Submitted for the Biotechnology section of *Applied and Environmental Microbiology*

REVISED AEM02312-09

Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*

Harry R. Beller*† [1,2], Ee-Been Goh† [1,3], Jay D. Keasling [1,3,4]

† H.R.B. and E.B.G. contributed equally to this study

[1]Joint BioEnergy Institute, 5885 Hollis Avenue, Emeryville, CA 94608

[2]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[4]Departments of Chemical Engineering and of Bioengineering, University of California, Berkeley, CA 94720

Running title: Alkene biosynthesis genes in *Micrococcus luteus*

***Corresponding author:**

Harry R. Beller
Joint BioEnergy Institute
5885 Hollis Avenue
Emeryville, CA 94608

E-mail HRBeller@lbl.gov
Phone (510) 486-7321
Fax (510) 486-5686

Abstract

Aliphatic hydrocarbons are highly appealing targets for advanced cellulosic biofuels, as they are already predominant components of petroleum-based gasoline and diesel fuels. We have studied alkene biosynthesis in *Micrococcus luteus* ATCC 4698, a close relative of *Sarcina lutea* (now *Kocuria rhizophila*), which four decades ago was reported to biosynthesize *iso*- and *anteiso*-branched, long-chain alkenes. The underlying biochemistry and genetics of alkene biosynthesis were not elucidated in those studies. We show here that heterologous expression of a three-gene cluster from *M. luteus* (Mlut_13230-13250) in a fatty-acid overproducing *E. coli* strain resulted in production of long-chain alkenes, predominantly 27:3 and 29:3 (no. carbon atoms: no. C=C bonds). Heterologous expression of Mlut_13230 (*oleA*) alone produced no long-chain alkenes but unsaturated aliphatic monoketones, predominantly 27:2, and in vitro studies with the purified Mlut_13230 protein and tetradecanoyl-CoA produced the same C_{27} monoketone. Gas chromatography-time of flight mass spectrometry confirmed the elemental composition of all detected long-chain alkenes and monoketones (putative intermediates of alkene biosynthesis). Negative controls demonstrated that the *M. luteus* genes were responsible for production of these metabolites. Studies with wild-type *M. luteus* showed that the transcript copy number of Mlut_13230-13250 and the concentrations of 29:1 alkene isomers (the dominant alkenes produced by this strain) generally corresponded with bacterial population over time. We propose a metabolic pathway for alkene biosynthesis starting with acyl-CoA (or -ACP) thioesters and involving decarboxylative Claisen condensation as a key step, which we believe is catalyzed by OleA. Such activity is consistent with our data and with the homology (including the conserved Cys-His-Asn catalytic triad) of Mlut_13230 (OleA) to FabH (β-ketoacyl-ACP synthase III), which catalyzes decarboxylative Claisen condensation during fatty acid biosynthesis.

INTRODUCTION

Aliphatic hydrocarbons are favorable targets for advanced cellulosic biofuels, as they are already predominant components of petroleum-based gasoline and diesel fuels and thus would be compatible with existing engines and fuel distribution systems. Certain bacteria are promising sources of the enzymes necessary for conversion of saccharification products such as glucose to aliphatic hydrocarbons, as a number of strains capable of aliphatic hydrocarbon production have been reported (13). Although some of these reports have proven irreproducible and are in question (e.g., (23)), alkene biosynthesis was well documented in *Sarcina lutea* ATCC 533 (now *Kocuria rhizophila*), which four decades ago was reported by two research groups to biosynthesize *iso*- and *anteiso*-branched, long-chain (primarily C_{25} to C_{29}) alkenes (2, 3, 22). The biosynthetic pathway was postulated to involve decarboxylation and condensation of fatty acids, however the underlying biochemistry and genetics of alkene biosynthesis were not elucidated. We chose to study alkene biosynthesis in *Micrococcus luteus* ATCC 4698 (also NCTC 2665), a close relative of *S. lutea* for which a genome sequence is available (25) and in which we have observed long-chain alkene biosynthesis. In this article, we provide in vivo and in vitro evidence of proteins in *M. luteus* that catalyze production of long-chain alkenes (and a key alkene biosynthesis intermediate, a long-chain monoketone) when expressed heterologously in *E. coli*, and also report how expression of the three relevant genes relates to growth and alkene production in wild-type *M. luteus*.

MATERIALS AND METHODS

Bacterial strains, plasmids, oligonucleotides, and reagents. Bacterial strains and plasmids used in this study are listed in Table 1. Plasmid extractions were carried out using the QIAGEN (Valencia, CA) miniprep and midiprep kits. Oligonucleotides were designed using the web-based PrimerBlast program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHomeAd) and synthesized by Integrated DNA Technologies (San Diego, CA) or Bioneer (Alameda, CA). *M. luteus* locus tags (e.g., Mlut_13230) used in Table 1 and elsewhere in this article correspond to the whole-genome sequence available in the GenBank/EMBL database under accession no. CP001628.

Media and bacterial growth. *E. coli* was propagated as previously described (18), whereas *M. luteus* was propagated at 30°C in tryptic soy broth or on tryptic soy agar plates.

For most *M. luteus* studies described here and for studies of heterologous gene expression in *E. coli* DH1 strains, cells were grown in 15 ml of tryptic soy broth in a 30-ml glass tube with 200 rpm agitation at 30°C for up to 60 hours before being harvested for analysis. Cultures grown for protein purification were cultivated with an autoinduction medium containing Luria-Bertia broth, phosphate buffer, and carbon sources as described by Studier (20).

When required, antibiotics were added to the growth medium at the following final concentrations: chloramphenicol, 25 μg/ml; kanamycin, 50 μg/ml (100 μg/ml when autoinduction medium was used). A final concentration of 0.5 mM IPTG was added to media when induction of genes was required.

Plasmids and strain construction for heterologous expression in *E. coli*. To clone *M. luteus* genes into expression plasmids, genomic DNA was first isolated using the Genomic-DNA tips and Genomic DNA buffer set from QIAGEN and used as the template for PCR amplification

66 of the genes of interest. To reduce error rates in the DNA amplification reaction, Phusion DNA
67 polymerase (Finnzymes, Woburn, MA) was used. In addition, due to the high-GC (73%) DNA
68 content of *M. luteus,* 10% DMSO was included in the PCR reaction to eliminate any secondary
69 structure of the template. For templates that were more difficult to amplify, 1M betaine (final
70 concentration) was used instead of DMSO. All primers used to amplify target genes are listed in
71 Table 2. PCR products and plasmid DNA were digested with the appropriate restriction enzymes
72 and purified with QIAquick gel extraction and/or PCR purification kits (QIAGEN) before being
73 ligated and transformed into *E. coli*. Proper clone construction was confirmed by DNA
74 sequencing, which was performed by Quintara Biosciences (Berkeley, CA). Expression of *M.*
75 *luteus* genes in constructs was confirmed by extraction of proteins, tryptic digestion, and analysis
76 of the resulting peptides by electrospray ionization liquid chromatography-tandem mass
77 spectrometry (LC/MS/MS)(QSTAR Elite Hybrid Quadrupole TOF, Applied Biosystems).
78 **Purification of N-terminally His-tagged Mlut_13230 protein for in vitro assays.** *E.*
79 *coli* strain EGS220 (Table 1) was grown at 30°C in 200 ml of autoinduction medium for 20-24
80 hours before being harvested for protein purification. Cell lysis and protein purification were
81 carried out as described elsewhere (16) with a few modifications. Briefly, the harvested cell
82 pellet was resuspended in 50 mM Tris-Cl (pH 8.0) with 10% glycerol, 500 mM NaCl, 30 mM
83 imidazole, and 5 mM dithiothreitol (DTT). Cells were lysed by sonication followed by 3 freeze-
84 thaw cycles at -80°C in the presence of 1 mg/ml lysozyme and 0.1% TritonX-100. Clarified cell
85 lysates were incubated with Ni-NTA resin at 4°C for 1 hour with gentle rocking before being
86 applied to a gravity flow column. The column was washed with 50 mM Tris-Cl (pH 7.9)
87 containing 10% glycerol, 500 mM NaCl, 30 mM imidazole, and 5 mM DTT, and proteins were
88 eluted with the same buffer except that the imidazole concentration was increased to 200 mM.

89 Eluted proteins were concentrated and exchanged with 100 mM potassium phosphate buffer (pH
90 7.0) with 25 mM NaCl using Amicon Centrifugal Devices (Millipore). Purified proteins were run
91 on an 8-16% gradient SDS-PAGE gel, stained with Coomassie Blue dye, and observed to
92 contain a major band at ~40 kDa. This band was excised, an in-gel trypsin digest was
93 performed, and the digested peptides were analyzed by electrospray ionization LC/MS/MS
94 (QSTAR Elite Hybrid Quadrupole TOF, Applied Biosystems) to confirm that the 40-kDa protein
95 band corresponded to the Mlut_13230 protein. In-solution trypsin digests carried out on purified
96 OleA samples determined that OleA constituted at least 70-75% of the protein based upon
97 calculations using the exponentially modified Protein Abundance Index (emPAI) method (10).
98 **In vitro assays with purified Mlut_13230**. Assays (500 µl total volume) were
99 conducted in 4-ml screw-cap glass vials with polytetrafluoroethylene (PTFE)-lined septa. Assay
100 mixtures contained 1 mM myristoyl-CoA (Sigma), 100 µl freshly purified Mlut_13230 protein (2
101 - 4 mg/ml), and freshly prepared *E. coli* DH1 (wild-type) cell lysate in 0.1 M potassium
102 phosphate buffer (pH 7.0) containing 5 mM DTT. Preparation of cell lysates was performed as
103 follows: 10 ml of *E. coli* DH1 was grown overnight in LB at 37°C before being harvested by
104 centrifugation. Cell pellets were washed once with 0.1 M potassium phosphate buffer (pH 7.0)
105 and resuspended in 750 µl of the same phosphate buffer before being subjected to cell lysis by
106 sonication. Cell lysate was clarified by centrifugation and the supernatant used for the in vitro
107 assay.
108 Controls included assay mixtures without Mlut_13230 protein or without DH1 cell lysate.
109 Assay vials were gently shaken for 1.5 hours at 30°C. After incubation, assays were extracted
110 with high-purity hexane (OmniSolv; EMD Chemicals) that was amended with two internal
111 standards: decane-d_{22} and tetracosane-d_{50} (each at a final concentration of 40 ng/µl for gas

chromatography/mass spectrometry (GC/MS) analysis). 1 ml hexane was added to the assay solution, mixed well, allowed to sit for 30 min, and the vials were centrifuged at 2000 rpm for 10 min (20°C) in an Allegra 25R centrifuge with an A14 rotor (Beckman Coulter). The extraction step was repeated, the two 1-mL aliquots of hexane were combined, and the extracts were derivatized with ethereal diazomethane (5, 6) with high-purity diethyl ether (>99.8% purity, preserved with 2% ethanol, Fluka) and concentrated under a gentle stream of ultra high-purity N_2 to 50 µl for analysis by GC/MS. Throughout the entire procedure, the hexane contacted only glass or PTFE.

Extraction of long-chain aliphatic hydrocarbons and related metabolites from bacterial cultures. Fifteen-ml cultures (either *E. coli* constructs or wild-type *M. luteus*) in 30-ml glass tubes with PTFE-lined screw-cap closures were centrifuged at 3500 rpm for 15 min (20°C) in an Allegra 25R centrifuge with an A14 rotor and the aqueous phase was decanted. The pellet was amended with 100 µl of reagent water and the mixture was homogenized with a vortex mixer. Then 1 ml of high-purity methanol (B&J Brand, $\geq$99.9% purity) and 4 ml high-purity hexane were added to the cells; as discussed previously, the hexane was amended with perdeuterated alkane standards to assess sample-specific analytical recovery. The cell-solvent mixture was homogenized with a vortex mixer and sonicated in an ice bath for 15 min, allowed to sit for 10 min, and then centrifuged at 3500 rpm for 15 min (20°C). The hexane layer was then removed with a solvent-cleaned Pasteur pipette and transferred to a glass, 10-ml conical vial. The hexane was concentrated to 100 µl under a gentle stream of ultra high-purity N_2; 50 µl was transferred (*via* 100-µl gas-tight glass syringe) to a vial for GC/MS analysis and the remaining 50 µl was derivatized with ethereal diazomethane (as discussed previously) and then concentrated to 50 µl for GC/MS analysis. *M. luteus* extracts were not derivatized. As discussed

for the in vitro assay extractions, organic solvents contacted only glass or PTFE, and all glass and PTFE surfaces were rigorously pre-cleaned with high-purity acetone.

Analysis by GC/MS (quadrupole and time of flight, or TOF). For electron ionization (EI) GC/MS analyses with a quadrupole mass spectrometer, studies were performed with a model 7890A GC (Agilent) with a DB-5 fused silica capillary column (30-m length, 0.25-mm inner diameter, 0.25-μm film thickness; J & W Scientific) coupled to an HP 5975C series mass selective detector; 1 μl injections were performed by a model 7683B autosampler. The GC oven was typically programmed from 40°C (held for 3 min) to 300°C at 15°C/min and then held for 20 min; the injection port temperature was 250°C, and the transfer line temperature was 280°C. The carrier gas, ultra high-purity helium, flowed at a constant rate of 1 ml/min. Injections were splitless, with the split turned on after 0.5 min. For full-scan data acquisition, the MS typically scanned from *m/z* 50 to 600 at a rate of 2.7 scans per s. Selected ion monitoring (SIM) acquisition was used for certain studies when additional sensitivity was required; specific ions monitored for SIM are discussed in the Results section, when applicable.

Selected samples were subjected to GC-TOF analysis to confirm the elemental composition of key metabolic products. GC-chemical ionization (CI)-TOF analyses were performed with a Waters/MicroMass GCT instrument scanning from *m/z* 65 to 800 with GC conditions as described previously; positive-ion CI mode was used and the reagent gas was methane. GC-EI-TOF analyses were carried out with a Waters GCT Premier instrument scanning from *m/z* 35 to 650 (with Dynamic Range Enhancement) with GC conditions as described previously. Elemental compositions were calculated with MassLynx software.

Transcriptional studies of *M. luteus* with reverse transcription-quantitative PCR (RT-qPCR) analysis. For transcriptional studies, RNA in *M. luteus* cultures was preserved

158 immediately before harvesting by adding an ethanol solution containing 5% phenol. Extraction
159 and purification of RNA were carried out with QIAGEN RNeasy kits. Concentration and
160 integrity of RNA were determined with a Thermo Scientific Nanodrop ND-1000
161 spectrophotometer and an Agilent 2100 BioAnalyzer, respectively.

162 Synthesis of cDNA for RT-qPCR analysis was carried out using 2 µg of total RNA
163 primed with 10 µg of random hexamers and reverse transcribed using SuperScript III enzyme
164 (Invitrogen, Carlsbad, CA). The reverse transcription reaction was carried out for 2 hours at
165 50°C before the RNA was hydrolyzed with 2 U of RNaseH (Invitrogen) at 37°C for 30 min.
166 qPCR analyses were then conducted with an Applied Biosystems StepOne system using 2 µl of
167 the reverse transcription reaction and gene-specific primers (Table 2), and the PerfeCTa SYBR
168 Green FastMix (Quanta Biosciences). Quantitative PCR cycle parameters were as follows: initial
169 denaturation at 95°C for 10 min, followed by 40 cycles of 15 s denaturation at 95°C and 1 min
170 annealing and extension at 60°C. Fluorescence measurements were taken between each cycle. At
171 the conclusion of the qPCR cycle, melting curve analysis was conducted by denaturing the PCR
172 products from 60°C to 95°C and making fluorescence measurements at 0.3°C increments. All
173 reactions were performed in duplicate. Transcripts were quantified with reference to a standard
174 curve generated by serial dilution of pEG142 (from 10^5 to 10^{10} copies/reaction).

175 **RESULTS**

176 **Identification of condensing enzymes as potential candidates for a key alkene**
177 **biosynthesis step.** After unsuccessful attempts to identify candidate genes for alkene
178 biosynthesis *via* transcriptomics in *M. luteus* (not addressed in this article), we drew upon
179 findings of previous *S. lutea* research and hypothesized that enzymes catalyzing fatty acid
180 decarboxylation and condensation would play an important role in alkene biosynthesis.

Specifically, some important observations for *S. lutea* were that (i) the dominant alkenes in *S. lutea*, namely, *iso*- and *anteiso*-branched C_{29} monoalkenes with the double bond near the center (at C-13), are very plausibly derived from decarboxylation and "head-to-head" condensation of the dominant fatty acids in that bacterium, namely, *iso*- and *anteiso*-branched C_{15} saturated acids (1, 2) and (ii) in vitro studies with cell-free *S. lutea* extracts, palmitate-16-^{14}C, palmitate-1-^{14}C, and their CoA derivatives, indicated that decarboxylation of acyl-CoAs was important in alkene biosynthesis (3).

Based upon these observations for *S. lutea*, we hypothesized that homologs of "condensing enzymes" involved in fatty acid biosynthesis [i.e., β-ketoacyl-ACP (acyl carrier protein) synthases] could be involved in alkene biosynthesis from fatty acids, as these enzymes catalyze decarboxylation of activated aliphatic acids (malonyl-ACP) and nucleophilic attack by the resulting carbanion on an acyl-CoA or acyl-ACP thioester (Claisen condensation) (9, 24).

A search of the draft genome sequence of *M. luteus* for genes associated with fatty acid metabolism revealed three possible condensing enzymes: Mlut_09290 (β-ketoacyl-ACP synthase II, or FabF), Mlut_09310 (β-ketoacyl-ACP synthase III, or FabH), and Mlut_13230 (a possible FabH homolog). Alignments of the translated products of these three genes and the most similar sequences from *E. coli* and two sequenced, Gram-positive, close relatives of *M. luteus* (*Arthrobacter aurescens* TC1 and *Arthrobacter* sp. strain FB24) revealed the presence of three key conserved, active-site residues characteristic of condensing enzymes (24) in all sequences (Fig. 1): Cys-His-Asn for the FabH homologs (Mlut_09310 and Mlut_13230) and Cys-His-His for the FabF homolog (Mlut_09290). Furthermore, based on their gene neighborhood, it seems likely that Mlut_09290 and Mlut_09310 are respectively *fabF* and *fabH*, which encode key condensing enzymes involved in fatty acid biosynthesis; the six-gene cluster containing

Mlut_09310 and Mlut_09290 includes a number of other genes critical to biosynthesis of branched-chain fatty acids, including a putative branched-chain α-keto acid decarboxylase (Mlut_09340), malonyl-CoA:ACP transacylase (*fabD*; Mlut_09320), and acyl carrier protein (ACP; Mlut_09300). In addition, it is clear from Fig. 1 that Mlut_09310 and Mlut_09290 have relatively high sequence identity to known copies of FabH and FabF, respectively, in contrast to Mlut_13230, which has relatively low sequence identity to *E. coli* FabH.

Thus, the putative condensing enzymes Mlut_09290, Mlut_09310, and Mlut_13230 were selected as candidates for catalyzing an important reaction in alkene biosynthesis.

Long-chain alkenes and unsaturated monoketones resulting from heterologous expression of *M. luteus* condensing enzymes (and associated genes) in fatty acid-overproducing *E. coli*. To test the hypothesis that one or more of the putative condensing enzymes in *M. luteus* has a role in alkene biosynthesis, we expressed Mlut_13230, Mlut_09290, and Mlut_09310 in a fatty acid-overproducing *E. coli* strain (strains EGS180, EGS210, and EGS212, respectively; Table 1) and analyzed the metabolites by GC/MS. Comparison of total ion chromatograms (TIC) from extracts of strains EGS210 and EGS212 with those of a negative control (empty vector; strain EGS084; Table 1) did not reveal any new peaks resulting from the presence of Mlut_09290 or Mlut_09310 (data not shown). However, the TIC representing strain EGS180 did reveal some noteworthy peaks relative to the negative control (peaks labeled 27:2, 27:1, 29:2, and 29:1 in Fig. 2A; these labels represent X:Y, where X = carbon number and Y = number of C=C bonds). The 27:2 peak was particularly prominent. The TIC in Fig. 2A represents an extract derivatized with diazomethane. The extracts were derivatized to reduce baseline noise by converting abundant and strongly tailing free fatty acids to fatty acid methyl esters, which were well resolved chromatographically and had minimal tailing. The labeled

227 peaks in Fig. 2A were also present in the TIC of underivatized samples but were less prominent
228 (e.g., Fig. 2B); thus, derivatization did not create these compounds but merely enhanced their
229 detectability. Mass spectra of these peaks (e.g., Fig. 3A) were consistent with mono- and di-
230 unsaturated C_{27} and C_{29} monoketones, as the nominal molecular ions for peaks 27:2, 27:1, 29:2,
231 and 29:1 were at *m/z* 390 ($C_{27}H_{50}O$), 392 ($C_{27}H_{52}O$), 418 ($C_{29}H_{54}O$), and 420 ($C_{29}H_{56}O$),
232 respectively. Although authentic standards are not available for these compounds, GC-EI-TOF
233 and GC-CI-TOF analyses confirmed the elemental compositions just described. For the 27:2,
234 27:1, 29:2, and 29:1 monoketones, measured masses agreed with the calculated masses within
235 0.4 mDa absolute error and 1.0 ppm relative error.
236 Because C_{27} and C_{29} unsaturated monoketones are plausible intermediates in a
237 hypothesized pathway of alkene biosynthesis from C_{14} and C_{16} fatty acids (see Discussion), we
238 further pursued the possible role of Mlut_13230 in alkene biosynthesis. We constructed a
239 plasmid containing the native three-gene cluster that includes Mlut_13230 (i.e., Mlut_13230-
240 13250) and expressed it in a fatty acid-overproducing *E. coli* strain (strain EGS145; Table 1).
241 GC/MS analysis of the extract from strain EGS145 revealed peaks in the TIC that were not
242 present in strain EGS180 (with Mlut_13230 alone) or in the negative control (strain EGS084)
243 (Fig. 2B). Mass spectra of these peaks (e.g., Fig. 3B) were consistent with di- and tri-unsaturated
244 C_{27} and C_{29} alkenes, as the nominal molecular ions for peaks 27:3, 27:2, 29:3, and 29:2 were at
245 *m/z* 374 ($C_{27}H_{50}$), 376 ($C_{27}H_{52}$), 402 ($C_{29}H_{54}$), and 404 ($C_{29}H_{56}$), respectively. Authentic
246 standards are not commercially available for di- and tri-unsaturated C_{27} and C_{29} alkenes.
247 However, for the most abundant ions in these spectra (dominated by a series of ions differing by
248 14 amu, or CH_2 groups), fragmentation patterns were consistent with National Institute of
249 Standards and Technology library spectra of shorter alkenes for which standards are available;

for example, the best library match for the spectrum of the peak labeled 27:3 was a shorter tri-unsaturated alkene (22:3). Furthermore, GC-EI-TOF analyses confirmed the elemental compositions just described. For the 27:3, 27:2, and 29:3 alkenes, measured masses agreed with the calculated masses within 0.3 mDa absolute error and 0.8 ppm relative error (for the 29:2 alkene, these errors were 1.0 mDa and 2.5 ppm, respectively). The total concentration of the four alkenes was on the order of 0.5 mg/l (a 14:1 alkene standard was used for this estimate, as authentic standards were not commercially available).

In Fig. 2B, there are two peaks in the EGS180 extract (Mlut_13230 only) that co-elute with the peaks labeled 29:3 and 29:2 in the EGS145 extract (Mlut_13230-13250). The two co-eluting peaks for strain EGS180 are not alkenes. This is demonstrated in the two insets in Fig. 2B, in which extracted ion chromatograms characteristic of the 29:3 and 29:2 alkenes (molecular ions at *m/z* 402 and 404, respectively) clearly show that these alkenes are present in the extract from strain EGS145 but are not detectable in the extract of strain EGS180 (or the negative control, strain EGS084).

To provide more information on the possible roles of the three *M. luteus* genes that, in combination, enable alkene biosynthesis in *E. coli*, we constructed more strains that heterologously expressed either Mlut_13240 or Mlut_13250. As summarized in Table 3, heterologous expression of Mlut_13240 or Mlut_13250 alone did not result in formation of the long-chain monoketones or alkenes observed with Mlut_13230 and Mlut_13230-13250.

In vitro studies with the purified Mlut_13230 protein. To confirm that long-chain, unsaturated monoketones observed during in vivo studies (Fig. 2A, 3A) derive from fatty acid condensation and that Mlut_13230 catalyzes this reaction, we conducted in vitro studies with N-terminally His_6-tagged Mlut_13230 protein (Table 1) and acyl-CoA (specifically, tetradecanoyl-

CoA). In addition to the purified protein and acyl-CoA, the assays were amended with wild-type *E. coli* DH1 cell-free lysates, because the alkene pathway likely includes several preliminary steps that are not catalyzed by Mlut_13230 (see Discussion). Briefly, these preliminary steps may involve conversion of tetradecanoyl-CoA to 3-oxotetradecanoyl-CoA (e.g., via the early steps of beta-oxidation); we propose such β-ketoacyl-CoA (or -ACP) thioesters as substrates for the Mlut_13230 protein.

Assays including purified Mlut_13230 protein, tetradecanoyl-CoA, and DH1 lysate resulted in the formation of the same 27:2 monoketone that was prominent during in vivo studies of strain EGS180 (Fig. 2A, 4). Negative control assays conducted without Mlut_13230 protein or without DH1 lysate did not result in clearly detectable 27:2 monoketone (Fig. 4). These results indicate that acyl-CoAs (or their derivatives) are the source of the long-chain monoketones observed in vivo and that the Mlut_13230 protein is responsible for long-chain monoketone production.

To attain the necessary sensitivity for long-chain monoketone detection during in vitro assays, mass spectral data were acquired in the SIM mode employing prominent and characteristic ions for the 27:2 monoketone (*m/z* 291; Fig. 3A) and the 27:1 monoketone (*m/z* 293; Fig. 3A). Evidence supporting that the 27:2 monoketone peak observed in the in vivo studies was the same compound observed in the in vitro studies includes identical GC retention times and agreement of full-scan mass spectral patterns (albeit of lower quality for the in vitro studies because of lower concentration). The use of SIM for in vitro studies leaves open the possibility that additional metabolites were formed but not detected.

Long-chain alkene production and transcription of Mlut_13230-13250 in *M. luteus*. Our initial studies of *M. luteus* ATCC 4698 confirmed that it produced long-chain alkenes, which

296 were dominated by two C_{29} monoalkene peaks (hereafter referred to as alkene 1 and alkene 2,
297 where alkene 1 eluted approximately 0.3 min before alkene 2 on GC/MS). For both alkene 1 and
298 alkene 2, GC/MS analysis demonstrated a nominal molecular mass of 406 (consistent with
299 $C_{29}H_{58}$) and a fragmentation pattern characteristic of alkenes. GC-CI-TOF analysis for alkene 1
300 determined a molecular mass of 406.4536, which is within 0.3 mDa absolute error and 0.7 ppm
301 relative error of the calculated mass of 406.4539 for $C_{29}H_{58}$. Similar results were obtained for
302 alkene 2.

303 Alkene 2 appears to be more *anteiso*-substituted than alkene 1, based upon experiments
304 in which isoleucine was added to the growth medium. In bacteria like *M. luteus* that synthesize
305 *iso*- and *anteiso*-branched fatty acids, isoleucine is a precursor for α-keto-β-methylvalerate,
306 which in turn serves as a primer for *anteiso*-branched fatty acids (11, 15). When the TSB
307 medium was amended with isoleucine (2 mM initially and 4 mM in early stationary phase), the
308 alkene 1 concentration at 48 hr was comparable to that of an unamended control (within 15%),
309 whereas the alkene 2 concentration was more than 3-fold higher than in the unamended control.
310 Thus, it seems likely that alkene 2 is *anteiso*-substituted at both ends (i.e., it is the product of
311 condensation of two *anteiso*-substituted fatty acids).

312 Examination of alkene 1 and 2 production throughout growth revealed that concentration
313 trends generally corresponded to growth (OD_{600}) and that the alkene 2:alkene 1 ratio increased
314 considerably from late exponential phase (15 hr) through early stationary phase (24 hr to 48 hr)
315 (Fig. 5A). In Fig. 5A, OD_{600} and alkene 1 and 2 concentrations are normalized to their
316 maximum values, and the insets (chromatograms showing alkenes 1 and 2 at 15, 24, and 48 hr)
317 are all shown with the same *y*-axis scale. The apparent decrease in alkene 1 and 2 concentrations
318 between 24 and 48 hr is likely a result of reduced extraction efficiency at the higher cell density

319 at 48 hr (OD_{600} ~ 6.1) compared to 24 hr (OD_{600} ~ 2.4), rather than the result of alkene
320 degradation, as genes associated with alkane degradation were not found in the genome. Such
321 decreases in C_{29} alkene concentration in post-exponential phase have been observed in the
322 related *Arthrobacter chlorophenolicus* A6 (7).

323 Expression of the 3-gene cluster associated with alkene production (Mlut_13230-13250)
324 generally corresponded to growth (Fig. 5B), as did C_{29} alkene production. Transcript copy
325 number for Mlut_13230, 13240, and 13250 as determined by RT-qPCR analysis is normalized to
326 the maximum observed value for each gene in Fig. 5B. Expression of these three genes does not
327 appear to vary much through the period of maximum alkene production and into stationary
328 phase. Based upon similar expression profiles for these three genes and predictions using a
329 method described by Price and co-workers (17), it appears that Mlut_13230-13250 constitutes an
330 operon.

331 **DISCUSSION**

332 We have shown that heterologous expression of three genes from *M. luteus* (Mlut_13230-13250)
333 in a fatty-acid overproducing strain of *E. coli* resulted in production of long-chain alkenes,
334 predominantly 27:3 and 29:3. Heterologous expression of Mlut_13230 alone produced
335 unsaturated monoketones, predominantly 27:2, and in vitro studies with the purified Mlut_13230
336 protein, tetradecanoyl-CoA, and wild-type *E. coli* DH1 lysate produced the same monoketone.
337 Recently, in an international patent application, Friedman and Da Costa (8) showed similar
338 results for homologous genes from other bacteria. For example, heterologous expression of
339 *oleACD* from *Stenotrophomonas maltophilia* in *E. coli* resulted in long-chain alkenes,
340 predominantly 27:3, 27:2, 29:3, and 29:2. The four genes in *S. maltophilia* strain R551-3 that
341 Friedman and Da Costa named *oleABCD* are apparent homologs of *M. luteus* genes featured in

342 this study; the OleA and OleD sequences from *S. maltophilia* strain R551-3 are each 39%
343 identical to the translated sequences of Mlut_13230 and Mlut_13250, respectively, and
344 Mlut_13240 appears to be a fusion of *oleB* and *oleC*. Also, similar to our results for
345 heterologous expression of Mlut_13230 in *E. coli*, Friedman and Da Costa reported that
346 heterologous expression of *oleA* from *S. maltophilia*, *Xanthomonas axonopodis*, or *Chloroflexus*
347 *aggregans* in *E. coli* resulted in predominantly 27:2, 27:1, and 27:0 monoketones. Finally,
348 Friedman and Da Costa observed that in vitro studies with purified OleA, tetradecanoyl-CoA,
349 and *E. coli* C41(DE3) lysate produced a C_{27} monoketone. In contrast to the present study,
350 Friedman and Da Costa did not assess heterologous expression in a Gram-negative host of
351 *oleABCD* genes from Gram-positive bacteria (like *M. luteus*) that produce *iso*- and *anteiso*-
352 branched fatty acids and alkenes, nor did they provide detailed evidence confirming the identity
353 of the alkenes and monoketones, such as the GC-TOF analyses reported here.
354 We propose a pathway for alkene biosynthesis from fatty acyl-CoAs (or -ACPs) that is
355 based largely on enzyme activities homologous to those essential for fatty acid biosynthesis (Fig.
356 6). For brevity throughout the following discussion, we discuss CoA thioesters with the
357 acknowledgment that ACP thioesters may actually be involved. We hypothesize that the first
358 key step in alkene biosynthesis involves not two fatty acyl-CoAs as substrates but rather a fatty
359 acyl-CoA and a β-ketoacyl-CoA. Thus, as suggested in Fig. 6, a fatty acyl-CoA could be
360 converted to a β-ketoacyl-CoA by early steps of β-oxidation (e.g., *via* an acyl-CoA
361 dehydrogenase, an enoyl-CoA hydratase, and a 3-hydroxyacyl-CoA dehydrogenase). The first
362 step of alkene biosynthesis in *M. luteus* (and other bacteria), catalyzed by OleA (e.g.,
363 Mlut_13230), could be decarboxylation of the β-ketoacyl-CoA and nucleophilic attack by the
364 resulting carbanion on an acyl-CoA to form an aliphatic diketone (Fig. 6). Such decarboxylative

Claisen condensation would be consistent with the homology (Fig. 1C) of Mlut_13230 to FabH (β-ketoacyl-ACP synthase III), which catalyzes decarboxylation of malonyl-ACP and its condensation to acetyl-CoA. In fact, the FabH active-site Cys-His-Asn residues conserved in the Mlut_13230 sequence (Fig. 1C) specifically suggest catalysis of decarboxylation by OleA; based upon structural studies of FabH in *E. coli* and *Mycobacterium tuberculosis*, the conserved Cys residue has been associated with binding of the acyl intermediate and the conserved His-Asn residues are associated with decarboxylation (9, 24). Following formation of the aliphatic diketone by OleA, alkene biosynthesis could follow a series of reductase and dehydratase reactions (Fig. 6) homologous to those catalyzed by β-ketoacyl-ACP reductases (e.g., FabG), β-hydroxyacyl-ACP dehydratases (e.g., FabZ), and enoyl-ACP reductases (e.g., FabI). In addition to carbon chain length, a key characteristic that distinguishes most intermediates in the proposed alkene biosynthesis pathway from those in the fatty acid biosynthesis pathway is the absence of an ACP thioester (for intermediates following condensation).

The data presented here are consistent with, but do not prove, the pathway proposed in Fig. 6. In vitro studies with the purified Mlut_13230 protein (OleA), tetradecanoyl-CoA, and wild-type *E. coli* DH1 lysate produced an unsaturated C_{27} monoketone, which would be consistent with the proposed pathway starting with two C_{14} thioesters (e.g., decarboxylation and condensation would yield a C_{27} compound). The need to form a β-ketoacyl-CoA as a substrate for OleA could explain why in vitro controls without cell lysate yielded negligible product (Fig. 4) - - the relevant β-oxidation genes needed to be supplied by the lysate. The cell lysate may also explain why the monoketone was observed rather than the diketone (i.e., FabG and FabZ present in the lysate may have been able to act on the diketone and β-hydroxyketone). In vivo studies of heterologous expression of *oleA* (Mlut_13230) in a fatty-acid overproducing strain of

388 *E. coli* also produced predominantly an unsaturated C_{27} monoketone (Fig. 2), suggesting the
389 physiological relevance of the in vitro studies. The C_{29} monoketones also observed in these in
390 vivo studies (Fig. 2) would be consistent with condensation of a C_{14} and C_{16} substrate. The
391 predominant di-unsaturated monoketone (27:2) observed in the in vitro and in vivo studies with
392 OleA appears inconsistent with the proposed pathway, as 27:1 or 27:0 monoketones would be
393 expected (Fig. 6). One possible explanation for the additional double bond is that two β-
394 ketoacyl-CoAs could serve as substrates rather than one β-ketoacyl-CoA and one acyl-CoA; this
395 would lead to the formation of a triketone and a di-unsaturated monoketone following reactions
396 analogous to those shown in Fig. 6. Our data for heterologous expression of Mlut_13230-13250
397 (*oleABCD*) in *E. coli* is consistent with an aliphatic monoketone being an intermediate of alkene
398 biosynthesis. Heterologous expression of Mlut_13230-13250 (strain EGS145) resulted in 27:3
399 and 29:3 as the predominant alkenes, whereas expression of Mlut_13230 alone (strain EGS180)
400 resulted in 27:2 and 29:2 monoketones. Thus, the monoketones and alkenes had the same carbon
401 backbones but the alkenes had one additional double bond (which would be expected if the enoyl
402 reductase in the proposed pathway was not present or active). In this light, comparison of the
403 number of observed double bonds in these heterologous expression studies (strain EGS145)
404 versus alkenes produced by wild-type *M. luteus* is instructive, as *M. luteus* produces
405 monoalkenes (29:1) and thus apparently has effective enoyl reductase activity, in contrast to
406 strain EGS145.
407 Whereas multiple lines of evidence suggest that the probable role of OleA (Mlut_13230)
408 in alkene biosynthesis is catalysis of decarboxylative Claisen condensation, the roles of OleBC
409 (Mlut_13240) and OleD (Mlut_13250) are not clear from our data. Heterologous expression of
410 Mlut_13240 or13250 alone did not produce aliphatic monoketones or alkenes (Table 3).

411 However, alkene production seems to require the expression of all 3 genes, Mlut_13230-13250
412 (Table3). Mlut_13250 (OleD) was annotated as a nucleoside-diphosphate-sugar epimerase
413 (GenBank) and BLASTp searches (4) of the translated Mlut_13250 sequence against the
414 GenBank nonredundant database revealed a conserved domain of the NADB Rossmann
415 superfamily. It is thus possible that OleD is an NADH- or NADPH-dependent reductase. As
416 discussed previously, the Mlut_13240 gene appears to be a fusion of *oleB* and *oleC* (which are
417 separate genes in *S. maltophilia*, studied by Friedman and Da Costa (8)). From BLASTp
418 analysis, it appears that an N-terminal region of the Mlut_13240 protein is equivalent to OleB,
419 which has homology to the alpha/beta hydrolase fold family; a C-terminal region is equivalent to
420 OleC, which has homology to the AMP-dependent synthetase/ligase family. The role of such
421 proteins in the proposed pathway is unclear, but it seems that an AMP-dependent
422 synthetase/ligase would precede Claisen condensation, after which no metabolites would contain
423 carboxylic acid or thioester moieties. Finally, it seems possible that the enoyl reductase
424 putatively used by *M. luteus* to generate monoalkenes (Fig. 6) is encoded by a gene not included
425 in the Mlut_13230-13250 cluster, as di- and trienes (not monoenes) were observed during
426 heterologous expression of Mlut_13230-13250. If OleA, OleBC, and OleD do not include an
427 enoyl reductase, and if the pathway proposed in Fig. 6 is accurate, it follows that Mlut_13250
428 (OleD) is a keto reductase. Further study will be required to elucidate the roles of OleA, OleBC,
429 and OleD in alkene biosynthesis.

430 **ACKNOWLEDGMENTS**

431 We thank Yisheng Kang and Eric Steen (JBEI) for providing the fatty acid-
432 overproducing *E. coli* DH1 strain; Charles Greenblatt (Hebrew University) for providing early
433 access to the draft genome sequence of *Micrococcus luteus*; Alyssa Redding and Tanveer Batth

(Functional Genomics Department, Technology Division, JBEI) for mass spectrometric analysis of protein samples; Rudy Alvarado, Vladimir Tolstikov, and Saeed Khazaie (University of California at Davis Genome Center) as well as Doug Stevens and Steven Lai (Waters Corporation) for providing GC-TOF analyses; and Taek Soon Lee and Steven Singer (JBEI) for helpful comments on the manuscript.

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

REFERENCES

1. **Albro, P. W.** 1971. Confirmation of the identification of the major C-29 hydrocarbons of *Sarcina lutea*. J. Bacteriol. **108:**213-8.
2. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. I. Characterization of the hydrocarbons of *Sarcina lutea* and the isolation of possible intermediates of biosynthesis. Biochemistry **8:**394-404.
3. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. IV. Characteristics of synthesis by a cell-free preparation of *Sarcina lutea*. Biochemistry **8:**3317-24.
4. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-402.
5. **Beller, H. R., and A. M. Spormann.** 1997. Anaerobic activation of toluene and *o*-xylene by addition to fumarate in denitrifying strain T. J. Bacteriol. **179:**670-6.
6. **Fales, H. M., T. M. Jaouni, and J. F. Babashak.** 1973. Simple device for preparing ethereal diazomethane without resorting to codistillation. Anal. Chem. **45:**2302-2303.
7. **Frias, J. A., J. E. Richman, and L. P. Wackett.** 2009. C_{29} olefinic hydrocarbons biosynthesized by *Arthrobacter* species. Appl. Environ. Microbiol. **75:**1774-7.
8. **Friedman, L., and B. Da Costa.** 2008. Hydrocarbon-producing genes and methods of their use. International Patent Application **WO 2008/147781**.
9. **Heath, R. J., and C. O. Rock.** 2002. The Claisen condensation in biology. Nat. Prod. Rep. **19:**581-96.

10. **Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann.** 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell Proteomics **4:**1265-72.

11. **Kaneda, T.** 1991. *Iso*- and *anteiso*-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. Microbiol. Rev. **55:**288-302.

12. **Kirchner, O., and A. Tauch.** 2003. Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. J. Biotechnol. **104:**287-99.

13. **Ladygina, N., E. G. Dedyukhina, and M. B. Vainshtein.** 2006. A review on microbial synthesis of hydrocarbons. Process Biochemistry **41:**1001-1014.

14. **Meselson, M., and R. Yuan.** 1968. DNA restriction enzyme from *E. coli*. Nature **217:**1110-4.

15. **Oku, H., K. Fujita, T. Nomoto, K. Suzuki, H. Iwasaki, and I. Chinen.** 1998. NADH-dependent inhibition of branched-chain fatty acid synthesis in *Bacillus subtilis*. Biosci. Biotechnol. Biochem. **62:**622-7.

16. **Petty, K. J.** 1996. Metal Chelate Affinity Chromatography, p. 10.11.10-10.11.24. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current Protocols in Molecular Biology. John Wiley & Sons, Inc.

17. **Price, M. N., K. H. Huang, E. J. Alm, and A. P. Arkin.** 2005. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucleic Acids Res. **33:**880-92.

18. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

19. **Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling.** In Press. Microbial production of fatty acid-derived chemicals from plant biomass. Nature. DOI 10.1038/nature08721

20. **Studier, F. W.** 2005. Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. **41:**207-234.

21. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113-30.

22. **Tornabene, T. G., E. Gelpi, and J. Oro.** 1967. Identification of fatty acids and aliphatic hydrocarbons in *Sarcina lutea* by gas chromatography and combined gas chromatography-mass spectrometry. J. Bacteriol. **94:**333-43.

23. **Wackett, L. P., J. A. Frias, J. L. Seffernick, D. J. Sukovich, and S. M. Cameron.** 2007. Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. Appl. Environ. Microbiol. **73:**7192-8.

24. **White, S. W., J. Zheng, Y. M. Zhang, and Rock.** 2005. The structural biology of type II fatty acid biosynthesis. Annu. Rev. Biochem. **74:**791-831.

25. **Young, M., V. Artsatbanov, H. R. Beller, G. Chandra, K. F. Chater, L. G. Dover, E.-B. Goh, T. Kahan, A. S. Kaprelyants, N. Kyrpides, A. Lapidus, S. R. Lowry, A. Lykidis, J. Mahillon, V. Markowitz, K. Mavrommatis, G. V. Mukamolova, A. Oren, J. S. Rokem, M. C. M. Smith, D. I. Young, and C. L. Greenblatt.** In press. Genome sequence of the Fleming strain of *Micrococcus luteus*, a simple free-living actinobacterium. J. Bacteriol.

TABLE 1. Bacterial strains, plasmids and primers used in this study

Strain or plasmid	Relevant characteristics	Source or reference
E. coli strains		
BL21 (DE3)	F^- *ompT gal dcm lon hsdSB*($r_B^- m_B^-$) λ (DE3)	(21)
DH1	*endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR*17($r_K^- m_K^+$) λ^-	(14)
LT-ΔfadE	DH1 *ΔfadE* with pKS1	(19)
EGS084	LT-ΔfadE with pEC-XK99E	This study
EGS145	LT-ΔfadE with pEG142	This study
EGS180	LT-ΔfadE with pEG174	This study
EGS210	LT-ΔfadE with pEG200	This study
EGS212	LT-ΔfadE with pEG205	This study
EGS220	BL21(DE3) with pEG185	This study
EGS244	LT-ΔfadE with pEG225	This study
EGS300	LT-ΔfadE with pEG275	This study
Micrococcus luteus strains		
ATCC 4698	Wild type	ATCC
Plasmids		
pEC-XK99E	Km^r; *E. coli* - *C. glutamicum* shuttle expression vectors with ColE1 origin of replication and *trc*	(12)

	promoter	
pKS1	Cm^r; p15a derivative containing *LtesA* under the *lacUV5* promoter	(19)
pSKB3	Km^r; A derivative of the expression vector pET-28a with the thrombin protease site replaced by a TEV protease site.	Burley[a]
pEG142	Km^r; ~5.2 kb containing Mlut_13230-13250 cloned into pEC-XK99E at KpnI & XbaI sites	This study
pEG174	Km^r; ~1kb containing Mlut_13230 #1 into pEC-XK99E at EcoRI and XbaI sites.	This study
pEG185	Km^r; ~1kb fragment of Mlut_13230 #2 cloned into pSKB3 at NdeI and SalI sites.	This study
pEG200	Km^r; ~1.2kb fragment of Mlut_09290 cloned into pEC-XK99E at EcoRI and XbaI sites.	This study
pEG205	Km^r; ~1kb fragment of Mlut_09310 cloned into pEC-XK99E at EcoRI and XbaI sites.	This study
pEG225	Km^r; ~3kb fragment of Mlut_13240 cloned into pEC-XK99E at XbaI site.	This study
pEG275	Km^r; ~1.1kb fragment of Mlut_13250 cloned into pEC-XK99E at XbaI and SbfI sites	This study

a Stephen K. Burley.

TABLE 2. Primers used in this study

Target genes	Primer name	Primer Sequence[a] (5'→3')
Primers used for target gene amplification		
Mlut_13230 #1	MLprimer_F1	TACTGAATTCGAAGGAGGGTCCTGGTGACGAACGTGT
	MLprimer_R1	TTCATCTAGACCCACTAGTTGGCTCCTTCAGC
Mlut_13230 – Mlut_13250	MLprimer_F2	AGACACTAGTAGGAGGATTGGTCCTGGTGACGAACG
	MLprimer_R2	ATCTCTAGAGTTTCCCGAACGAAAGCTC
Mlut_09290	MLprimer_F3	CATGGAATTCAGACCCAGAGGCAGCAGACC
	MLprimer_R3	GCCCACTAGTCTCCGGCTCAGACGCTGC
Mlut_09310	MLprimer_F4	CATGGAATTCGGCAGAGAGAGGCACCATGA
	MLprimer_R4	GCCCACTAGTGTACGTGGACAGTGAATCAGACGG
Mlut_13230 #2	MLprimer_F5	TCACATATGGTGACGAACGTGTCCGGCAA
	MLprimer_R5	AGTGTCGACTTGGCTCCTTCAGCTCACCA
Mlut_13240	MLprimer_F6	CTATACTAGTGCTCGAGATCGAATGGTGAGCTGAAGG
	MLprimer_R6	ATCATCTAGAGTCGAGGCCGACGTCGTAGCCGAAG
Mlut_13250	MLprimer_F7	GATTTCTAGACGCCGGCGGGAAGGTGGGTG
	MLprimer_R7	AAGTCCTGCAGGGGACGGGCGCTCGTTCCGGC

Primers used for real-time PCR		
Mlut_13230	qPCR_MLprimer_F1	CCTGATCAAGGACGGTCTCG
	qPCR_MLprimer_R1	CTGGTGGGTGATGAATCGGT
Mlut_13240	qPCR_MLprimer_F2	ACACCGACCAGCAGAGCAAG
	qPCR_MLprimer_R2	GTGGTGATCACGTGCTGGAG
Mlut_13250	qPCR_MLprimer_F3	AGTACGAGGCCGTGAACGTG
	qPCR_MLprimer_R3	GGGAGGAGACGTGGACGAAG

[a] Underlined sequences indicate restriction sites used for cloning purposes.

TABLE 3. Summary of long-chain alkenes and monoketones detected in *E. coli* heterologously expressing *M. luteus* genes[a]

Genes[b]	Strain	Alkenes produced				Monoketones produced			
		27:3[c]	27:2	29:3	29:2	27:2	27:1	29:2	29:1
Empty vector	EGS084								
13230	EGS180					+	+	+	+
13240	EGS244								
13250	EGS300								
13230 to 13250	EGS145	+	+	+	+	+	+	+	+
09290 (*fabF*)	EGS210								
09310 (*fabH*)	EGS212								

a + indicates that the compound was detected.

b For brevity, "Mlut_" has been removed from the locus tags (e.g., 13230 = Mlut_13230).

c Number of carbon atoms : Number of C=C bonds.

FIGURE LEGENDS

FIG. 1. Partial amino acid alignments of three translated *M. luteus* genes with homology to condensing enzymes involved in fatty acid biosynthesis: (A) Mlut_09290 (FabF), (B) Mlut_09310 (FabH), and (C) Mlut_13230. Alignments include the most similar sequences from *E. coli* and close relatives *Arthrobacter* sp. strain FB24 and *Arthrobacter aurescens* TC1. Three conserved active-site residues (see text) are highlighted: Cys-His-His (panel A) or Cys-His-Asn (panels B and C). Gray areas indicate sequence identity. GenBank accession numbers: (A) *E. coli* – NP_415613, strain TC1 – YP_948166, strain FB24 – YP_831948; (B) *E. coli* – NP_415609, strain TC1 – YP_948164, strain FB24 – YP_831946; (C) *E. coli* – NP_415609, strain TC1 – YP_947743, strain FB24 – YP_832433.

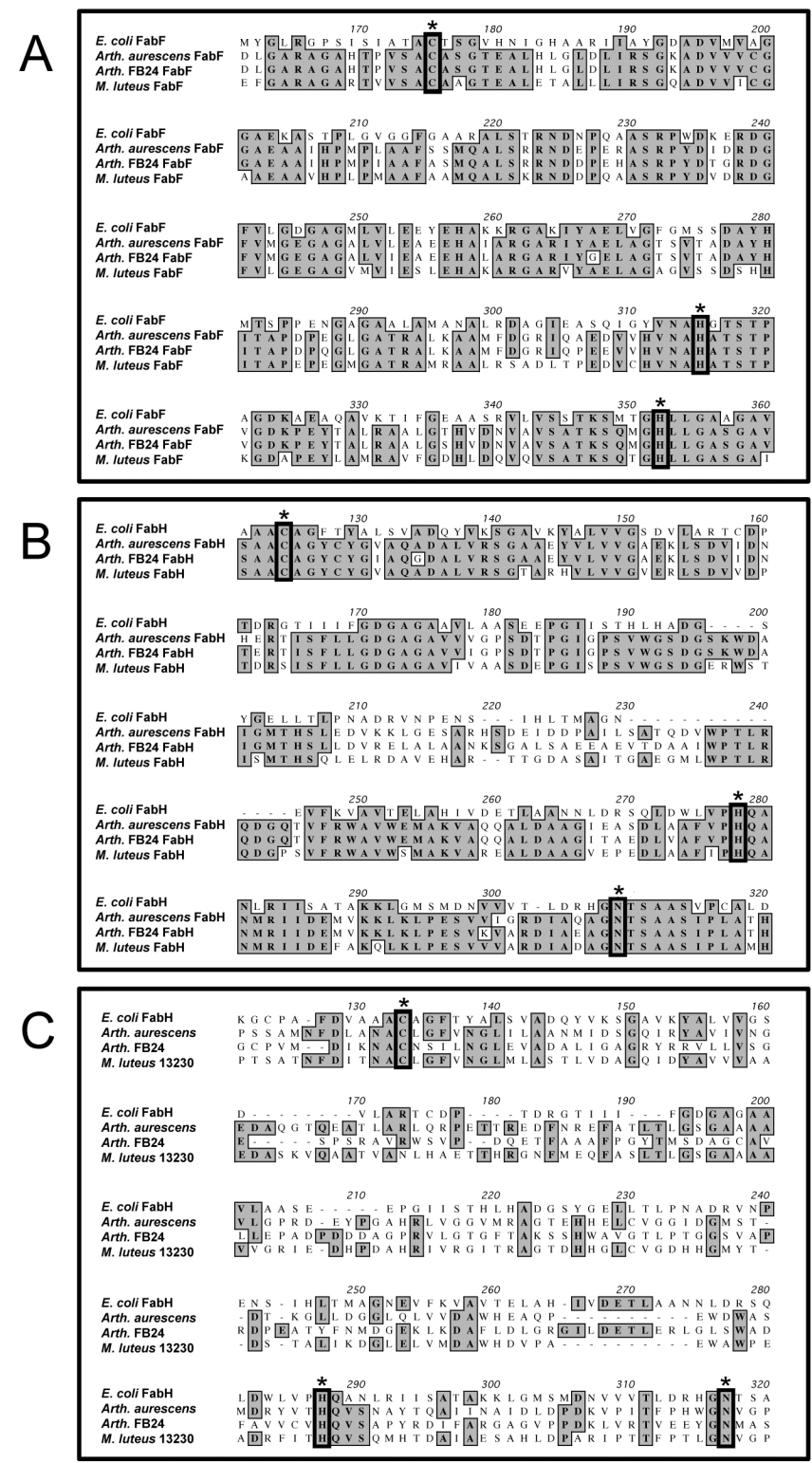
FIG. 2. (A) Total ion chromatograms of diazomethane-derivatized extracts of fatty acid-overproducing *E. coli* expressing Mlut_13230 (strain EGS180; blue) or no *M. luteus* genes (strain EGS084; black). Long-chain **ketones** (27:2, 27:1, 29:2, 29:1; blue fill) were observed when Mlut_13230 was expressed and were not observed in the control. (B) Total ion chromatograms of underivatized extracts of fatty acid-overproducing *E. coli* expressing Mlut_13230-13250 (strain EGS145; red), Mlut_13230 (strain EGS180; blue), or no *M. luteus* genes (strain EGS084; black). Long-chain **alkenes** (27:3, 27:2, 29:3, 29:2; red fill) were observed only when Mlut_13230-13250 were present, and were not observed with Mlut_13230 alone or in the negative control. There are peaks from strain EGS180 that co-elute with 29:3 and 29:2 alkenes, however, inspection of extracted ion profiles for the molecular ions of these alkenes (*m/z* 402 and 404) demonstrates that the alkenes are not present in strain EGS180 (insets).

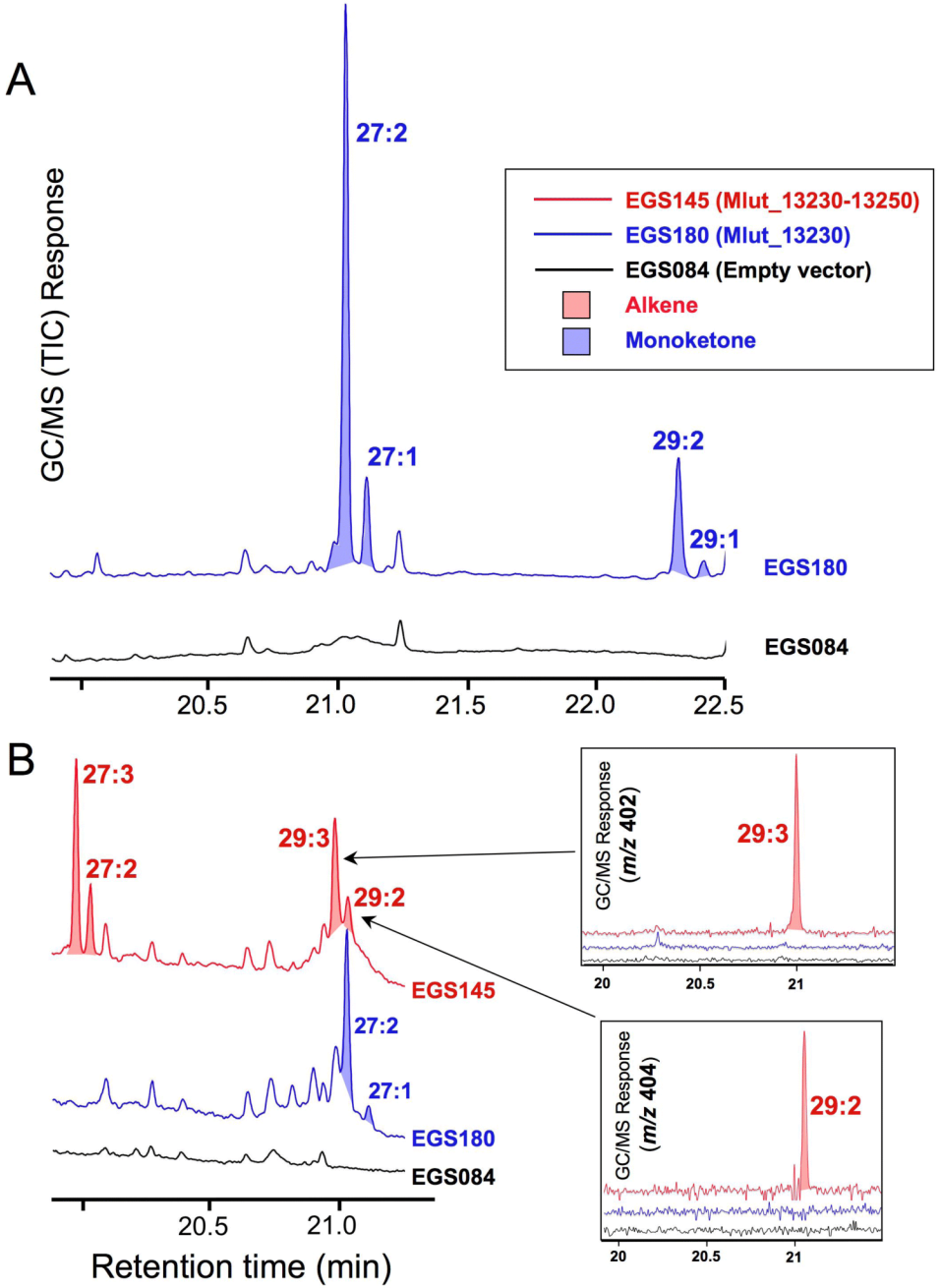
FIG 3. (A) 70-eV electron ionization mass spectra of the two unsaturated C_{27} monoketones (labeled 27:2 and 27:1) in Figure 2A. (B) 70-eV electron ionization mass spectra of the two C_{27} alkenes (labeled 27:3 and 27:2) in Figure 2B.

FIG. 4. Extracted ion chromatograms (*m/z* 291) of extracts from in vitro studies with purified Mlut_13230 protein. Duplicate results are shown for assays including Mlut_13230, tetradecanoyl-CoA, and crude lysate from wild-type *E. coli* DH1 (red); controls without DH1 lysate (blue); and controls without Mlut_13230 protein (black). The peak has the same retention time as the 27:2 monoketone observed during in vivo studies with Mlut_13230 (Fig. 2A), and *m/z* 291 is characteristic of that compound (Fig. 3A).

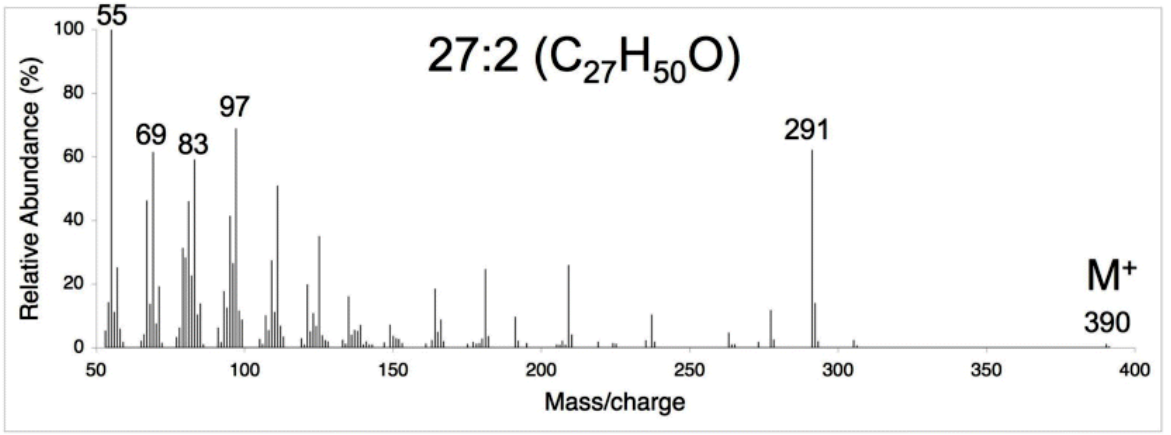
FIG. 5. Alkene production (A) and expression of alkene biosynthesis genes (B) through different growth stages of *M. luteus*. All variables are plotted as a percent of their maximum values and duplicate results are shown. Alkenes 1 and 2 (A) are 29:1 alkenes (see text); inset chromatograms showing the relative enhancement of alkene 2 over time are shown. In (B), results of RT-qPCR analysis of Mlut_13230, 13240, and 13250 over time are shown.

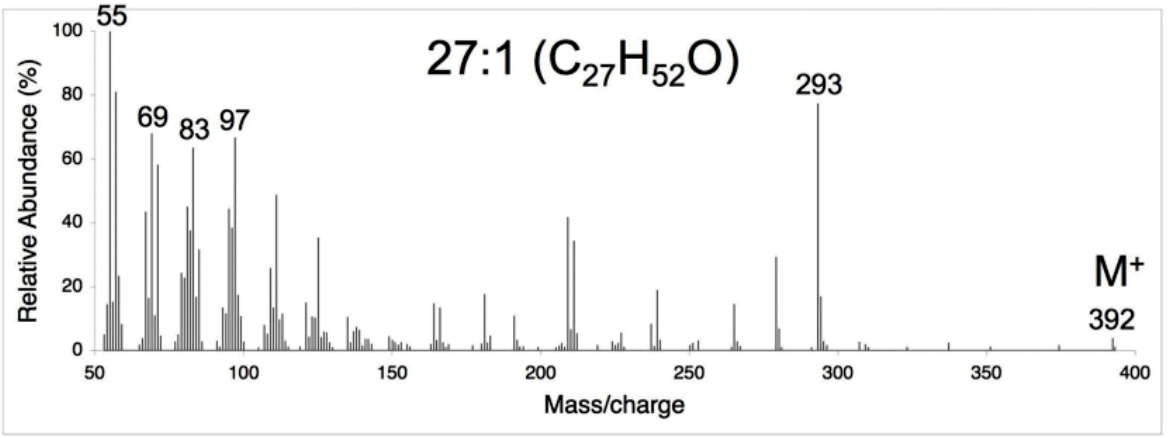
FIG. 6. Proposed pathway for alkene biosynthesis from condensation of fatty acids. Compounds shown as CoA thioesters may in fact be ACP thioesters. The unsaturated monoketones observed in this study (Fig. 2, 3, 4) correspond to the metabolite following the first dehydratase reaction. In *M. luteus*, the starting compounds are likely *iso*- and *anteiso*-branched C_{15} fatty acids and the predominant products are *iso*- and *anteiso*-branched C_{29} monoalkenes.


A
E. coli FabF
Arth. aurescens FabF
Arth. FB24 FabF
M. luteus FabF
B
E. coli FabH
Arth. aurescens FabH
Arth. FB24 FabH
M. luteus FabH
C
E. coli FabH
Arth. aurescens
Arth. FB24
M. luteus 13230

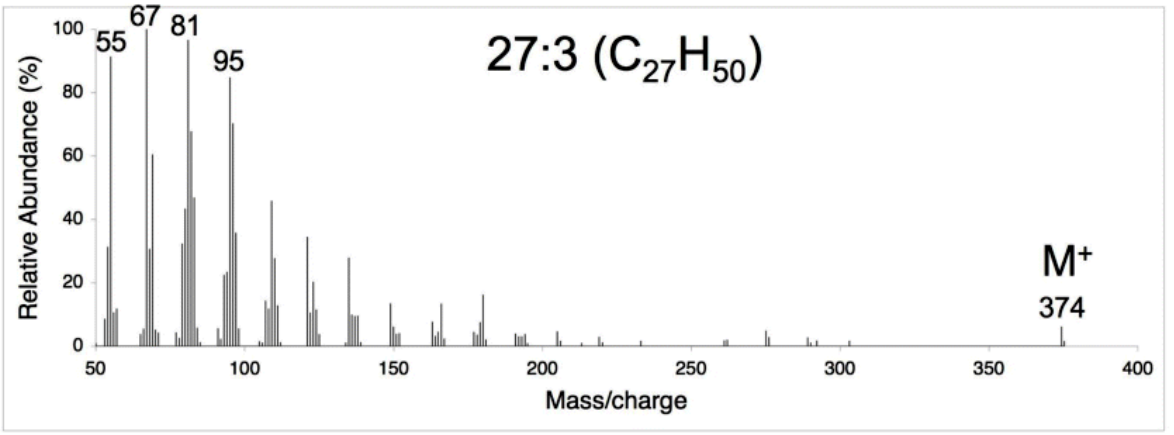

A
GC/MS (TIC) Response
27:2
27:1
29:2
29:1
EGS180
EGS084
EGS145 (Mlut_13230-13250)
EGS180 (Mlut_13230)
EGS084 (Empty vector)
Alkene
Monoketone
20.5
21.0
21.5
22.0
22.5
B
27:3
27:2
29:3
29:2
EGS145
27:2
27:1
EGS180
EGS084
20.5
21.0
Retention time (min)
GC/MS Response (m/z 402)
29:3
20
20.5
21
GC/MS Response (m/z 404)
29:2
20
20.5
21

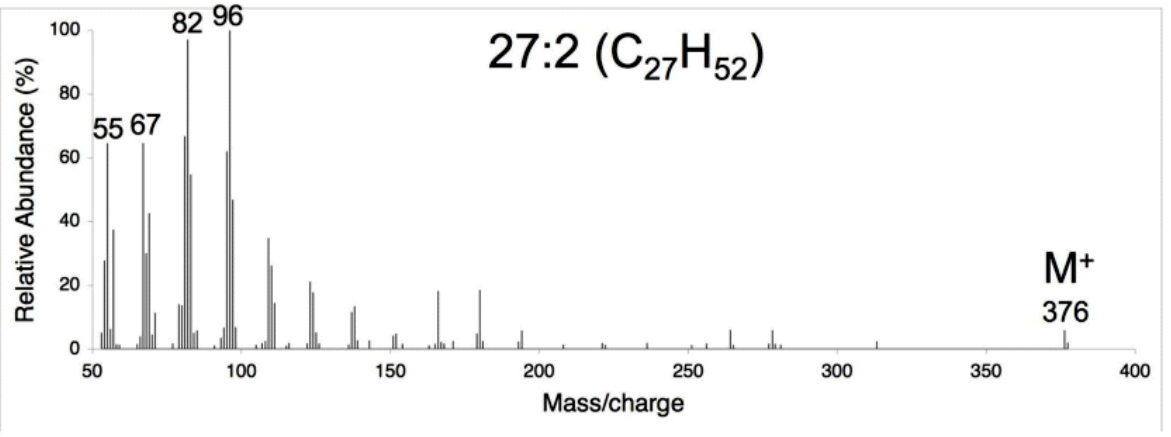
A

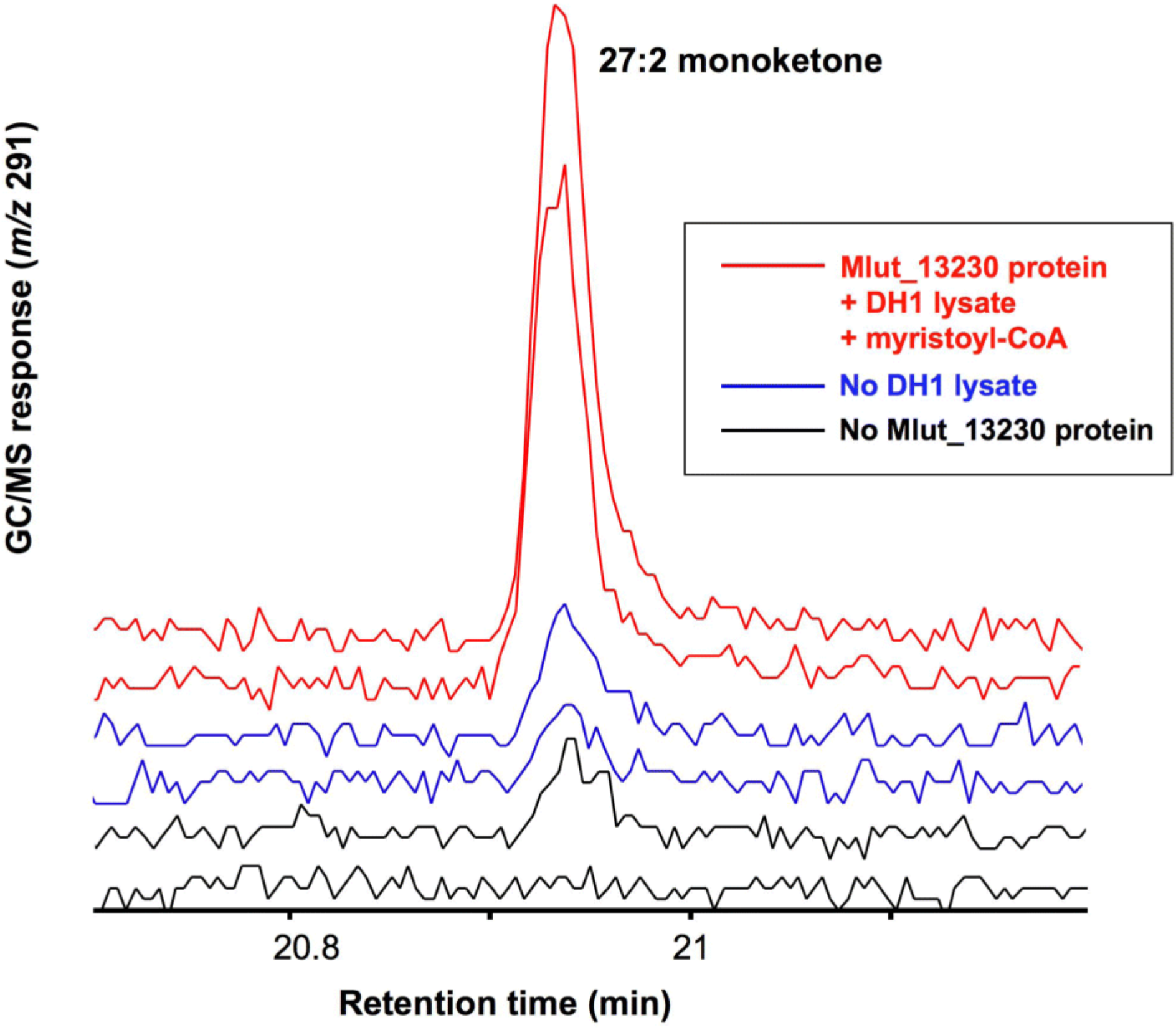

27:2 ($C_{27}H_{50}O$)
55
69
83
97
291
M+
390
Relative Abundance (%)
Mass/charge


27:1 ($C_{27}H_{52}O$)
55
69
83
97
293
M+
392
Relative Abundance (%)
Mass/charge

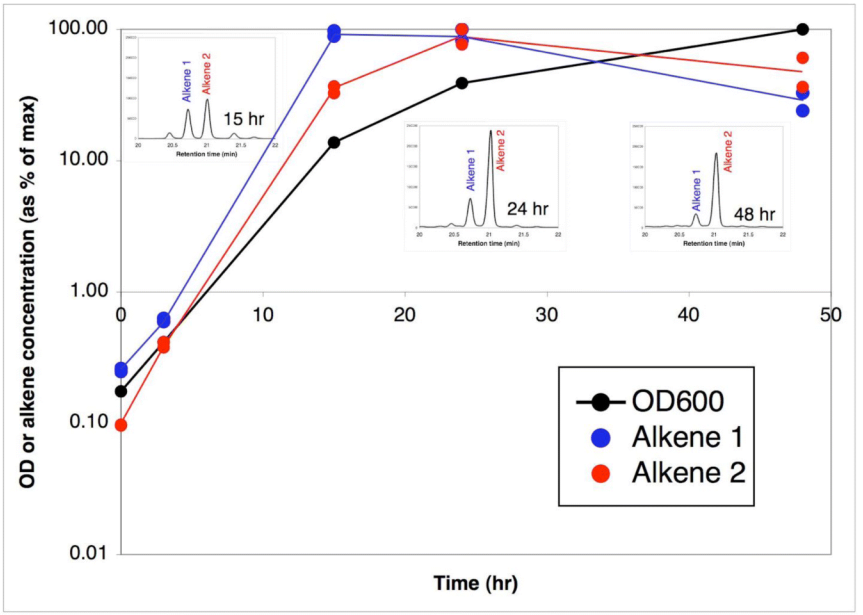
B


27:3 ($C_{27}H_{50}$)
55
67
81
95
M+
374
Relative Abundance (%)
Mass/charge

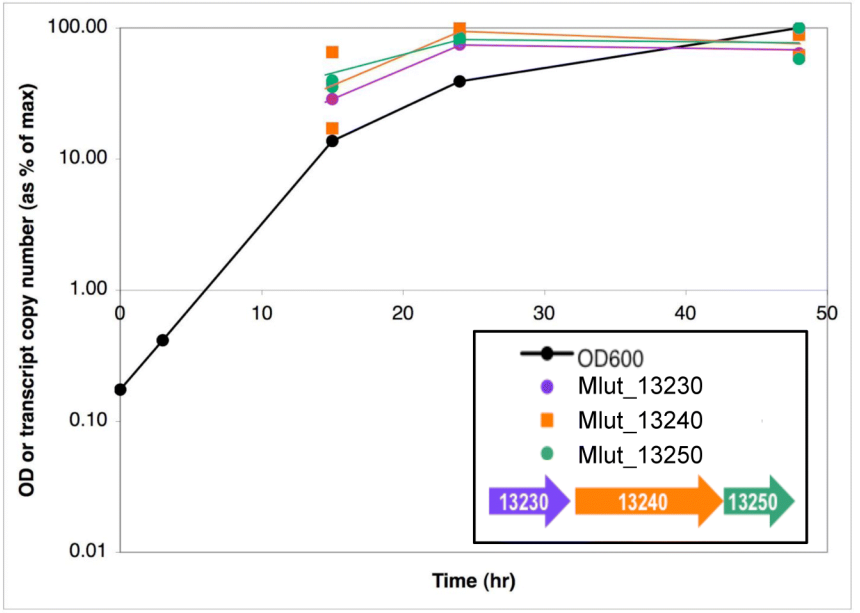

27:2 ($C_{27}H_{52}$)
55
67
82
96
M+
376
Relative Abundance (%)
Mass/charge

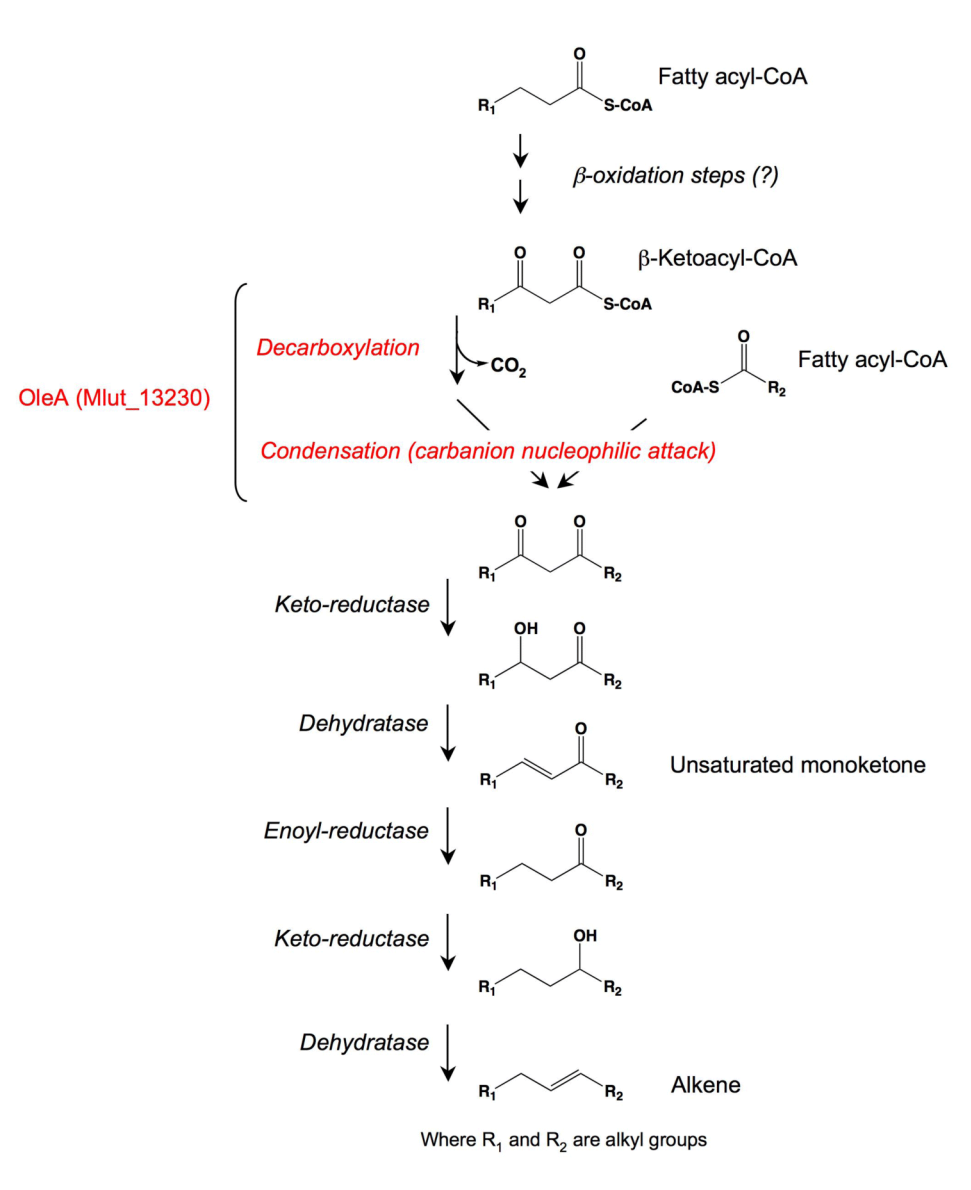

27:2 monoketone
GC/MS response (m/z 291)
Mlut_13230 protein
+ DH1 lysate
+ myristoyl-CoA
No DH1 lysate
No Mlut_13230 protein
20.8
21
Retention time (min)

A


100.00
10.00
1.00
0.10
0.01
OD or alkene concentration (as % of max)
0
10
20
30
40
50
Time (hr)
Alkene 1
Alkene 2
15 hr
24 hr
48 hr
Retention time (min)
OD600
Alkene 1
Alkene 2

B


100.00
10.00
1.00
0.10
0.01
OD or transcript copy number (as % of max)
0
10
20
30
40
50
Time (hr)
OD600
Mlut_13230
Mlut_13240
Mlut_13250
13230
13240
13250


Fatty acyl-CoA
β-oxidation steps (?)
β-Ketoacyl-CoA
Decarboxylation
CO2
Fatty acyl-CoA
OleA (Mlut_13230)
Condensation (carbanion nucleophilic attack)
Keto-reductase
Dehydratase
Unsaturated monoketone
Enoyl-reductase
Keto-reductase
Dehydratase
Alkene
Where R1 and R2 are alkyl groups