# Supporting Information

## Howard et al. 10.1073/pnas.1215966110

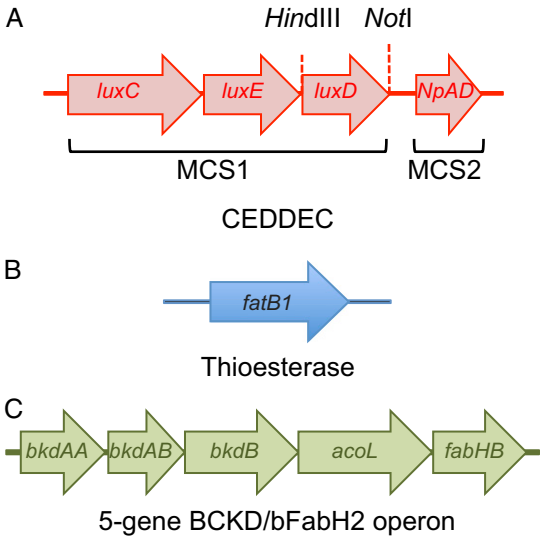


**Fig. S1.** Genetic constructs used in this study. (*A*) Graphical representation of the *luxC*, *luxE*, *luxD*, and the aldehyde decarbonylase gene (CEDDEC) construct used to introduce the alkane biosynthetic pathway (red in Fig. 1) into *Escherichia coli*. The amino acid sequences for LuxC, LuxE, and LuxD (accession AAD05355.1, AAD05359.1, and P19197.1, respectively) were reverse-translated and codon-optimized for expression in *E. coli* using DNA2.0 GeneGPS Technology and synthesized as a single operon by DNA2.0 into the pACYCDuet-1 expression vector (Novagen) multiple cloning site (MCS) 1. The restriction enzyme sites HindIII and NotI were used to remove the *luxD* gene as indicated in the results and discussion. The decarbonylase gene was cloned from genomic DNA extracted from *Nostoc punctiforme* using the FastDNA SPIN Kit (MP Biomedicals) following the manufacturer's instructions and purified further by phenol-chloroform extraction and precipitation using ethanol and sodium acetate. The gene encoding aldehyde decarbonylase (*Np*AD) was amplified using Phusion high-fidelity DNA polymerase using cyanobacterial genomic DNA as template. Primers used were 5′ CATATGCAGCAGCTTACAGACCAAT 3′ and 5′ CTCGAGTTAAGCACCTATGAGTCCGTAGG 3′ allowing direct cloning into MCS2 of either pACYCDuet-1 or pCDFDuet-1 using the NdeI and XhoI sites (underlined). *Np*AD sequence was confirmed by DNA sequencing (Geneservice) of purified plasmids. The *lux* operon sequence and *Np*AD sequence (collectively CEDDEC) were subsequently subcloned into pCDFDuet-1 MCS1 and MCS2 using NcoI and NotI (CED) and NdeI and XhoI (*Np*AD) sequences, respectively, for coexpression with thioesterase and branched-chain modules. (*B*) The thioesterase amino acid sequence Q39473.1 minus the chloroplast transit peptide sequence (blue in Fig. 1) was reverse-translated and codon-optimized for expression in *E. coli* using DNA2.0 GeneGPS Technology, synthesized by DNA2.0, and cloned into the MCS1 of pETDuet-1 using the NcoI and BamHI restriction enzyme sites. (*C*) Graphical representation of the five-gene branched-chain α-keto acid dehydrogenase (BCKD)/bFabH2 operon used to generate branched fatty acids (green in Fig. 1 and Fig. S2) in *E. coli*. All genes were reverse-translated and codon-optimized for *E. coli* from the *Bacillus subtilis* amino acid sequences. The amino acid sequences for BCKD components (NP_390285.1, NP_390284.1, NP_390283.1, and ZP_03600867.1) and for β-ketoacyl-ACP synthase III (KASIII) FabH2 (NP_388898.1) were reverse-translated and codon-optimized in silico using DNA2.0 GeneGPS Technology, synthesized by DNA2.0, and cloned into pETDuet-1 MCS1 using NcoI and NotI sites.

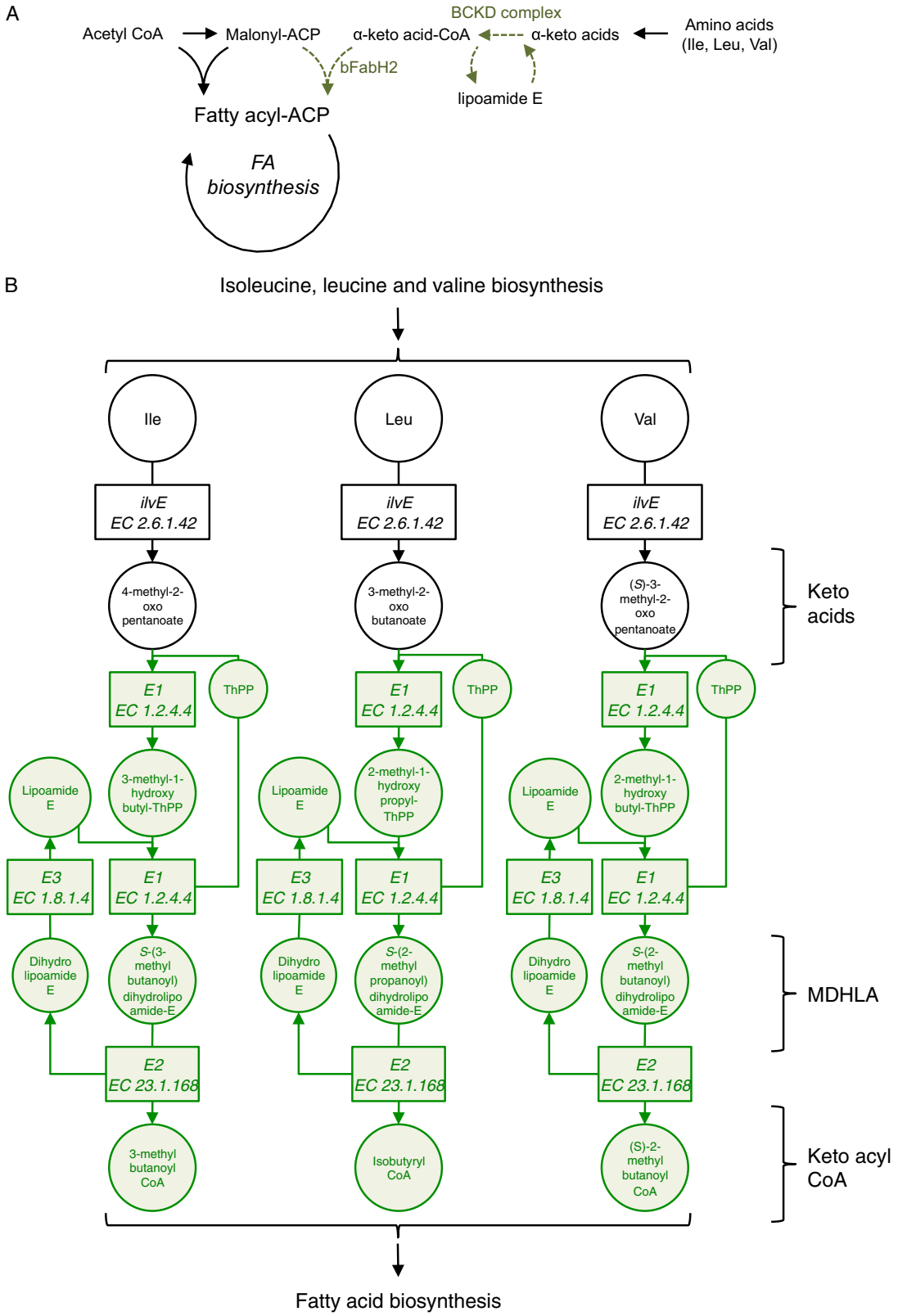


**Fig. S2.** Creation of branched fatty acid (FA) biosynthesis pathway in *E. coli*. (*A*) Overview of the metabolic changes required for the introduction of branched primers into the endogenous *E. coli* FA biosynthesis elongation cycle (green in Fig. 1). (*B*) Detail of the metabolic steps catalyzed by the enzymes of the BCKD protein complex [adapted from the valine, leucine, and isoleucine degradation pathway Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg) pathway 00280]. ACP, fatty acyl-acyl carrier protein; MDHLA, *S*-(3-methylbutanoyl)-dihydrolipoamide E, *S*-(2-methylpropanoyl)-dihdrolipoamide E, or *S*-(2-methylbutanoyl)-dihydrolipoamide E. *ilvE* encodes branched-chain amino transferase.

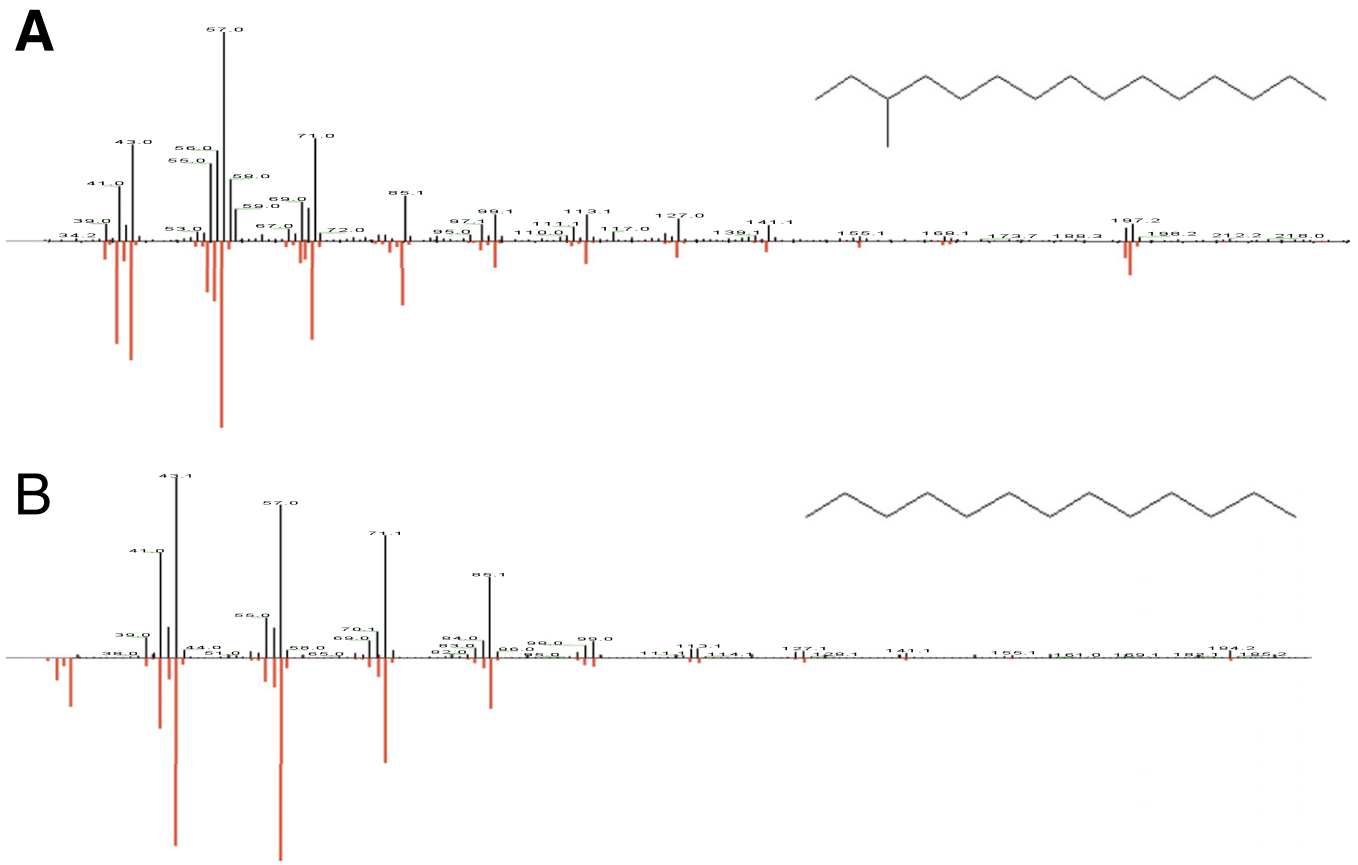


**Fig. S3.** Mass spectral data for methyl pentadecane and tridecane identified in Figs. 4 and 5, respectively. (*A*) The mass spectral data for methyl pentadecane (black) and for the 15:0* (Fig. 4*B*) peak (red). The structure of methyl pentadecane is inset. (*B*) The mass spectral data for tridecane (black) and for the 13:0 peak (red). The structure of tridecane is inset.

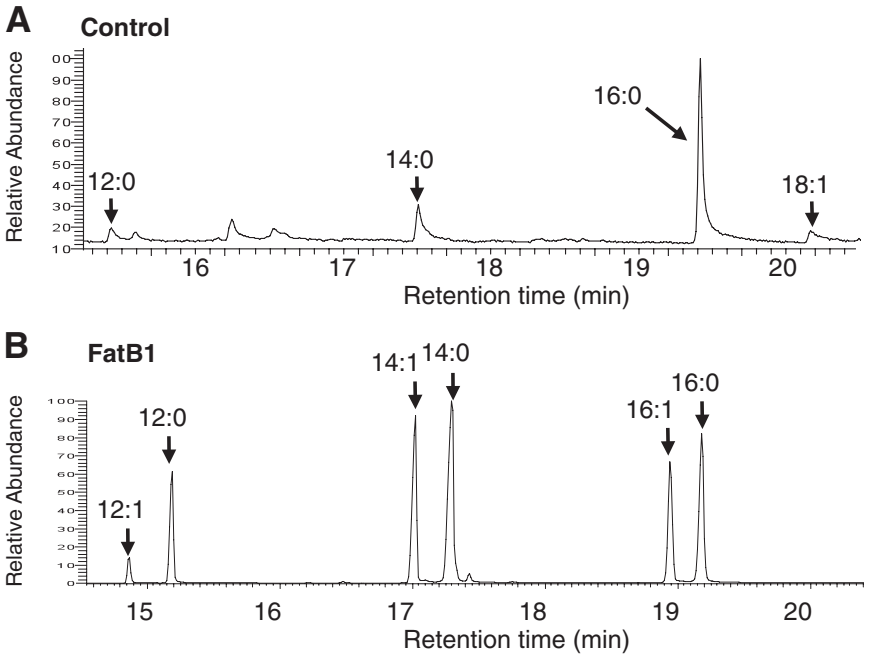


**Fig. S4.** Creation of a tetradecanoic acid substrate pool in *E. coli. E. coli* cells harboring the *fat*B1 thioesterase gene without transit peptide sequence from *C. camphor* in MCS1 of the expression vector pETDuet-1 were grown and induced as described in *Experimental Procedures*. (*A*) GC-MS total ion chromatograms (TIC) of FA extracted from control cells. (*B*) GC-MS TIC of FA extracted from cells expressing *fat*B1.

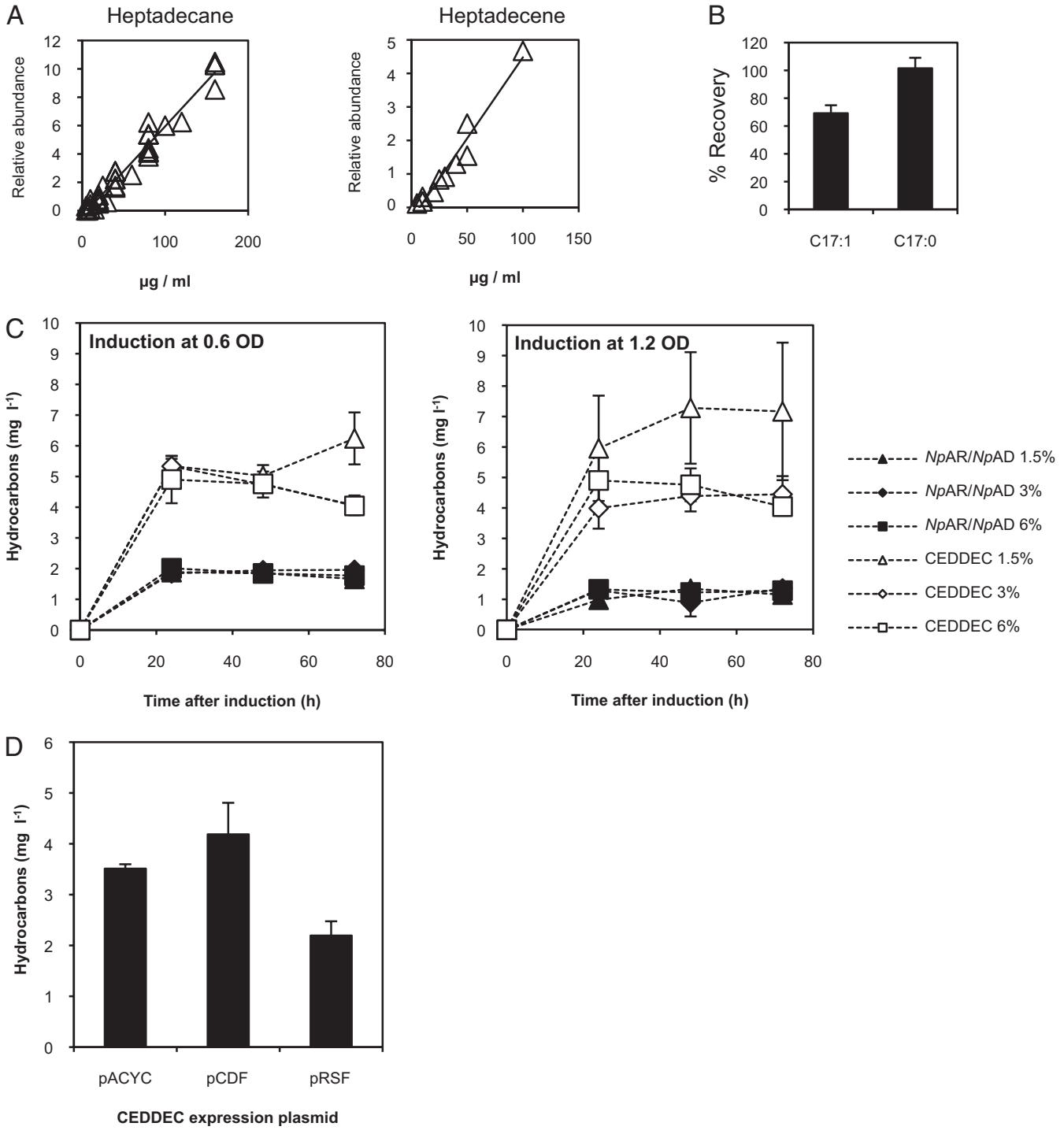


**Fig. S5.** Detection and optimization of alkane biosynthesis in BL21* (DE3) cells using the pETDuet expression system. (*A*) Heptadecane and heptadecene standards were linear down to 5 µg·mL$^{-1}$. (*B*) Standards were recovered at 100% and 70%, respectively, from samples spiked with 12.5 µg heptadecane or heptadecene. (*C*) Cells carrying either the cyanobacterial pathway [black symbols, reported previously (1)], or the artificial CEDDEC pathway described here (white symbols) were grown in 1.5% (triangle), 3% (diamond), or 6% glucose (square), and protein expression was induced at OD 0.6 (*Left*) or OD 1.2 (*Right*) with 20 µM isopropylthiogalactoside (IPTG). (*D*) Alkanes recovered from cells expressing the CEDDEC pathway from plasmids pACYC (low copy number), pCDF (medium copy number), or pRSF (high copy number).

1. Schirmer A, Rude MA, Li XZ, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329(5991):559–562.

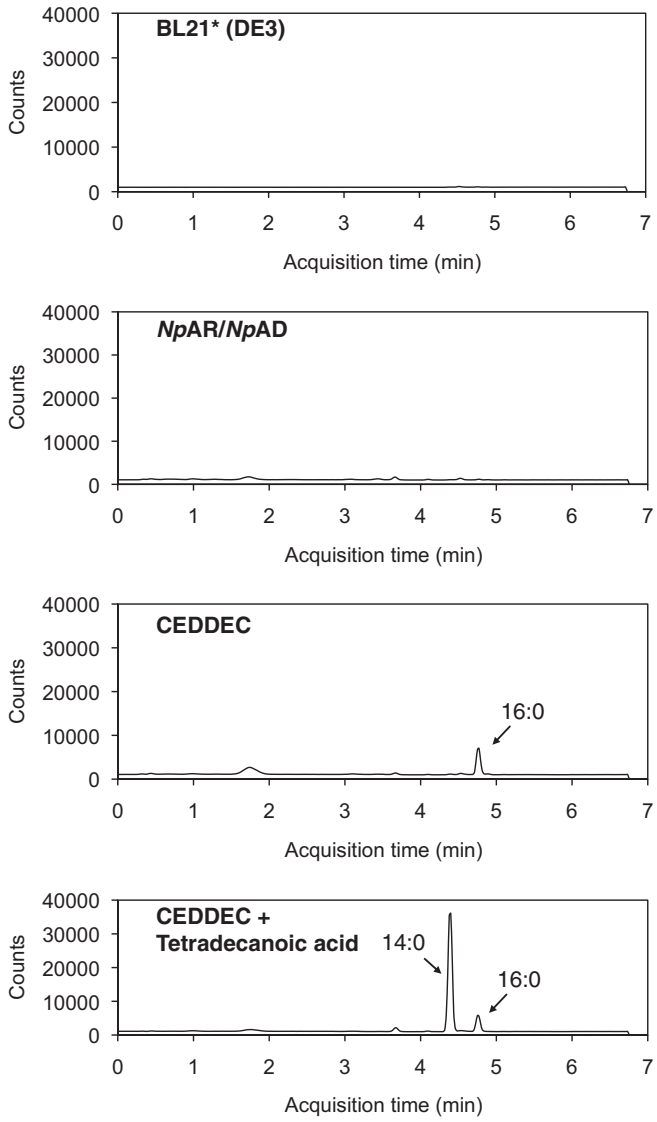


**Fig. S6.** Fatty aldehyde pools in *E. coli* cells. Liquid chromatography mass spectrometry (LCMS) TIC of fatty aldehydes (*Top* to *Bottom*) from BL21* (DE3) cells with no plasmid, cells expressing the cyanobacterial pathway, cells expressing the artificial CEDDEC pathway, and cells expressing the CEDDEC pathway in the presence of exogenous tetradecanoic acid at 100 $\mu g \cdot mL^{-1}$ supplementation. Cells grown in the presence of tetradecanoic acid but without induction did not reveal the presence of fatty aldehydes.