_STRS3        ---------PQDVDFHIP-----------LPGRQSPDHARAEAEQLAWPR 30
                            :  :     .  . : :*  : : ::   :* :  *  .: :    :*

tr|Q9K499|Q9K499_STRCO      EKRLMPADKVEEYADGLCYTDLMAGYYLGAPDEVLQAIADYSAWFFVWDD 100
sp|Q55012|PTLS_STRS3        SLGLIRSDAAAERHLRGGYADLASRFYPHATGADLDLGVDLMSWFFLFDD 80
                            .  *: :* . *      *:** : :*  *..  *:  .*  :***::**

tr|Q9K499|Q9K499_STRCO      RHDRDIVHGRAGAWRRLRGLLHTALDSPGDHLHHEDTLVAGFADSVRRLY 150
sp|Q55012|PTLS_STRS3        LFDGPRGE-NPEDTKQLTDQVAAALDGP--LPDTAPPIAHGFADIWRRTC 127
                             .*    . ..   ::* . : :***.*    .   .:. ****  **

tr|Q9K499|Q9K499_STRCO      AFLPATWNARFARHFHTVIEAYDREFHNRTR-GIVPGVEEYLELRRLTFA 199
sp|Q55012|PTLS_STRS3        EGMTPAWCARSARHWRNYFDGYVDEAESRFWNAPCDSAAQYLAMRRHTIG 177
                              :..:* ** ***::. ::.*  * ..*   .   .. :** :** *:.

tr|Q9K499|Q9K499_STRCO      HWIWTDLLEPSSGCELPDAVRKHPAYRRAALLSQEFAAWYNDLCSLPKEI 249
sp|Q55012|PTLS_STRS3        VQPTVDLAERAGRFEVPHRVFDSAVMSAMLQIAVDVNLLLNDIASLEKEE 227
                                .** * :.  *:*. * . ..      :: :.    **:.** **

tr|Q9K499|Q9K499_STRCO      AGDEVHNLGISLITHHSLTLEEAIGEVRRRVEECITEFLAVERDALRFAD 299
sp|Q55012|PTLS_STRS3        ARGEQNNMVMILRREHGWSKSRSVSHMQNEVRARLEQYLLLESCLPKVGE 277
                            * .* :*: : *  .*. : ..::..::..*.  : ::* :*    :..:

tr|Q9K499|Q9K499_STRCO      ELADGTVRGKELSGAVRANVGNMRNWFSSVYWFHHESGRYMVDSWDDRST 349
sp|Q55012|PTLS_STRS3        IYQLDTAERERALE---RYRTDAVRTVIRGSYDWHRSSGRYDAEFALAAGA 324
                                .*.. : *.  :* ... :*. : . * :*:.**** .:     .:

tr|Q9K499|Q9K499_STRCO      PPYVNNEAAGEK 361
sp|Q55012|PTLS_STRS3        QGYLEELGSSAH 336
                              *::: .:. :
```

Figure S1. CLUSTAL W (1.83) multiple sequence alignment (http://www.ebi.ac.uk/clustalw/)[1]

Materials and Methods. General materials and methods were as described.[2]

Expression and Purification of *Streptomyces coelicolor* epi-isozizaene synthase (SCO5222p). Cosmid SC7E4 from *S. coelicolor* A3(2) was a gift from Professor Keith Chater of the John Innes Centre. PCR was used to amplify SCO5222 using the forward (5′-GGT**CATATG**GTGCATGCTTTCCCACACGGC-3′) and reverse (5′-CGGACT**CTCGAG**TCATTTCTCACCTGCCGCTTC-3′) primers to introduce *Nde*I and *Xho*I restriction sites (**bold**) flanking the normal start and stop codons, respectively. The double digested PCR product was the ligated into *Nde*I/*Xho*I digested pET-28a(+) with T4 DNA ligase using a 30:1 molar ratio of insert:vector (16 °C, 16 h) and the ligation mixture was transformed into competent cells of *Escherichia coli* XL1_Blue by standard procedures. The resulting pET28/SCO5220 plasmids, which were screened by appropriate restriction digests and verified by DNA sequencing, were transformed into expression strain *E. coli* BL21(DE3). A 1-L LB-kanamycin (50 μg/mL) culture of *E. coli* BL21(DE3)/SCO5222 was grown at 37 °C to an OD_{600} of 0.5, then induced with 0.1 mM IPTG at 20 °C for 18 h. After two passages through a French-Press (10,000 psi), the cell lysate was centrifuged at 10400 rpm (12930*g*) for 30 min to remove cell debris. The supernatant was subsequently loaded to Ni-NTA column, and the N-terminal His6-tag protein was eluted with buffer consisting of 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 70 mM imidazole, 20% (vol/vol) glycerol, and 5 mM 2-mercaptoethanol. The eluent containing SCO5222 protein were concentrated by YM-10 centriprep and were dialyzed against kinetics buffer (50 mM PIPES, pH 6.5, 20% (vol/vol) glycerol, 10 mM $MgCl_2$, 100 mM NaCl, 5 mM 2-mercaptoethanol) using a PD-10 desalting column. Typical yield is 13 mg recombinant protein per L culture.

Epi-isozizaene synthase assay. To test the metal cofactor requirement, SCO5222 cyclase was assayed in 1 mL modified assay buffer (50 mM PIPES, pH 6.5, 20% (vol/vol) glycerol plus 10 mM divalent cation as the chloride salt) as previous described at 30 °C.[2] Different metal cations tested were Mg^{2+}, Mn^{2+}, Co^{2+}, Fe^{2+}, Fe^{3+}, Cu^{2+}, Ni^{2+} and Zn^{2+}. In kinetic studies, 100 mM NaCl and 5 mM 2-

mercaptoethanol were added to the assay buffer and the product was extracted by hexanes. Assays were performed with FPP (**1**) (164 mCi/mmol) at concentrations of 83, 166, 332, 498, 1245, and 2490 nM at 30 °C for 15 min. The reaction was quenched by the addition of 75 μL of 500 mM EDTA (pH 8.0) and the mixture was vortexed for 30 s. The organic layer was loaded onto a silica gel column (1 cm) in a Pasteur pipette and expelled with an N_2 stream into a scintillation vial containing 7 mL of Opti-Fluor. The samples were extracted with a further two 1-mL portions of hexanes. After passage of the extracts through the same minicolumn, the silica gel was finally washed with 750 μL of hexanes. The steady-state kinetic parameters, k_{cat} and K_m, were calculated by fitting the liquid scintillation data to the Michaelis-Menten equation. In small scale incubations for GC-MS analysis, HPLC-grade pentane and ether were used for extraction respectively, and both extracts displayed a > 95% pure single peak in GC-MS with an $[M]^+$ of *m/z* 204 and a base peak of 119.[5] In a typical preparative-scale incubation, a total 0.6 μmol of recombinant enzyme was incubated with 12.6 mg of **1** in 500 mL of kinetic buffer with 5 mM $MgCl_2$ for 18 h at 30 °C. The enzyme was added in three equal portions every 6 hours and reaction mixture was extracted with HPLC-grade pentane.

Analysis of volatile extracts of *S. coelicolor* M145. An attempt was made to detect **2** from liquid or agar solid cultures of *S. coelicolor* M145. For solid media extraction, *S. coelicolor* mycelia were grown directly on wet dialysis membranes overlaid on SFM-agar plates.[3] The mycelia were therefore isolated from media by peeling off whole membranes. The extraction was performed as previously reported and geosmin, germacrene D and germacradienol were all detected by capillary GC-MS.[4] Dialysis membranes were submerged in tap water. After they were completely wet, they were sandwiched between wet Whatman filter papers and wrapped in aluminum foil. The whole package was autoclaved using a liquid cycle. The sterilized membranes were overlaid flat on 25-mL SFM-agar plates and 20 μL *S. coelicolor* M145 spore suspension was deposited in the center of the membranes. 13 plates were grown at 28 °C for ~11 days.

NMR – General. NMR spectra were obtained on Bruker Avance NMR spectrometers operating at 300, 400, and 600 MHz ^{1}H frequency. Chemical shifts are referenced to $CDCl_3$ at room temperature.

Epi-isozizaene. NMR assignments. ^{1}H NMR ($CDCl_3$, 300.15 MHz) δ 2.2 (ddq, *J* =9.21, 17.07, 1.35 Hz, 1 H, H-5a), 2.06 (dddq, 1 H, H-5b), 1.81 (m, 1 H, H-1), 1.77 (m, 2 H, H-6b, H-7), 1.73 (m. 1 H, H-10a), 1.56 (m, 1 H, H-10b), 1.47 (dd, *J*= 5.1, 10.5 Hz, 1 H. H-11$_{anti}$), 1.42 (bs, 3 H, H-14), 1.39 (d, *J* = 10.5 Hz, 1 H, H-11$_{syn}$), 1.35 (m, 1 H, H-9b), 1.22 (m, 1 H, H-6a), 1.16 (m, 1 H, H-9a), 0.99 (s, 3 H, H-12), 0.96 (s, 3 H, H-13), 0.90 (d, *J* = 6.3 Hz, 3 H, H-15); ^{13}C NMR ($CDCl_3$, 75.48 MHz, ppm): 143.0 (C, C-4), 127.4 (C, C-3), 52.6 (C, C-8), 47.1 (CH, C-1), 40.4 (C, C-2), 39.7 (CH, C-7), 36.9 (CH_2, C-11), 32.5 (CH_2, C-6), 28.6 (CH_2, C-9), 28.3 (CH_3, C-12), 27.3 (CH_2, C-5), 25.0 (CH_3, C-13), 24.3 (CH_2, C-10), 14.0 (CH_3, C-15), 12.8 (CH_3, C-14).

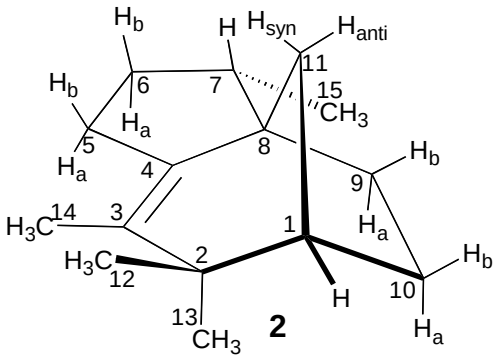
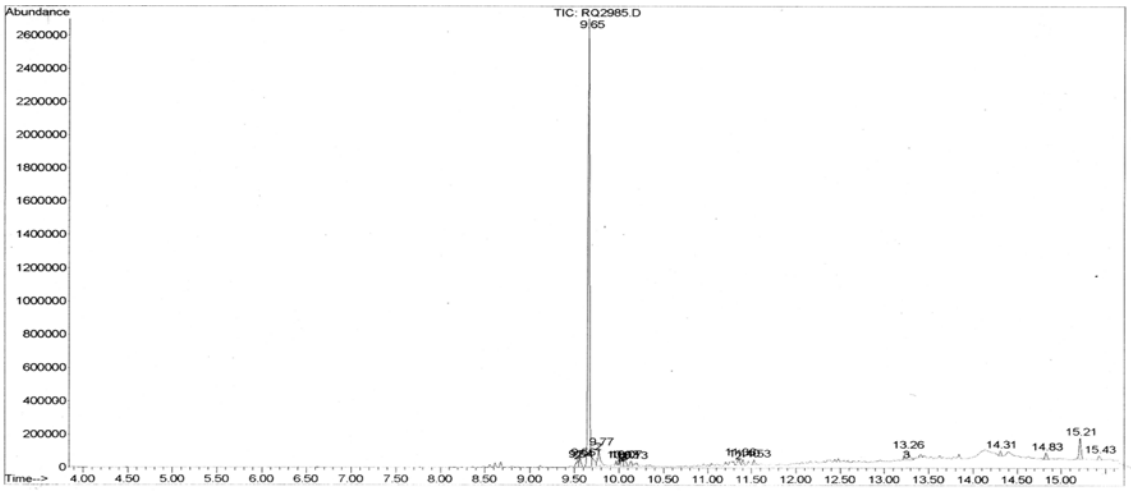


Figure S2. Epi-isozizaene (**2**).

Comparison of enzymatically-generated 2 with synthetic epi-isozizaene and isozizaene. The 300 MHz ^{1}H NMR spectrum of enzymatically-generated **2** was recorded in CCl_4 and the observed chemical shifts of the 4 methyl groups were used to simulate the 60 MHz spectrum using MestReC NMR software (Mestrelab Research, www.mestrec.com). The chemical shifts for the 4 methyl groups (δ 0.937 (d, *J*=6.6 Hz, H-15), 0.986 (s, H-13), 1.011 (s, H-12), 1.427 (bs, H-14) closely matched the original 60 MHz spectrum of synthetic epi-isozizaene provided by Prof. Niels H. Andersen: (δ 0.925 (d, *J*=6.5 Hz, H-15), 0.975 (s, H-13), 1.017 (s, H-12), 1.41 (bs, H-14). Synthetic isozizaene: δ 0.825 (d, *J*=6.6 Hz, H-15), 0.96 (s, H-13), 0.98 (s, H-12), 1.43 (bs, H-14).

A



B

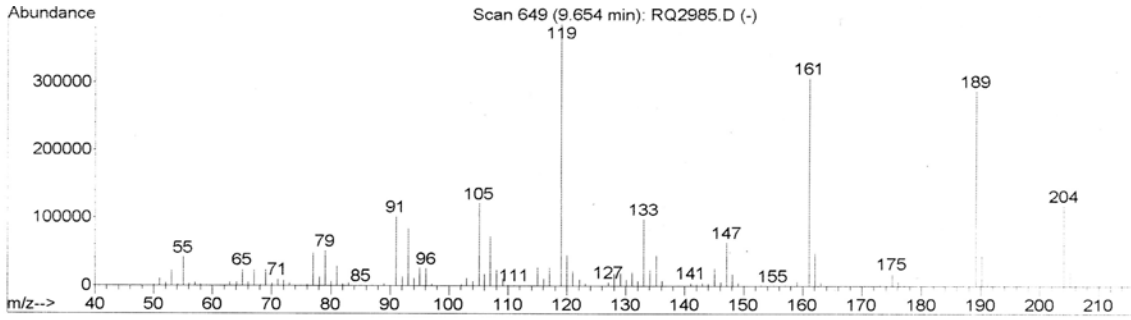


Figure S3. GC/MS spectra of **2** generated from farnesyl diphosphate (FPP, **1**) by epi-isozizanene synthase (SCO5222p). Agilent 6890 GC/JEOL JMS-600H mass spectrometer, using a 30 m x 0.25 mm HP5MS capillary column in EI (positive) mode using a temperature program of 60-280 °C, with a gradient of 20 °C/min and a solvent delay of 3.5 min. A. GC/MS TIC. B. EI-Mass spectrum of **2**.

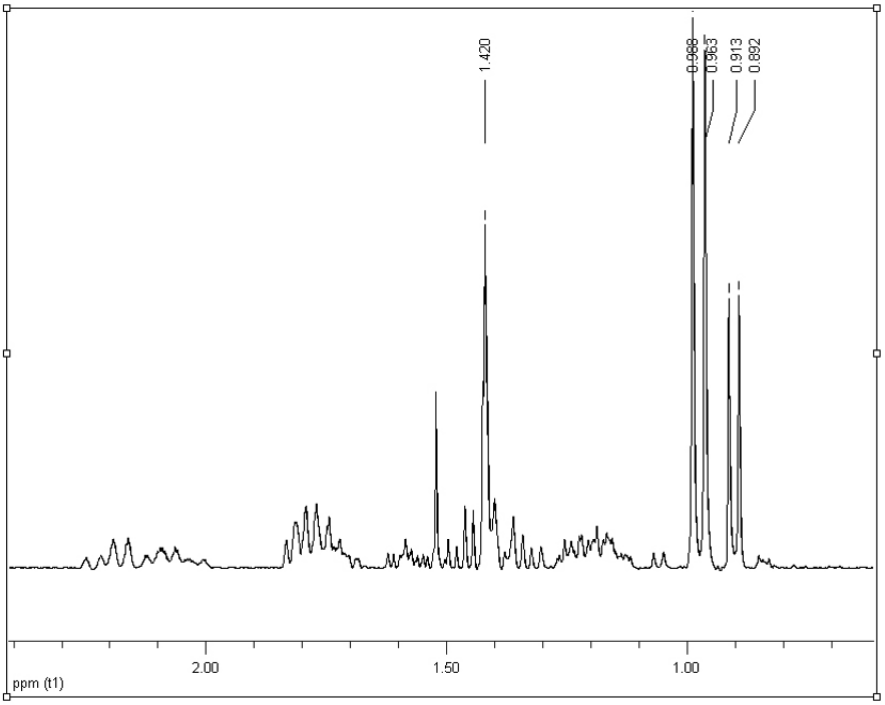


Figure S4. ^{1}H NMR ($CDCl_3$, 300.15 MHz) spectrum of **2** (Peak at δ 1.52 is water.)

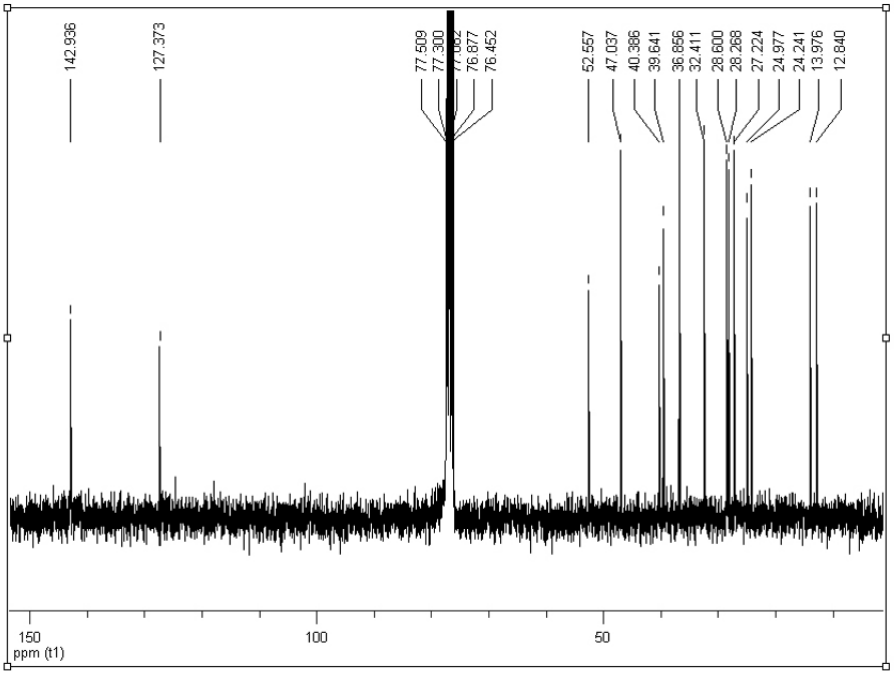


Figure S5. ^{13}C NMR ($CDCl_3$, 75.48 MHz) spectrum of **2**.

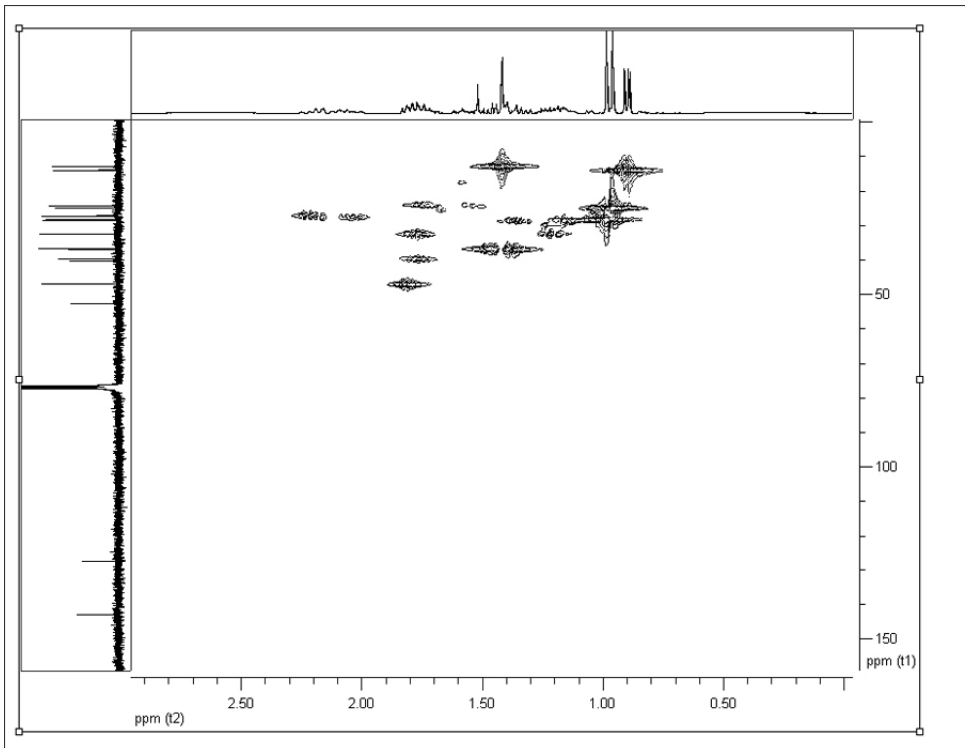


Figure S6. HMQC NMR (400.13 MHz, 100.62 MHz) spectrum of **2**.

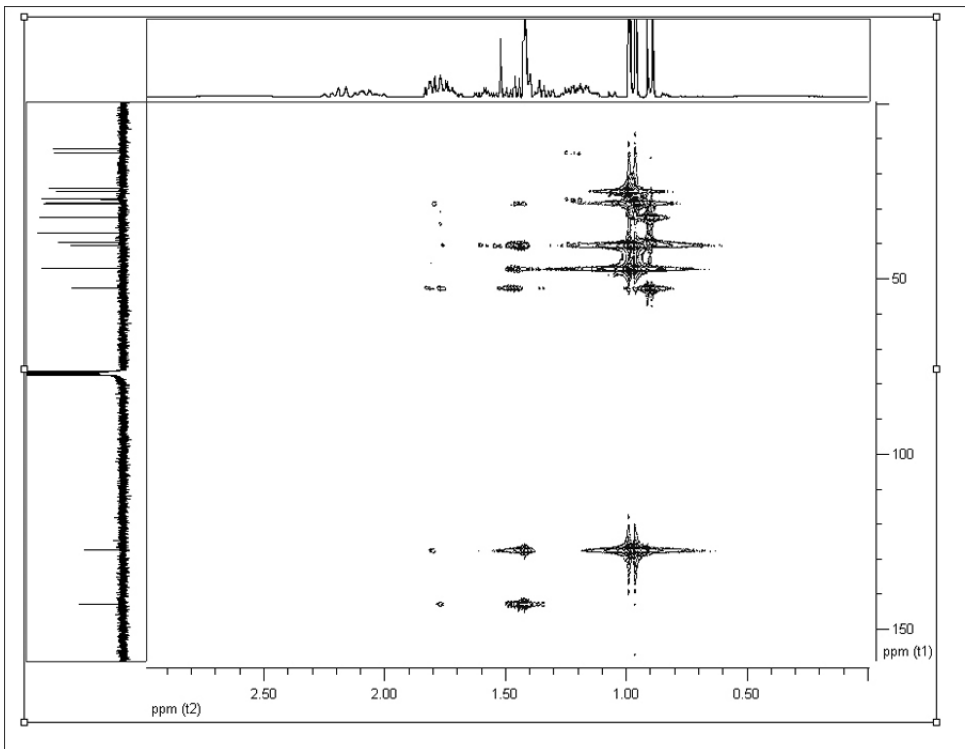


Figure S7. HMBC NMR (400.13 MHz, 100.62 MHz) spectrum of **2**. CIGAR-HMBC parameters: J(XH) min CNST6 – 130Hz; J(XH) max CNST7 – 160Hz; J(XH) long range (min) CNST14 – 6Hz; J(XH) long range (max) CNST 15 – 12Hz.

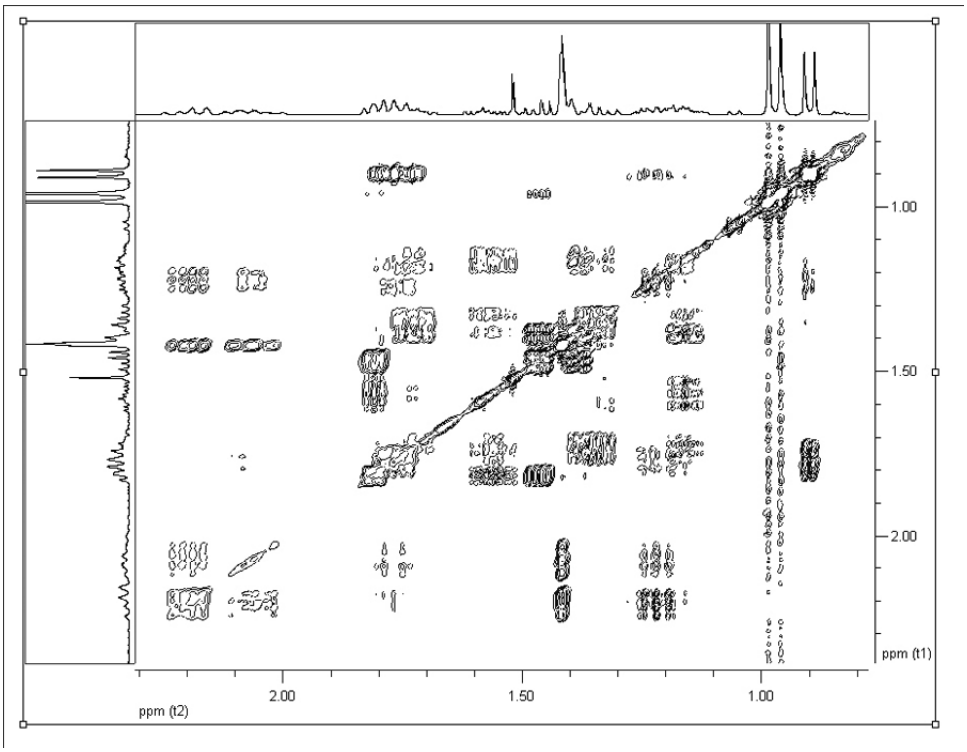


Figure S8. ^{1}H-^{1}H COSY NMR (400.13 MHz) spectrum of **2**.

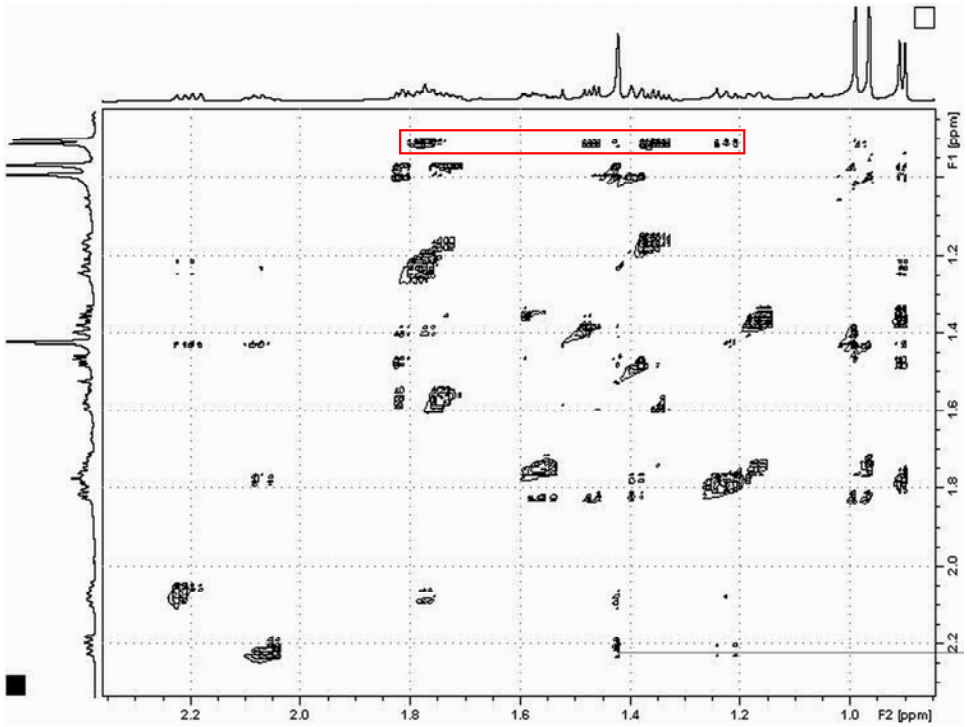


Figure S9. 600 MHz NOESY NMR spectrum of **2**. NOESY parameters: A mixing time of 1.5 seconds was used with a D1 of 2.5s.

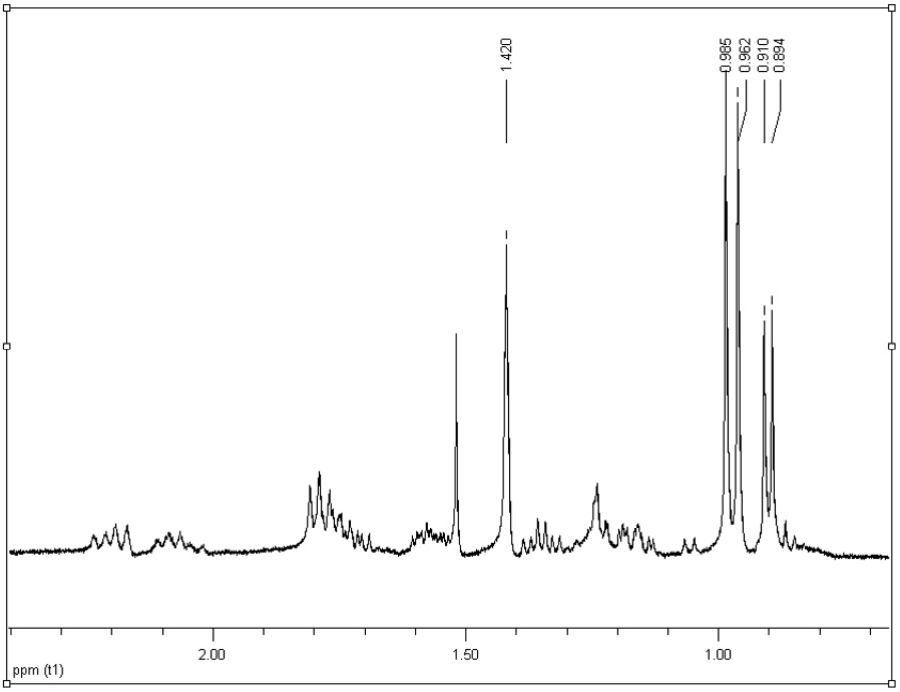


Figure S10. ^{1}H NMR (400.13 MHz) spectrum of [11,11-$^{2}H_2$]-(+)-epi-isozizaene (**2a**).

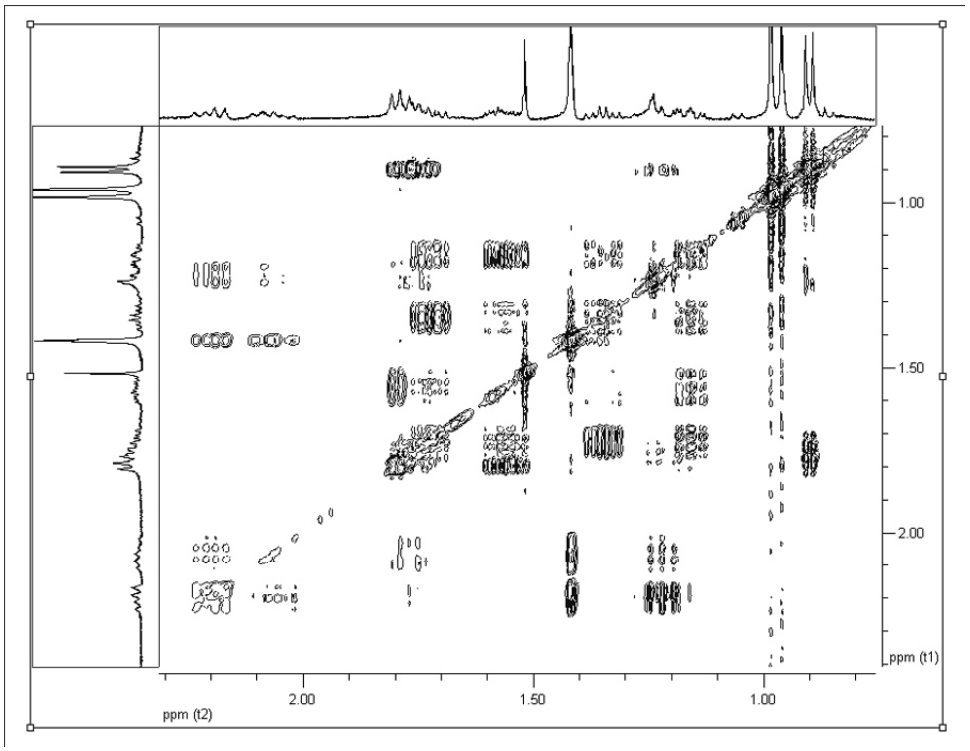


Figure S11. ^{1}H-^{1}H COSY (400.13 MHz) spectrum of [11,11-$^{2}H_2$]-(+)-epi-isozizaene (**2a**).

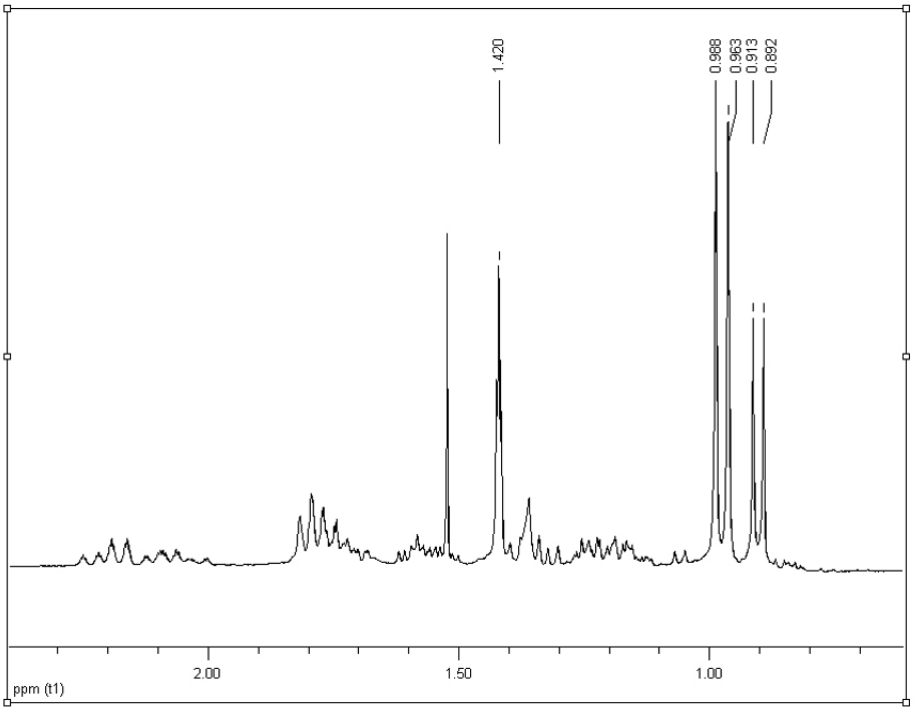


Figure S12. ^{1}H NMR (300.15 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2b**).

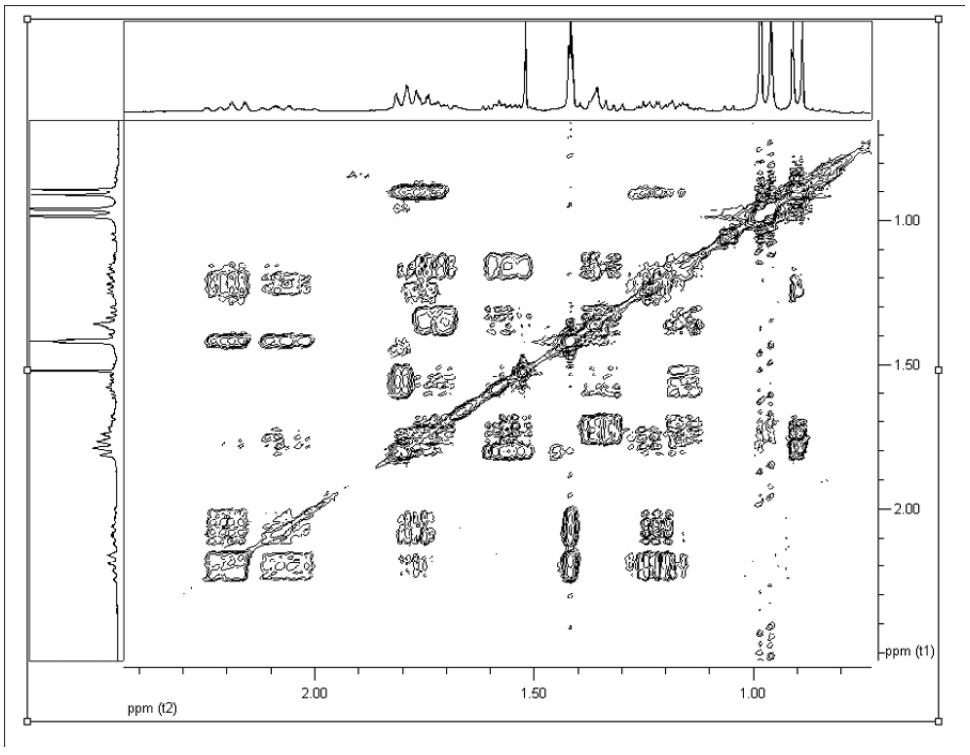


Figure S13. ^{1}H-^{1}H COSY (400.13 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2b**).

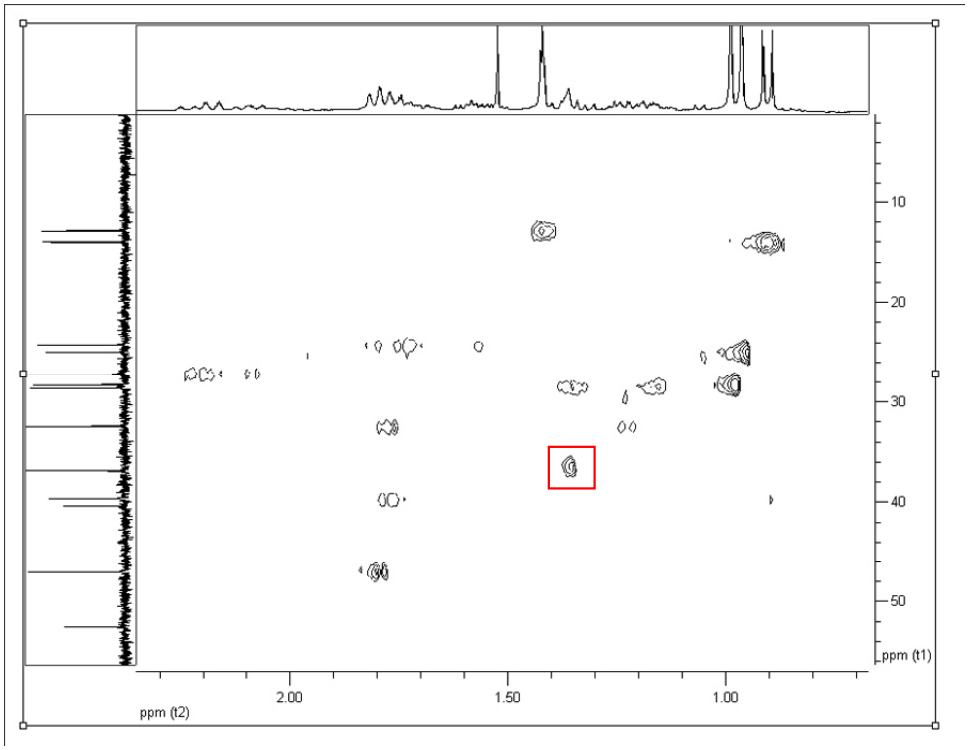


Figure S14. HSQC NMR (400.13 MHz, 100.62 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2b**). Crosspeak between H-11$_{syn}$ (δ 1.36) and C-11 (36.5 ppm).

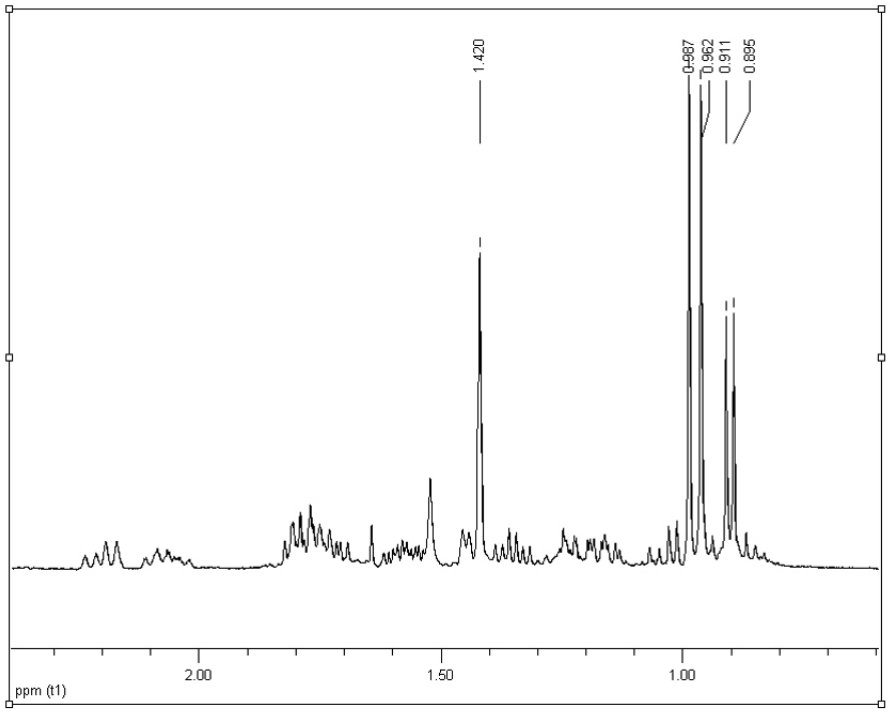


Figure S15. ^{1}H NMR (400.13 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2c**).

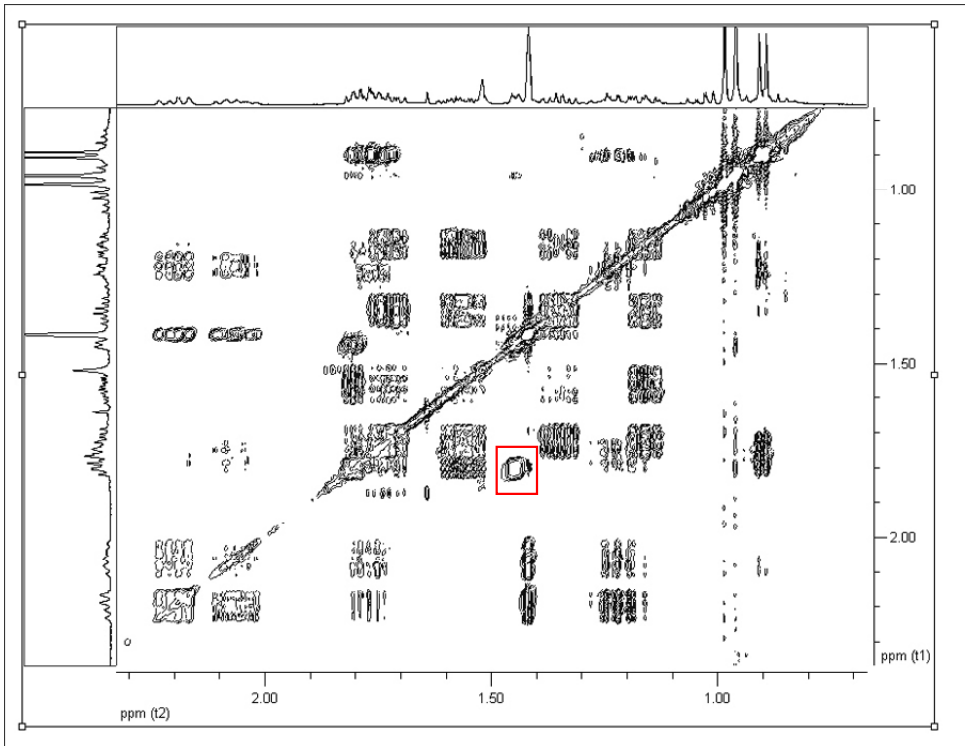


Figure S16. ^{1}H-^{1}H COSY (400.13 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2c**). Crosspeak between H-11$_{anti}$ (δ 1.45) and H-1 (δ 1.81).

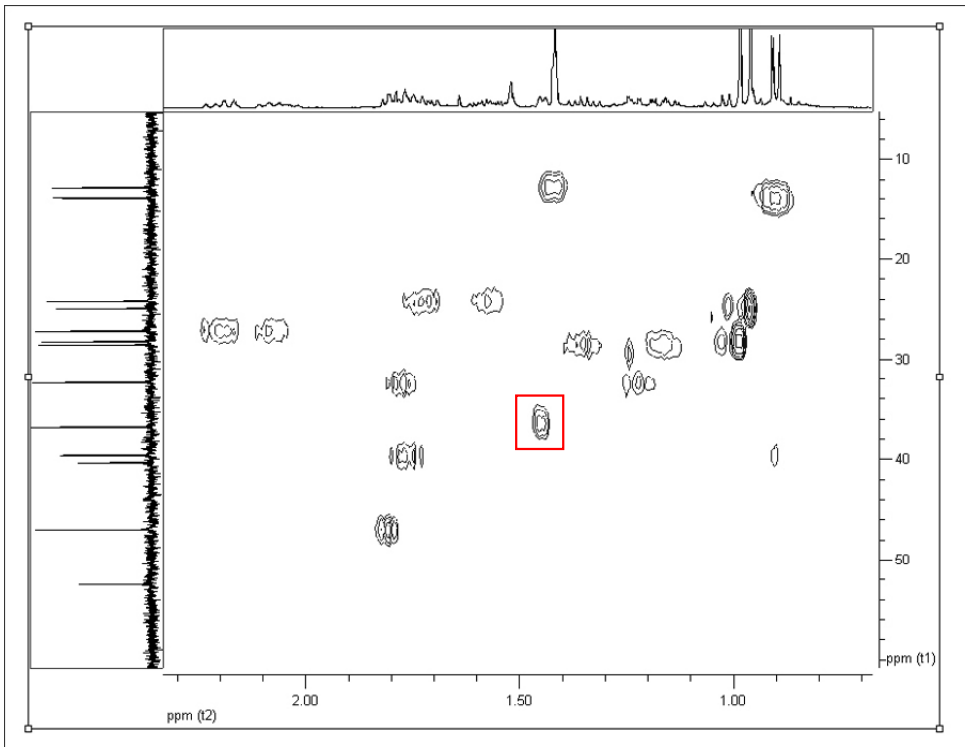


Figure S17. HSQC NMR (400.13 MHz, 100.62 MHz) spectrum of [11-^{2}H]-(+)-epi-isozizaene (**2c**). Crosspeak between H-11$_{anti}$ (δ 1.45) and C-11 (36.3 ppm).

References.

Ref 5 (full citation): Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. *Nature* **2002**, *417*, 141-147.

(1) Chenna, R.; Sugawara, H.; Koike,T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.;, Thompson, J. D. *Nucl. Acids Res.* **2003**, *31*, 3497-3500.

(2) (a) Cane, D. E.; Watt, R. M. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1547-1551. (b) Tetzlaff, C. N.; You, Z.; Cane, D. E.; Takamatsu, S.; Omura, S.; Ikeda, H. *Biochemistry* **2006**, *in press.*

(3) (a) Wolf, H.; Schoppmann, H. *FEMS Microbiol. Lett.*, **1989**, *57*, 259-264. (b) Hanahan, D.; Meselson, M. *Gene* **1980**, *10*, 63-67.

(4) (a) Krings, U.; Kelch, M.; Berger, R. G. *J. Chem. Tech. Biotechnol.* **1993**, *58*, 293-299. (b) Pollak, F. C.; Berger, R. G. *Appl. Env. Microbiol.* **1996**, *62*, 1295-1299.