HHS Public Access

Author manuscript

ACS Synth Biol. Author manuscript; available in PMC 2019 September 21.

Published in final edited form as:

ACS Synth Biol. 2018 September 21; 7(9): 2205–2215. doi:10.1021/acssynbio.8b00215.

Highly active C8-acyl-ACP thioesterase variant isolated by a synthetic selection strategy

Néstor J. Hernández Lozada[a,d], **Rung-Yi Lai**[a,d], **Trevor R. Simmons**[a], **Kelsey A. Thomas**[a], **Ratul Chowdhury**[b], **Costas D. Maranas**[b], and **Brian F. Pfleger**[a,c]

[a]Department of Chemical and Biological Engineering; University of Wisconsin-Madison; 1415 Engineering Drive; Madison, WI, 53706

[b]Department of Chemical Engineering; Pennsylvania State University; 158 Fenske Laboratory; University Park, PA, 16802

[c]To whom correspondence should be addressed.

[d]These authors contributed equally to this work.

Abstract

Microbial metabolism is an attractive route for producing medium chain-length fatty acids, e.g. octanoic acid, used in the oleochemical industry. One challenge to this strategy is the lack of enzymes that are both highly active in a microbial host and selective towards substrates with desired chain-length. Of the many steps in fatty acid biosynthesis, the thioesterase is the most widely used enzyme for controlling chain length. Thioesterases hydrolyze the thioester bond between fatty acids and the acyl-carrier protein (ACP) or coenzyme A (CoA) co-factor. The functional role of thioesterases varies between organisms (i.e. bacteria vs. plant) and therefore so do the substrate specificities. As a result, microbial biocatalysts that utilize a heterologous thioesterase either produce high titers of fatty acids with mixed chain-lengths or low titers of products with a narrow chain-length distribution. To search for highly active enzymes that selectively hydrolyze octanoyl-ACP, we developed a genetic-selection based on the lipoic acid requirement of *Escherichia coli*. We used the selection to identify variants in a randomly mutagenized library of the C_8-specific *Cuphea palustris* FatB1 thioesterase. After optimizing expression of the thioesterase, *E. coli* cultures produced 1.7 g/L of octanoic acid with >90% specificity from a single chromosomal copy of this thioesterase. *In vitro* studies confirmed the mutant thioesterase possessed a 15-fold increase in k_{cat} compared to its native sequence. The high level of specific activity allowed for low levels of expression while maintaining fatty acid titer. The low expression requirement will allow metabolic engineers to use more cellular resources to address other limitations in the pathway and maximize overall productivity.

Corresponding Author: Tel: 608-890-1940. pfleger@wisc.edu.

Author Contributions: N.H.L, R.Y.L., and B.F.P. designed the experiments. N.H.L, R.Y.L., T.R.S. and K.A.T. performed the experiments. N.H.L, R.Y.L., R.C., C.D.M and B.F.P. evaluated the data and wrote the manuscript.

Note: The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXXX

Keywords

Escherichia coli; thioesterase; octanoic acid; selection; protein engineering; acyl-ACP

Oleochemicals are a large class of industrial chemicals used for making products in the bioenergy, plastics, surfactants, and personal care sectors. Oleochemicals include molecules such as free fatty acids (FFA), fatty acid methyl esters (FAME), fatty alcohols, and organosulfates (e.g. sodium dodecyl sulfate)[1,2]. While the most desirable oleochemicals contain medium chains (C_6-C_{12}), most natural oleochemical sources are dominated by long acyl chains (>C_{16}). Of the major oil seed crops, only coconut and palm kernel oils have large fractions of medium chain fatty acids (MCFA), with C_{12} being the most abundant chain length. Further, there are few high-volume sources of the desired C_8 and C_{10} fatty acids (Figure S1). Some plant oils contain high proportions of medium-chain acids, e.g. *Umbellularia californica* (California Bay laurel) contains high proportion of C_{12} fatty acids[3] and *Cuphea* species often contain high percentages of C_8 fatty acids[4], but most are not cultivated in volumes capable of meeting oleochemical demand. The low availability of MCFA has motivated the development of microbial conversion strategies where renewable sugars are converted to specific oleochemicals in the medium chain category[5]. In this approach, metabolic engineering principles are used to redirect carbon flux through fatty acid biosynthesis to specific products[6,7] leveraging heterologous enzymes capable of catalyzing desired biochemical transformations.

While many oleochemicals have been produced in microbes, there remains a dearth of enzymes capable of directing flux to products containing a specific chain length, i.e. C_8. The notable exception is the thioesterase[8] which cleaves acyl-thioesters (CoA or acyl-carrier protein, ACP) to release FFA from biosynthetic pathways. Thioesterases are expressed in many organisms for various purposes. Microbial thioesterases often have proofreading roles in the cell and therefore act on a broad substrate range. In some cases, thioesterase selectivity can be tailored via protein engineering[9,10,11,12,13], but complete selectivity and high activity remains an unmet challenge. In contrast, many plant thioesterases act on a narrow set of substrates[14]. Plants synthesize fatty acids in the chloroplast and lipids in the cytsol[15]. In order to transport acyl-chains across the chloroplast membrane, plants express thioesterases to release FFA in the chloroplast and reactivate them as acyl-CoA thioesters in the cytosol. Therefore, the substrate specificity of thioesterases often dictates the composition of plant oils. For this reason, plants have become a preferred source for isolating thioesterases with desired substrate preference. These enzymes can then be used in either transgenic crops or microbes to produce oleochemicals with desired chain lengths[7]. This approach is often made difficult in *Escherichia coli* by a loss of activity when plant thioesterases are heterologously expressed. Therefore, researchers remain motivated to isolate, evolve, and/or engineer improved thioesterases with desired selectivity and activity.

One challenge to thioesterase engineering is the lack of good screening methods to differentiate products with different chain-lengths. The analysis of fatty acid chain-length typically uses gas chromatography to separate fatty acid methyl esters (FAME) derived from biological samples. While accurate, this method requires considerable sample preparation

time and instrument time that limits the number of samples that can be processed to less than a 200 per day per instrument. In protein engineering projects, a library size typically ranges from 10^3 to 10^8 samples. Therefore, gas chromatography is not an applicable method for screening large libraries for increased activity. Without high throughput screens mutagenesis is limited to rational design. Although rational design of thioesterases for C_8 chains using computational methods[9] and domain swapping[12,13] has given some degree of success, there is still considerable room for improving these enzymes to be both highly active and specific for a single substrate. One alternative to achieving this goal is the development of biosensors that use screening (change in observable phenotype) or selection (live/dead) as a way to differentiate improved enzymes from the rest. Biosensors for detecting fatty acids and other aliphatic molecules have been developed by others using transcriptional regulators linked to fluorescent proteins[16] and G-protein coupled receptors[17,18]. However, these approaches have limited ability to tailor chain-length specificity. Here, we developed a genetic selection for acyl-chains containing exactly eight carbons using the lipoic acid requirement of *E. coli* under aerobic conditions.

Lipoic acid is an essential vitamin in most organisms. It is an important cofactor for function of several key enzymes involved in aerobic metabolism, such as pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, the glycine cleavage system, and the branched-chain 2-oxoacid dehydrogenase[19]. Pyruvate dehydrogenase contains a lipoyl group in its E2 domain that translocates an activated acetyl moiety to the thiol of coenzyme A to form acetyl-CoA. This lipoyl group synthesis proceeds via one of two pathways in *E. coli*. The endogenous biosynthesis pathway branches from a central intermediate in fatty acid biosynthesis octanoyl-ACP (Figure 1A) which donates its octanoyl group to the E2-domain of pyruvate dehydrogenase via the enzyme LipB (octanoyl-ACP *N*-lipoyltransferase). Next, LipA, an iron-sulfur cluster enzyme, catalyzes a radical reaction to install two sulfur atoms at C_6 and C_8 positions of the octanoyl group to complete this biosynthesis. *E. coli* Δ*lipB* mutants are unable to grow on minimal media. However, *E. coli* has a salvage pathway to obtain and assimilate free lipoic acid from the environment. LplA, a lipoate-protein ligase, is an ATP-dependent enzyme that activates lipoic acid as well as octanoic acid and transfers it to the E2-domain of pyruvate dehydrogenase. As seen in Figure 1A, growth of *E. coli* Δ*lipB* mutants can be rescued by supplying lipoic acid as well as octanoic acid in the media[20]. The authors proposed that LplA could transfer octanoic acid to the E2 domain of pyruvate dehydrogenase and rejoin the native lipoic acid biosynthesis pathway where a LipA-catalyzed reaction completes the pathway. Conversely, other chain-length carboxylic acids were not able to rescue growth of the mutant (not shown). Based on these intriguing results, we hypothesized that an *E. coli* Δ*lipB* strain could be leveraged as a growth-based biosensor for octanoic acid presence.

In this manuscript, we describe how we used this novel screening approach to select for improved variants from a randomly mutagenized library of *Cuphea palustris* FatB1 thioesterase (CpFatB1) genes[21]. The best variants led to 5–7 fold improvements in octanoic acid titer while sustaining the enzyme's high selectivity towards 8-carbon chains. The best variant CpFatB1-M4 demonstrated a 15-fold increased k_{cat}. The increased specific activity enabled us to place the gene on the chromosome in a single copy and achieve the same octanoic acid titer achieved by plasmid containing strains.

Results and Discussion

Establishment of a baseline thioesterase

Cuphea palustris FatB1 thioesterase (CpFatB1) is highly selective for C8:0-ACP when expressed in *E. coli*[14] albeit with lower activity relative to *E. coli* 'TesA and other commonly used thioesterases[6,8,23–27]. The lower activity likely comes from a combination of poor expression and/or poor specific activity. Plant thioesterases are often associated with the chloroplast membrane and native genes contain membrane localization sequences. When heterologously expressed, these sequences can lead to insoluble or aggregated proteins. Therefore, one must construct a N-terminal truncation of a plant thioesterase to obtain high levels of soluble protein. We constructed three N-terminal truncations of CpFatB1 based on prior work[28] and sequence alignment-CpFatB1.2, CpFatB1.3, and CpFatB1.4. Each was cloned into a high copy plasmid, pTRC99a (Figure S2, Table S1) and transformed into *E. coli* RL08ara (Δ*fadD*) for testing. Each strain was grown in LB supplemented with 0.4% glycerol and 1 mM IPTG. Expression of CpFatB1.2 generated the highest titer of octanoic acid under these conditions. Therefore, we used this gene sequence as a starting point in our mutagenesis studies.

Development of a lipoic acid-based selection

As discussed above, *E. coli* requires small amounts of lipoic acid to enable pyruvate dehydrogenase activity under aerobic conditions. As little as 50 μM (7.2 mg/L) octanoic acid can restore growth of an *E. coli* Δ*lipB* strain[29]. This amount is less than the ~200 mg/L of octanoic acid produced from the plasmid-based CpFatB1.2 described above. Therefore, to use the lipoic acid requirement as a selection, the overall activity of the thioesterase must be reduced, such that the baseline enzyme cannot complement a Δ*lipB* mutation. To do this we reduced the expression of CpFatB1.2 by swapping the promoter for a weaker P_{araBAD} and moved the expression cassette to a plasmid maintained at a lower copy number (pACYC origin). Unfortunately, *E. coli* Δ*lipB* pBAD33-CpFatB1.2 grew on MOPS-minimal media-agar[30] containing 0.2% arabinose to induce expression (Figure S3). In a second attempt to reduce activity, we cloned the original pTrc-CpFatB1.2 expression cassette onto a low-copy plasmid, pBTRCK (pBBR1 origin), and used a series of low IPTG concentrations to vary expression (Figure S4A). Interestingly, cells grown on plates with 20 μM IPTG took an extra day to grow (4 days), compared to cells grown (3 days) on the same media with 30 μM IPTG (Figure S4A). Given this difference in growth rates, we hypothesized that we had found a window in which cells expressing thioesterase variants with improved specific activity or mutations that increased protein production would be identified before the cells carrying the parent gene became visible colonies. Therefore, we used the low-copy pBTRCK plasmid and 20 μM IPTG in our mutagenesis study.

Library of CpFatB1.2 mutants

To introduce mutations, we generated a library of CpFatB1.2 variants by error-prone PCR covering the full coding sequence. PCR products were cloned into pBTRCK, plasmids were transformed into *E. coli ΔlipB*, and cells were plated on MOPS minimal media containing 20 μM IPTG. Hundreds of colonies appeared after three days and ninety were picked for further studies (Figure S4B). To validate each hit, plasmids were isolated from each colony,

retransformed into fresh *E. coli ΔlipB* cells, and cells were grown under selecting conditions. All of the variants rescued growth within three days (not shown), indicating that growth-conferring mutations were plasmid-based. When cultured in liquid media, cells harboring the variant thioesterases produced more octanoic acid than cultures expressing the original CpFatB1.2 (Figure 1B). In particular, variants CpFatB1.2-M3, CpFatB1.2-M4, and CpFatB1.2-M9 exhibited 4-fold, 5.3-fold and 2.3-fold improvements, respectively, when expressed from the low-copy plasmid. After analyzing the first 10 mutants, we repeated the protocol on the remainder of the 90 putative mutants, finding several additional improved variants, but none superior to M3 or M4 (Figure S5, Table S2).

Plasmids isolated from each of the hits that generated more than a 2-fold increase in octanoic acid were sequenced. A small family of mutations was observed in these hits. Interestingly, one mutation, D293V appeared independently in five of the 20 sequenced mutants, CpFatB1.2-M20, CpFatB1.2-M40, CpFatB1.2-M47, CpFatB1.2-M66, and CpFatB1.2-M73. Mutant CpFatB1.2-M40 and CpFatB1.2-M66 contained only the D293V mutation, indicating that it provided on average a 2.3-fold increase in activity over CpFatB1.2. In addition to two point mutations (N28S, I65M), CpFatB1.2-M4, the best variant, contained a frame-shifting deletion which introduced a premature stop codon. CpFatB1.2-M3 contained two mutations (A59S and K296R) that were also found in other mutants. Given the superior performance of CpFatB1.2-M3, CpFatB1.2-M4, and CpFatB1.2-M9, we focused the remainder of the study on these variants.

Mutants CpFatB1.2-M3, CpFatB1.2-M4, and CpFatB1.2-M9 were subcloned into high copy plasmid pTRC99a to determine if the improvements found under screening conditions would be maintained under optimal production conditions. Plasmids were transformed into *E. coli* RL08ara (Δ*fadD*) and cells were grown in MOPS media enriched[22] with tryptone, yeast extract, and 1 mM IPTG to maximize induction. Cells expressing the M3 and M4 variants produced 1751 mg/L and 1263 mg/L of octanoic acid respectively. These titers represent a 3–4 fold-increase relative to cells expressing CpFatB1.2 which produced 375 mg/L (Figure 1D). Cells expressing the M9 variant produced approximately 500 mg/L, a smaller relative value to CpFatB1.2 than seen under screening conditions. Conveniently, the increased activity did not come at the expense of octanoic acid selectivity as a percentage of MCFA produced. Variants M3 (94 %mol) and M4 (94 %mol) generated equivalent if not larger percentages of octanoic acid as a percentage MCFA compared to CpFatB1.2 (92 %mol MCFA). Variant M9 produced 91% octanoic acid. These data show that the lipoic acid selection was capable of identifying useful mutations.

Characterization of M4 variant *in vivo*

The CpFatB1.2-M4 variant contained two point mutations and an early nucleotide deletion (ΔA_{54}, N28S, I65M) that led the original open reading frame to an early stop codon (Figure 2A). Since CpFatB1.2-M4 is active, we suspected that translation was restarting at a different in-frame methionine using a suboptimal ribosome binding site (RBS). If true, we hypothesized that the M4 variant could generate more activity, if expressed with an optimal RBS. To determine which protein was being made, we cloned five in-frame CpFatB1.2-M4 variants based on the next five in frame methionines as start codons, creating CpFatB1.2-

M4–287, −288, −289, −290 and −291. These genes were cloned into a pBTRCk vector to position the new start adjacent to the original RBS. Plasmids harboring each variant were individually transformed into RL08ara (*ΔfadD*) strain and cells were grown in LB supplemented with 0.4% glycerol and 1 mM IPTG. Only variant CpFatB1.2-M4–287 showed a significant level of octanoic acid production (Figure 2B). Moreover, variant CpFatB1.2-M4–287 generated more octanoic acid than the isolated variant CpFatB1.2-M4. In addition, we added a histidine tag to the ptrc99a-CpFatB1.2-M4 construct and purified the protein in a Nickel column. We performed Electrospray Ionization Mass analysis on this sample and identified truncations M4–287, M4–288, and M4–289 as being the only present translated sequences on the construct. Out of these, M4–289 was a major component and 287 and 288 were in lower amounts (see Figure S6). This suggests that the hypothesis of CpFatB1.2-M4 being translated from a non-specific RBS was correct.

Further, the CpFatB1.2-M4–287 demonstrated a ~20-fold increase in octanoic acid production under the low expression conditions tested (compared to 4-fold for M3 and 7.5-fold for M4 in the same conditions) (Figure 2B, Table S1), suggesting that more activity could be obtained if overexpressed. Therefore, we cloned CpFatB1.2-M4-287 onto a high copy plasmid, ptrc99a, transformed the plasmid into *E. coli* RL08ara (*ΔfadD*), and cultured cells in MOPS media enriched[22] with tryptone and yeast extract (see methods) and different concentrations of IPTG to optimize expression (Figure 2C). The resulting fatty acid profiles extracted from these cultures showed that a subsaturating concentration of 50 μM IPTG gave maximum activity. At saturating concentrations of 1 mM IPTG, we observed a growth defect and a drastic decrease in C_{16} fatty acid species (Figure 2C). This data strongly suggested that we had dramatically increased the specific activity of CpFatB1, because it could no longer be maximally expressed. The same relationship was observed when titrating expression of the FatB1 from *U. californica* with plasmid copy number[24] or induction[31]. Impaired growth was connected to reduced membrane fitness and increased cell permeability caused by the production of dodecanoic acid[32].

Finally, we made combinations of the constitutive mutations found in CpFatB1.2-M4–287 (N28S, I65M, 287-truncation) to determine which contributed to enhanced activity. CpFatB1.2 variants containing 1, 2 or all three mutations were cloned into pTRC99A and cultured in *E. coli* RL08ara as described for Figure 2C. As a negative control, we cloned a variant with a H224A mutation that renders the enzyme catalytically inactive. Interestingly, only I65M was observed to increase activity above the baseline variant, CpFatB1.2 (Figure 2D). Only the triple mutant was able to drastically outperform the CpFatB1.2. This also suggests a possible explanation for the CpFatB1.2-M4–288 truncation, which excludes the N28S mutation being less active than the CpFatB1.2-M4 (Figure 2B).

Characterization of M4–287 variant *in vitro*

Our *in vivo* data suggested that the M4 variant had increased specific activity towards C_8-acyl ACPs. To prove this hypothesis, we measured the reaction rate *in vitro* using Ellman's reagent (DTNB) to monitor release of free thiols in holo-ACP (Figure 3A). As hydrolysis of octanoyl-ACP occurs, holo-ACP formed reacts with DTNB forming the colored compound TNB that absorbs light at 412 nm. For substrate, we synthesized octanoyl-ACP *in vitro* from

apo-ACP (purified from *E. coli)*, coenzyme A (which donated the 4'-phosphopantetheine prosthetic group to convert apo-ACP to holo-ACP), and octanoate, using methods described elsewhere[33]. Complete synthesis of octanoyl-ACP was confirmed by HPLC (Figure 3B). Using the DTNB assay, we measured the initial rate of the thioesterase reaction for a range of substrate concentrations. We found that mutant CpFatB1.2-M4–287 has dramatically improved activity towards octanoyl-ACP compared to both CpFatB1.2 as well as a TesA variant (TesA-R3.M4) that we designed computationally in previous work[9] to produce octanoic acid. The major contribution to CpFatB1.2-M4–287 improvement observed was due to a 15.7 fold increase in k_{cat} over CpFatB1.2 while the K_m remained relatively low (Figure 3C–D). Interestingly, we found that a TesA-R3.M4 variant developed in prior work[9] had a high k_{cat} as well but very low affinity for octanoyl-ACP with K_m 17-fold higher than CpFatB1.2-M4–287. These data confirm that our lipoic acid selection isolated a variant with improved specific activity.

Optimizing expression of M4–287 in *E. coli*

When building stable, industrially-relevant strains, it is beneficial to remove any requirement for antibiotics for maintaining plasmids and to reduce the cellular burden associated with protein overexpression. In other words, it is preferable to achieve a desired activity by increasing specific activity of essential enzymes such that each enzyme can be expressed at a modest level. Here, we wanted to test the ability of CpFatB1.2-M4–287 to provide thioesterase activity when expressed from low copy plasmids or the chromosome. Therefore, we created the low copy plasmid, pBTRCK-CpFatB1.2-M4–287, with an optimized RBS and tested FFA production under various induction levels (Figure 4A). CpFatB1.2-M4–287 was optimally expressed from this construct when no IPTG was added, suggesting that the copy number could be decreased further. Moreover, at high induction we again observed a growth defect, which reduced the final octanoic acid titers (Figure S7). Next, we took the same construct and inserted it into the *E. coli* chromosome in the *fadD* locus using CRISPR-Cas9 mediated homologous recombination. This yielded *E. coli* strain NHL17 (Figure 4B, Table S2). As can be seen from Figure 4B, a single copy CpFatB1.2-M4–287 in the chromosome when induced to 50 μM IPTG was sufficient to yield the levels of production obtained from the plasmid. Moreover, OD data of both the low copy plasmid and the chromosomally expressed CpFatB1.2-M4–287 shows that there is a growth defect associated with high expression (Figure S7), a phenomenon seen before[24]. Since in absolute terms both cases (low copy plasmid and single chromosomal copy) are relatively low expression, we hypothesize that this growth defect can be to depletion of the natural acyl-ACP pool in the cell, slowing down membrane synthesis. In addition, the conditions that yielded similar production levels also matched in growth (0 μM IPTG from pBTRCK-CpFatB1.2-M4–287 plasmid and 50 μM IPTG in NHL17) (**Figure A-B,** S7). When compared side by side, a single chromosomal copy of CpFatB1.2-M4–287 performs several fold higher than CpfatB1.2 in a high copy plasmid (promoter, RBS and induction level remained the same), both in rich media (Figure 4C) and minimal media (Figure 4D).

Structural modeling insights

It has been experimentally shown that the CpFatB1.2-M4–287 mutant exhibits 15-fold higher specific activity *in vivo* (Figure 2B) relative to the natural parent enzyme. The

CpFatB1.2-M4–287 enzyme variant consists of N28S, I65M double mutation and an eighteen residue N-terminal truncation. Each of these alterations in isolation or in pairs only marginally contributed to the higher activities observed when all three were combined (Figure 2D). To evaluate how these mutations impacted activity, we constructed a computational model of the mutant and parent enzymes. These were prepared by homology modeling using the published structure of *U. californica* UcFatB1 (PDB: 5X04[11]) as a template. Unfortunately, we could not predict the structure of the N-terminus of CpFatB1.2-M4–287 as the twelve N-terminal residues of M4–287 were not conserved with the crystallized protein and no empirical 3D template structure could be identified with >30% sequence identity to this region. This means that our CpFatB1.2 model structure has a 30 residue N-terminus truncation relative to CpFatB1.2 and thus excluding the important N28S mutation and the truncation introduced in variant M4–287. On the other hand, the model was able to project where the I65M mutation resided relative to the active site (Figure 5A–Figure B). Ile65 was positioned at one end of the acyl-binding crevice farthest from the crevice opening, where the catalytic residues line the periphery of the opening. We hypothesize that the bulkier side chain of the methionine is occluding the crevice end introducing steric clashes with the omega-1 acyl carbon of C_{12}-ACP, the preferred substrate of the template (PDB: 5X04) thioesterase. This residue is thus seemingly important for altering C_8-specificity, not activity. However, Mendonça *et al.*[34] show that mutations to amino acids connected to an active site residue by three or less non-covalent interactions can have a high impact on enzyme activity, a feat consistent with I65M according to the contact map for this residue (Figure 5C).

The CpFatB1.2-M3 also exhibited elevated enzymatic activities (~4 folds higher than wild-type *in vivo*). Our CpFatB1.2 computational model with the thirty-amino acid truncation at the N-terminus was used as the starting point to generate the variant model. Unlike CpFatB1.2-M4 we could capture the effects of both A59S and K296R mutations in this model. We hypothesize K296R has indirect and A59S has direct effects on enzyme activity based on the number of non-covalent bonds that connect these residues to one or more catalytic residues. Lys296 is a surface residue that has its side chain facing away from the acyl-binding pocket. However, a K296R mutation introduces a stable salt bridge interaction (~3.2 Å) between the positively charged N-atom of Arg296 side chain and side chain O-atom of Glu254 (Figure 6). Prior to mutation the orientation of the Lys296 side chain did not result in a salt-bridge (~5.3 Å apart). Contact map (Figure 6C) reveals Lys296 needs at least five non-covalent interactions to reach an active site residue which is shorter than that before the K296R mutation, but is less likely to have as much effect as I65M from CpFatB1.2-M4 (separated by 3 non-covalent interactions). However, CpFatB1.2-M3 accounts for its increase in activity by the A59S mutation which is adjacent to a Asp220 (a key catalytic residue). The side chain OH of Ser59 is linked to the backbone O-atom of Asp220 (~4.5 Å) (Figure 6C). The aliphatic methyl-side chain of Ala59 before mutation prevented any such polar contacts.

The CpFatB1.2-M4–287 and CpFatB1.2-M3 contact maps reveal the importance of Met69 (see Figures 5C and 6C) in connecting the mutated residues to the active site residues as it is the most connected node in the map. It is noteworthy that Met69 maintains these contact networks by mostly hydrophobic interactions higher up the cascade and polar interactions

lower down. The polar interactions are controlled by the backbone O-atom whereas the hydrophobic ones require the side chain C-atoms. We hypothesize that Met69 is in all probability important for ensuring a connected contact map for both these mutants and altering M69 to a charged amino acid or a small-side chain hydrophobic residue can significantly reduce the enhanced activity of the CpFatB1 mutants.

Using a lipoic acid selection, we isolated a mutated octanoyl-ACP thioesterase capable of high rates of hydrolysis while maintaining >90% specificity towards C_8 acyl chains. A cell harboring a single chromosomal copy of this thioesterase gene is capable of achieving the same high level of production observed from high-copy plasmids overexpressing the parent enzyme. Under the conditions tested we demonstrated a more than 3-fold improvement over the highest reported octanoic acid titers in the literature[5]. Moreover, our data suggests that thioesterase activity is no longer a limiting factor towards synthesizing C_8 compounds. Future work focusing on other considerations such as upstream pathway engineering and toxicity effects could yield higher octanoic acid titers and yields. This work also highlights the potential for optimizing the flux from octanoic acid to desired 8-carbon products with other chemical functionalities.

Materials and Methods:

Chemicals, Reagents, and Media

Chemicals were purchased from either Sigma Aldrich (St. Louis, MO) or Fisher Scientific (Waltham, MA). Oligonucleotides and gene fragments were purchased from Integrated DNA Technologies (Coralville, IA) or Thermo Fisher Scientific (Waltham, MA). Enzymes were purchased from New England Biolabs (Ipswich, MA). DNA purification kits were purchased from Qiagen (Venlo, Netherlands). All cultures were started from single colonies grown on LB agar isolated from freezer stocks stored in 15% glycerol. Overnight cultures of strains were grown in LB media at 37°C in a rotary shaker at 250 r.p.m. When a selective pressure was necessary to for plasmid retention, media was supplemented with the appropriate antibiotics (carbenicillin, 100 μg/mL; kanamycin, 50 μg/mL; chloramphenicol, 34 μg/mL).

DNA synthesis and cloning

Escherichia coli K12 MG1655 was used to create the Δ*lipB* selection strain. Here we used a CRISPR-Cas9 assisted homologous recombination protocol, modified from Li et al.,[35] to delete the *lipB* protein coding sequence. Standard lambda red recombination[36] was used to introduce the deletion and Cas9 guided to *lipB* was used to destroy unmodified chromosomes. The repair template contained 30 bases upstream and downstream of the lagging strand of *lipB* (Table S3).

NHL17 (*E. coli* K12 MG1655 Δ*araBAD* Δ*fadD::trc-CpFatB1.2-M4–287*) strain was created from *E. coli* K12 MG1655 Δ*araBAD* in the same manner by using a linear piece of dsDNA containing *lacI-trc-CpFatB1.2-M4–287* between 500 base pairs of homology upstream and downstream of *fadD* protein coding sequence (Table S3).

All plasmids made were constructed using Gibson Assembly of PCR products[37].

Random mutagenesis of CpFatB1.2

The CpFatB1.2 library was constructed by error-prone PCR following the manufacturer's instruction (GeneMorph II, Agilent). The mutation frequency chosen was low (0 to 4.5 mutations/kb). The plasmid backbone was amplified by PCR using high fidelity polymerase Phusion. CpFatB1.2 library was assembled with designated backbones by Gibson assembly method. Primers used in for the creation of the library contained the start and stop codons in order to prevent mutations on them.

Lipoic acid selection

In order to find suitable conditions for the Δ*lipB*-based selection method, purified plasmid pBTRCK- CpFatB1.2 was transformed into *E. coli* Δ*lipB* strain and plated in MOPS minimal media agarose plates containing 0.2% glucose. In addition, the plates contained kanamycin to maintain the plasmid and different IPTG concentrations (0 μM, 10 μM, 20 μM, 30 μM and 50 μM IPTG) to titrate the amount of CpFatB1.2 present in the cells. The 20 μM induction condition was chosen because at this level of CpFatB1.2 expression, the cells needed an extra day (4 days) for rescuing growth compared to higher induction levels (3 days).

Gibson assembly reaction mixtures (2 μL) containing a CpFatB1.2 library was transformed into 100 μL of electrocompetent *E. coli* Δ*lipB*. Following electroporation (1 mm cuvette, 2500 mV), 900 μL of fresh LB was added and cells were allowed to recover for 1.5 hours. In order to remove any remaining lipoic acid from the rich media, the cells were washed 3 times by spinning down at 9,400 g for 1 min. One milliliter of M9 minimal media was added and the cells were then resuspended. Finally, the washed cells were plated on MOPS minimal media-agarose plates containing 0.2% glucose and Kanamycin. The plates were incubated at 30°C and after 3 days, the chosen 90 putative mutants growing colonies were streaked on LB plates with Kanamycin. The plates were incubated at 37°C overnight to confirm the colony growth. Finally, the plasmids were purified and sequenced.

Fatty acid production

Plasmid based expression of thioestereases was performed in *E. coli* RL08ara[38] transformed with the appropriate plasmid. NHL17 strain contains a chromosomal copy of CpFatB1.2-M4–287 and therefore no plasmid was added.

For validating the 90 CpFatB1.2 variants (Figure 2C, Figure S3A, Table S2) as well as the truncations of CpFatB1 (Figure S1B) and truncations in CpFatB1.2-M4 (Figure 3B), isolated plasmids were transformed into *E. coli* RL08ara and single colonies were used to inoculate overnight cultures in LB media with kanamycin. Overnight cultures were used to inoculate 50 mL LB media with 0.4% glycerol and kanamycin at an initial optical density (OD_{600}) of 0.05. Cultures were incubated at 37°C with shaking in 250 mL shake flasks. When cultures reached an OD_{600} of 0.2–0.3 cells were induced with IPTG and moved to 30°C for 24hr. Mutants M1–M10 were tested using 20 μM IPTG (Figure 2C). Mutants M3, M4, and M11-M90 (Figure S3), truncations of CpFatB1 (Figure S1B) and truncations in CpFatB1.2-M4 (Figure 3B) were induced with 1 mM IPTG.

For experiments designed to test for high octanoic acid production (Figures 2E, 3C, 5A, 5B) overnight cultures as described above were used to inoculate 50 mL of medium described in Kim, et al.[22] with the following changes: 1.39 mM Na_2HPO_4, no biotin, thiamine or sodium selenite added.

Minimal media experiments were carried out in MOPS minimal media[30] containing 1% glucose and 0.240 mM K_2HPO_4 in order to create phosphate limiting conditions[39].

Fatty acid extraction and quantification

After 24 h post-induction, 2.5 mL of culture was transferred to 10 mL glass centrifuge tubes. 50 μL of 12.5 mg/mL nonanoic acid, and 1.25 mg/ml pentadecanoic acid in ethanol solution was added as an internal standard. The nonanoic acid internal standard was used to quantify octanoic acid and the pentadecanoic acid internal standard was used to quantify C_{10}-C_{18} chain lengths. Extraction and methylation process followed protocols described previously[9].

Protein expression and purification of Apo-Acyl Carrier Protein (ACP)

E. coli K12 MG1655 acyl carrier protein (ACP) was cloned into the pET28t vector system fused to a N-terminal polyhistidine tag coding the following peptide: MGSSHHHHHHSSENLYFQGGGG. The plasmid was transformed into BL21 (DE3) competent cells and grown LB media at 37°C until OD_{600} was 0.6–0.8. Cells were cooled to 18°C in ice water, induced with 1 mM IPTG and incubated overnight at 18°C with shaking. Cells were harvested by centrifugation at 8,000 × g and pellets were stored at −80°C for later use. Frozen pellets were resuspended in lysis buffer (50 mM Na_2HPO_4 pH8, 20 mM imidazole, 300 mM NaCl and 10% glycerol), sonicated, centrifuged at 12,000 RPM and filtered to clear the lysate. ACP was purified by Ni-NTA column following the manufacturer's instruction (GE Healthcare Life Sciences). To the ACP protein solution, Tev protease was added at a molar ratio of 1:20 and dialysed against 50 mM Tris, pH 7.5 overnight. Cleaved ACP was then passed through the Ni-NTA column to remove Tev protease and the His tag peptide. The flow through was dialyzed against 50 mM Na_2HPO_4 pH8, 10% glycerol for subsequent functionalization. The concentration of ACP was quantified via BCA assay (Thermo Fisher) using manufacturer's instructions.

Protein expression and purification of *Vibrio harveyi* AasS and *Bacillus subtilis* SfP, CpFatB1.2 and CpFatB1.2-M4

Vibrio Harveyi AasS, *Bacillus subtilis* SfP, CpFatB1.2 and CpFatB1.2-M4 were cloned into pET28t vector system with an N-terminal poly-histidine tag as described for ACP. Proteins were purified as described for ACP with the exception that no Tev protease reaction was performed. Following purification in Ni-NTA column, proteins were concentrated and buffer-exchanged into 50 mM Na_2HPO_4 pH 8, 30% Glycerol. Concentration of these proteins was quantified using the following extinction coefficients (280nm): 67520 M^{-1}cm-1 for AasS, 30620 $M^{-1}cm^{-1}$ SfP, 56295 $M^{-1}cm^{-1}$ for CpFatB1.2, and 50795 $M^{-1}cm^{-1}$ for CpFatB1.2-M4.

Synthesis of octanoyl-ACP

Octanoyl ACP synthesis was carried out by first functionalizing a 500 μM mixture of apo-ACP and holo-ACP from *E. coli* into holo-ACP by incubating at 37°C for 1 hr with 5 μM purified SfP, 10 mM $MgCl_2$, 5 mM Coenzyme A in 100 mM Na_2HPO_4 pH 8 as has been described elsewhere[33]. Next, 5 μM of purified AssS, 10 mM ATP and 5 mM sodium octanoate are added to the reaction mixture and incubated overnight at 37°C. Samples were taken in between steps for characterization by HPLC. After incubation, octanoyl-ACP was passed through a Ni-NTA column to remove both AasS and SfP followed by addition of an equimolar amount of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) in order to react all the CoA remaining prior to the assays. DTNB and yellow TNB produced in this step were subsequently dialyzed out against 100 mM Na_2HPO_4, pH 8, 10% glycerol before carrying out the enzymatic assays. Octanoyl-ACP concentration was quantified using BCA assay.

Liquid chromatography of octanoyl-ACP

To verify the functionalization and purity of Acyl carrier protein species, samples were separated via HPLC using a Harmony C4 column 2.1 × 150 mM, 3.5 μm (ES Industries). Mobile phases consisted of (1) aqueous solution of 0.05% (w/v) Trifluoroacetic acid and 0.05% (w/v) formic acid, and (2) 0.05% (w/v) Trifluoroacetic acid and 0.05% (w/v) formic acid in acetonitrile. The samples were separated over 20 min by imposing a gradient of 20% aqueous mobile phase to 98% acetonitrile mobile phase. The oven temperature was kept at 30°C and the flow rate was 0.2 ml/min with an injection volume of 10 μL. Prior to injection, the samples were buffer exchanged into 50 mM ammonium acetate and treated with 0.1% (w/v) formic acid

In-vitro analysis of CpFatB1.2-M4

Octanoyl-ACP thioesterase activity of CpFatB1.2, CpFatB1.2-M4 and TesA-R3.M4[9] was analyzed *in vitro* by tracking the formation of holo-ACP using the thiol-dependent reduction of 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB).TNB formation was monitored every 10 s for 2 min at Absorbance at 412 nm with a NanoDrop 2000c (Thermo Scientific) at a path length of 10 mm. Octanoyl-ACP was added to the assay in concentrations ranging 0–400 μM (quantified via BCA assay). The conditions for the assay were as follows: 40 nM thioesterase, 8 μg/mL BSA, 250 μM DTNB, 100 mM phosphate buffer pH 7.4, in1 mL reaction volume. Assay was started with the addition of the thioesterase. All concentrations except 400 μM were tested in triplicate.

Structural Modeling of CpFatB1.2-M3 and CpFatB1.2-M4

The CpFatB1 model was created using homology modeling of the CpFatB1 sequence and BTE (PDB:5×04 structure) as the template structure. Subsequently, the amino acid changes to create energy-minimized structures of the CpFatB1.2 mutants were made using Mutator[40]. The catalytic residues Asp220, Asn222, His224, Glu258, and Cys259 were identified using the *Umbellularia californica* thioesterase UaFatB1 (BTE) structure (PDB: 5×04) as a guide where the analogous residues have been reported to be Asp281, Asn283, His285, Glu319, and Cys320[11]. Cys320 mutants were seem to retain non-negligible catalytic activities, hence Cys259 was excluded from the list of catalytic residues in CpFatB1 model.

The octanoyl-ACP (substrate) was docked such that the carbonyl carbon (C=O) of the thioester bond of the acyl-ACP molecule was close to the side-chain O-atoms of Asp220 and Glu258. The catalytic distances corresponding to Asp220 and Glu258 were measured to be 3.5 Å and 3.7 Å, respectively. Subsequently, the BTE structure was used to identify the acyl-binding pocket residues which are important for controlling substrate specificity[11]. In order to understand the biophysical mechanism that underpins the catalytic activity in each of the enzyme variants, non-covalent contact maps were constructed similar to Mendonça *et al.*[34]. These contact networks (see Figures 5a and 6a) have been used to explain the possible path by which mutations away from the binding crevice or active site could affect enzyme activity. Python 2.7 scripts were written to identify hetero-atoms of residues within 6 Å of an altered residue of the CpFatB1 mutant. Each of these residues form nodes of the contact map and edges are drawn to show that a non-covalent contact exists. The same procedure is repeated for the nodes identified in the previous step and a cascade of interactions is mapped. The process is terminated when one or more catalytic and acyl-binding pocket residues are identified within 6 Å of a residue identified in previous step. Subsequently, we parse the obtained information to classify the non-covalent contacts (edges) as hydrophobic or polar, depending on the nature of atoms involved in the interaction. Finally, we visualize the contacts using PyMOL visualizing software.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

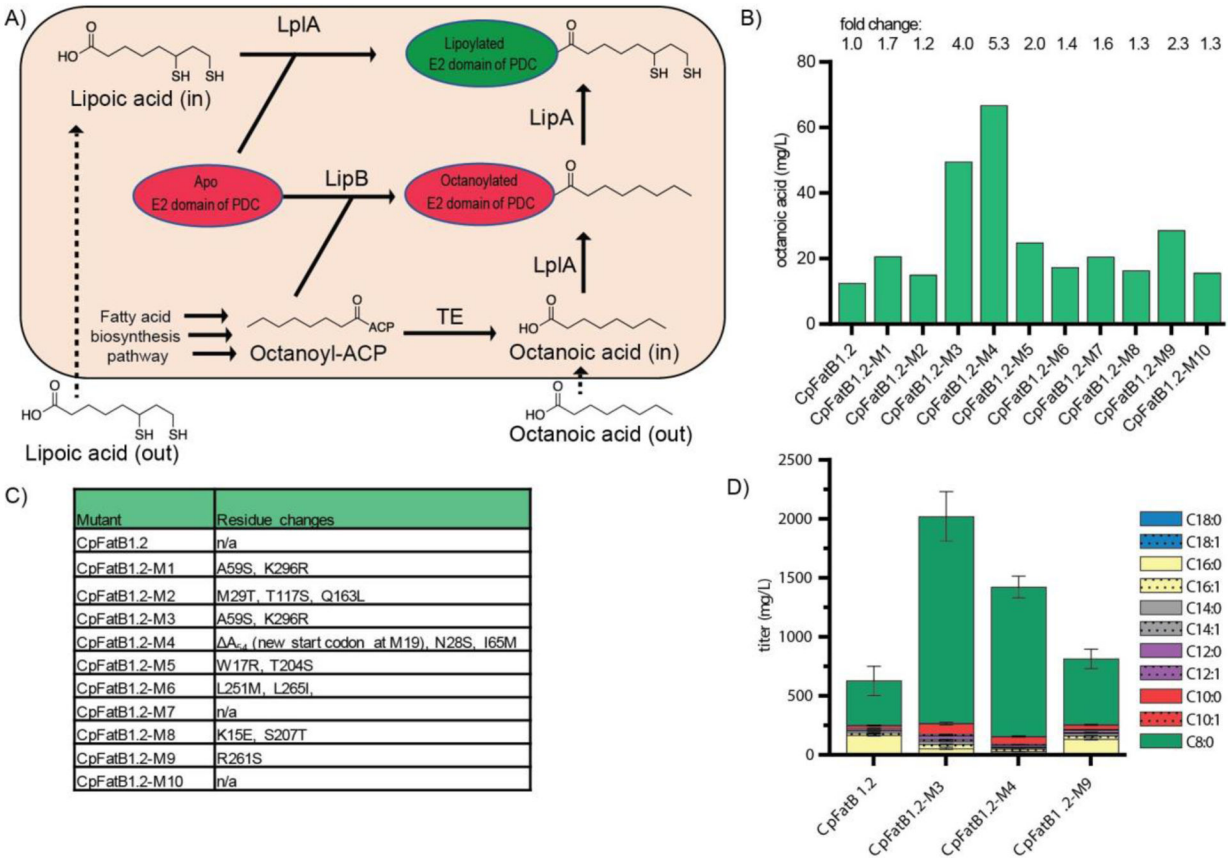
Support for this work was provided by the National Science Foundation (CBET-1703504) and Dow Chemical through a project grant to B.F.P. N.H.L. is the recipient of a NIH Chemistry-Biology Interface Training Program fellowship (No. T32 GM008505) and a Graduate Engineering Research Scholars fellowship from the UW−Madison College of Engineering. The authors would like to acknowledge the helpful discussions with Dr. Devon Rosenfeld and Dr. Christopher Stowers during the course of the project and the assistance of Dr. Travis Korosh in the octanoyl-ACP liquid chromatography analysis.

References:

1. Biermann U, Bornscheuer U, Meier MAR, Metzger JO & Schäfer HJ Oils and Fats as Renewable Raw Materials in Chemistry. Angew. Chemie Int. Ed 50, 3854–3871 (2011).
2. Lennen RM & Pfleger BF Microbial production of fatty acid-derived fuels and chemicals. Curr. Opin. Biotechnol 24, 1044–53 (2013). [PubMed: 23541503]
3. Voelker T. a & Davies HM Alteration of the specificity and regulation of Fatty Acid Synthesis of Escherichia coli by Expression of a Plant Medium- Chain Acyl-Acyl Carrier Protein Thioesterase. J. Bacteriol 176, 7320–7327 (1994). [PubMed: 7961504]
4. Graham SA, Hirsinger F & Robbelen G Fatty Acids of Cuphea (Lythraceae) Seed Lipids and Their Systematic Significance. Am. J. Bot 68, 908–917 (1981).
5. Sarria S, Kruyer NS & Peralta-Yahya P Microbial synthesis of medium-chain chemicals from renewables. Nat. Biotechnol 35, 1158–1166 (2017). [PubMed: 29220020]
6. Lennen RM & Pfleger BF Engineering Escherichia coli to synthesize free fatty acids. Trends Biotechnol. 30, 659–67 (2012). [PubMed: 23102412]
7. Pfleger BF, Gossing M & Nielsen J Metabolic engineering strategies for microbial synthesis of oleochemicals. Metab. Eng 29, 1–11 (2015). [PubMed: 25662836]
8. Cantu DC, Chen Y, Lemons ML & Reilly PJ ThYme: a database for thioester-active enzymes. Nucleic Acids Res. 39, D342–6 (2011). [PubMed: 21045059]

9. Grisewood MJ et al. Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. ACS Catal. 3837–3849 (2017). doi:10.1021/acscatal.7b00408 [PubMed: 29375928]
10. Yuan L, Voelker TA & Hawkins DJ Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. 92, 10639–10643 (1995).
11. Feng Y et al. Structural Insight into Acyl-ACP Thioesterase toward Substrate Specificity Design. ACS Chem. Biol 12, 2830–2836 (2017). [PubMed: 28991437]
12. Jing F, Zhao L, Yandeau-Nelson MD & Nikolau BJ Two distinct domains contribute to the substrate acyl chain length selectivity of plant acyl-ACP thioesterase. Nat. Commun 9, 860 (2018). [PubMed: 29491418]
13. Ziesack M et al. Chimeric Fatty Acyl-ACP Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression. Appl. Environ. Microbiol. AEM02868–17 (2018). doi:10.1128/AEM.02868-17
14. Jing F et al. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochem. 12, 44 (2011). [PubMed: 21831316]
15. Benning C Mechanisms of lipid transport involved in organelle biogenesis in plant cells. Annu. Rev. Cell Dev. Biol 25, 71–91 (2009). [PubMed: 19572810]
16. Zhang F, Carothers JM & Keasling JD Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat. Biotechnol 30, 354–9 (2012). [PubMed: 22446695]
17. Mukherjee K, Bhattacharyya S & Peralta-Yahya P GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. ACS Synth. Biol 4, 1261–1269 (2015). [PubMed: 25992593]
18. Sarria S, Bartholow TG, Verga A, Burkart MD & Peralta-Yahya P Matching Protein Interfaces for Improved Medium-Chain Fatty Acid Production. ACS Synth. Biol 7, 1179–1187 (2018). [PubMed: 29722970]
19. Cronan JE, Zhao X & Jiang Y Function , Attachment and Synthesis of Lipoic Acid in Escherichia coli Advances in Microbial Physiology 50, (Elsevier Masson SAS, 2005).
20. Zhao X, Miller JR, Jiang Y, Marletta MA & Cronan JE Assembly of the Covalent Linkage between Lipoic Acid and Its Cognate Enzymes. Chem. Biol 10, 1293–1302 (2003). [PubMed: 14700636]
21. Dehesh K, Edwards P, Hayes T, Cranmer AM & Fillatti J Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of Cuphea palustris seed oil. Plant Physiol. 110, 203–210 (1996). [PubMed: 8587983]
22. Kim S, Clomburg JM & Gonzalez R Synthesis of medium-chain length (C6–C10) fuels and chemicals via β-oxidation reversal in Escherichia coli. J. Ind. Microbiol. Biotechnol 42, 465–475 (2015). [PubMed: 25645093]
23. Steen EJ et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463, 559–62 (2010). [PubMed: 20111002]
24. Lennen RM, Braden DJ, West R. a, Dumesic J. a & e BF. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. Biotechnol. Bioeng 106, 193–202 (2010). [PubMed: 20073090]
25. Dellomonaco C, Clomburg JM, Miller EN & Gonzalez R Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. Nature 476, 355–9 (2011). [PubMed: 21832992]
26. Lu X, Vora H & Khosla C Overproduction of free fatty acids in E. coli: implications for biodiesel production. Metab. Eng 10, 333–9 (2008). [PubMed: 18812230]
27. Liu T, Vora H & Khosla C Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. Metab. Eng 12, 378–86 (2010). [PubMed: 20184964]
28. Torella JP et al. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. Proc. Natl. Acad. Sci. U. S. A 110, 11290–5 (2013). [PubMed: 23798438]
29. Zhao X, Miller JR, Jiang Y, Marletta MA & Cronan JE Assembly of the Covalent Linkage between Lipoic Acid and Its Cognate Enzymes. Chem. Biol 10, 1293–1302 (2003). [PubMed: 14700636]
30. Neidhardt FC, Bloch PL & Smith DF Culture Medium for Enterobacteria. J. Bacteriol 119, 736–747 (1974). [PubMed: 4604283]

31. Hoover SW et al. Bacterial production of free fatty acids from freshwater macroalgal cellulose. Appl. Microbiol. Biotechnol 91, 435–46 (2011). [PubMed: 21643704]
32. Lennen RM et al. Membrane stresses induced by overproduction of free fatty acids in Escherichia coli. Appl. Environ. Microbiol 77, 8114–28 (2011). [PubMed: 21948837]
33. Beld J, Finzel K & Burkart MD Versatility of acyl-acyl carrier protein synthetases. Chem. Biol 21, 1293–1299 (2014). [PubMed: 25308274]
34. Mendonça LMF & Marana SR Single mutations outside the active site affect the substrate specificity in a β-glycosidase. Biochim. Biophys. Acta - Proteins Proteomics 1814, 1616–1623 (2011).
35. Li Y et al. Metabolic engineering of Escherichia coli using CRISPR-Cas9 meditated genome editing. Metab. Eng 31, 13–21 (2015). [PubMed: 26141150]
36. Datsenko K. a & Wanner BL One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A 97, 6640–5 (2000). [PubMed: 10829079]
37. Gibson DG et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–5 (2009). [PubMed: 19363495]
38. Lennen RM & Pfleger BF Modulating membrane composition alters free fatty acid tolerance in Escherichia coli. PLoS One 8, e54031 (2013). [PubMed: 23349781]
39. Youngquist JT, Rose JP & Pfleger BF Free fatty acid production in Escherichia coli under phosphate-limited conditions. Appl. Microbiol. Biotechnol 97, 5149–59 (2013). [PubMed: 23619909]
40. Pantazes RJ, Grisewood MJ, Li T, Gifford NP & Maranas CD The Iterative Protein Redesign and Optimization (IPRO) suite of programs. J. Comput. Chem 36, 251–263 (2015). [PubMed: 25448866]



Mutant	Residue changes
CpFatB1.2	n/a
CpFatB1.2-M1	A59S, K296R
CpFatB1.2-M2	M29T, T117S, Q163L
CpFatB1.2-M3	A59S, K296R
CpFatB1.2-M4	ΔA_{54} (new start codon at M19), N28S, I65M
CpFatB1.2-M5	W17R, T204S
CpFatB1.2-M6	L251M, L265I,
CpFatB1.2-M7	n/a
CpFatB1.2-M8	K15E, S207T
CpFatB1.2-M9	R261S
CpFatB1.2-M10	n/a

Figure 1:

A) Pyruvate dehydrogenase complex (PDC) requires a lipolyated E2 domain (green) for *E. coli* to grow in aerobic conditions. This can be achieved via direct incorporation of lipoic acid or octanoic acid (C8:0) from the media through LplA and LipA (only for C8:0) or the LipB-mediated octanoylation of Apo-E2 domain (red) followed by insertion of sulfurs via LipA. In the absence of lipoic acid and octanoic acid from the media, a ΔlipB strain can be rescued by the action of an octanoyl-ACP thioesterase, which provides with intracellular C8:0 that can be used to lypolyate E2 domain of PDC. B) Plasmid from 10 of the 90 putative mutants were isolated and transformed into RL08ara (Δ*fadD*) strain and grown in LB 0.4% glycerol and 20 μM IPTG. Octanoic acid production for first ten mutants relative to baseline CpFatB1.2 identifies two mutants with several fold improvement over CpFatB1.2. C) Sequence identity of mutants 1–10 shows a mixture of CpFatB1.2 and CpFatB1.2 variants. Top three improved mutants (M3, M4 and M9) contain variations from CpFatB1.2. CpFatB1.2-M4 contains an early stop codon, which was characterized further in Figure 3. D) Mutants CpFatB1.2-M3, CpFatB1.2-M4, and CpFatB1.2-M9 were subcloned into high copy plasmid ptrc99a in order to remove possibilities of backbone mutations as well as increasing the expression levels to high MCFA production conditions. Plasmids were transformed into RL08ara (Δ*fadD*) strain and grown in Mops media enriched[22] with tryptone and yeast extract (see methods section) and 1 mM IPTG. Data shows several fold improvements over CpFatB1.2 in high expression conditions, consistent with the data in part B for low expression conditions.

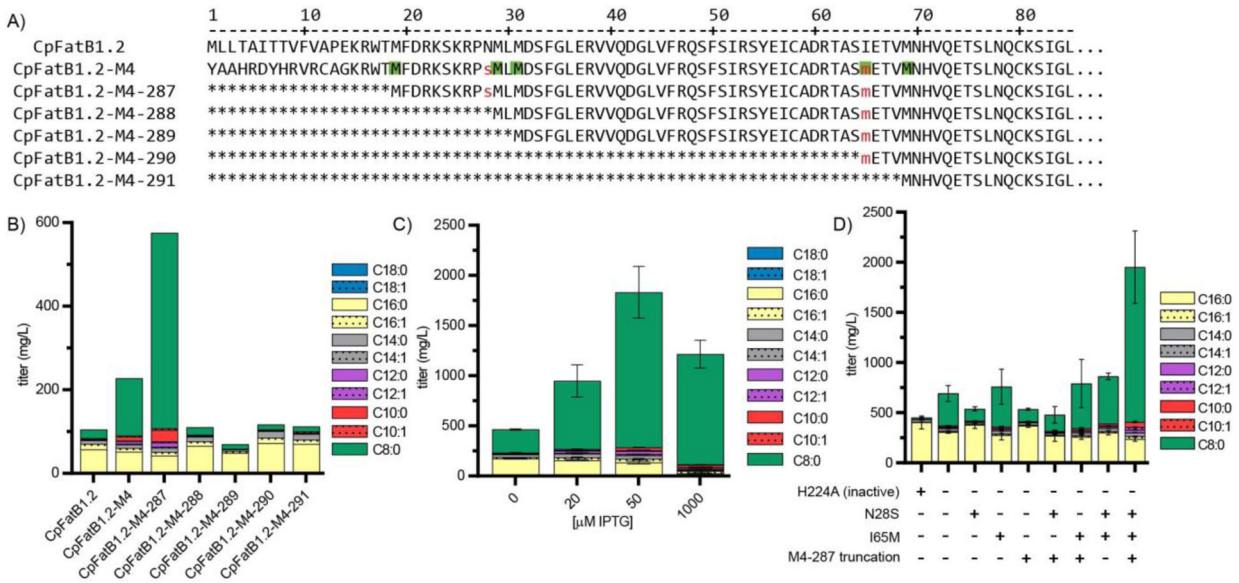
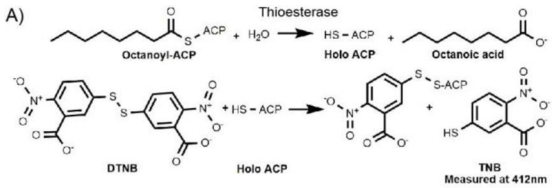
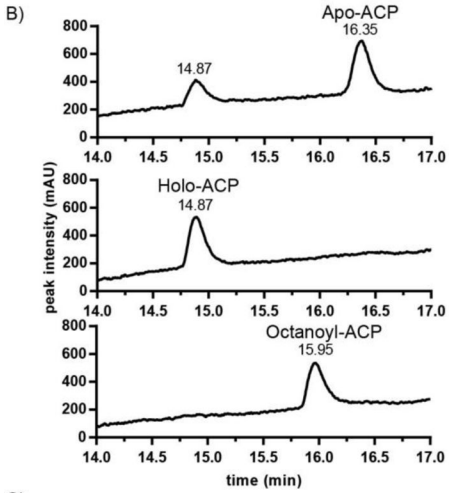
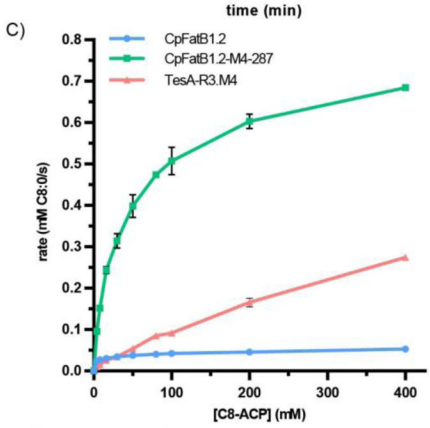


Figure 2:

Characterization of CpFatB1.2-M4. A) Alignment of CpFatB1.2 with CpFatB1.2-M4 (ΔA54, N28S, I65M) shows a frame shift which that rose from nucleotide A_{54} deletion, producing a stop codon. Given that the enzyme is active, we proposed that it is being translated from a different in-frame methionine from a non-specific ribosome binding site (RBS). We identified five methionines (highlighted in green) in frame with CpFatB1.2 and cloned in-frame CpFatB1.2-M4 variations with the methionines as start codon, CpFatB1.2-M4–287, 288, 289, 290 and 291. Mutated residues from CpFatB1.2 are in red font. B) Isolated plasmids of each variant in (A) were transformed into RL08ara (ΔfadD) strain and grown in LB 0.4% glycerol and 1 mM IPTG. Only variant CpFatB1.2-M4–287 showed high level of octanoic acid production characteristic of its parent CpFatB1.2-M4 sequence. Moreover, increased production observed in variant CpFatB1.2-M4–287 over CpFatB1.2-M4 suggests that the hypothesis of CpFatB1.2-M4 being translated from a non-specific RBS was correct. This places CpFatB1.2-M4–287 as having a ~20 fold increased production under low expression conditions tested. C) CpFatB1.2-M4–287 was then cloned into high copy plasmid ptrc99a and transformed into RL08ara (*ΔfadD*) strain and grown in Mops media enriched[22] with tryptone and yeast extract (see methods section) and different inductions of IPTG to optimize expression. This way we found that in for these conditions tested 50 μM IPTG gives maximum expression. At 1 mM IPTG we saw a growth defect that can be seen in the drastic decrease in C16 species. D) Finally, we made every single and double mutant combination contained in CpFatB1.2-M4–287 (N28S, I65M, 287-truncation) into CpFatB1.2 base thioesterase and studied FFA production in the same conditions of part (B). H224A is a catalytically inactive version of CpFatB1.2 ran as negative control. Interestingly, only I65M was observed to have minor contribution to activity as a single mutant and no significant additive effects were observed with double mutants. Only the triple mutant was able to drastically outperform the CpFatB1.2.







D)

Mutant	kcat (min^{-1})	Km (μM)	kcat/Km ($\mu M^{-1} min^{-1}$)
CpFatB1.2	68.6 ± 2.1	5.92 ± 1.1	11.6 ± 2.2
CpFatB1.2-M4-287	1076 ± 21	36.8 ± 2.3	29.2 ± 1.9
TesA-R3.M4	1028 ± 66	606 ± 56	1.70 ± 0.2

Figure 3:

A) Scheme for octanoyl-ACP assay. Holo-ACP generated by thioesterase activity reacts with DTNB forming TNB with absorbance at 412 nm. B) (top) apo-ACP and holo-ACP mixture as purified from E. coli; (center) holo-ACP after incubation of E. coli mixture with SfP; (bottom) octanoyl-ACP produced after incubation of holo-ACP with Aass (see methods). C) Activity of CpFatB1.2, CpFatB1.2-M4 and TesA-R3.M4–287[9] as a function of octanoyl-ACP concentration. D) Kinetic parameters for each enzyme in part B based on non-linear least squares fit of the curve.

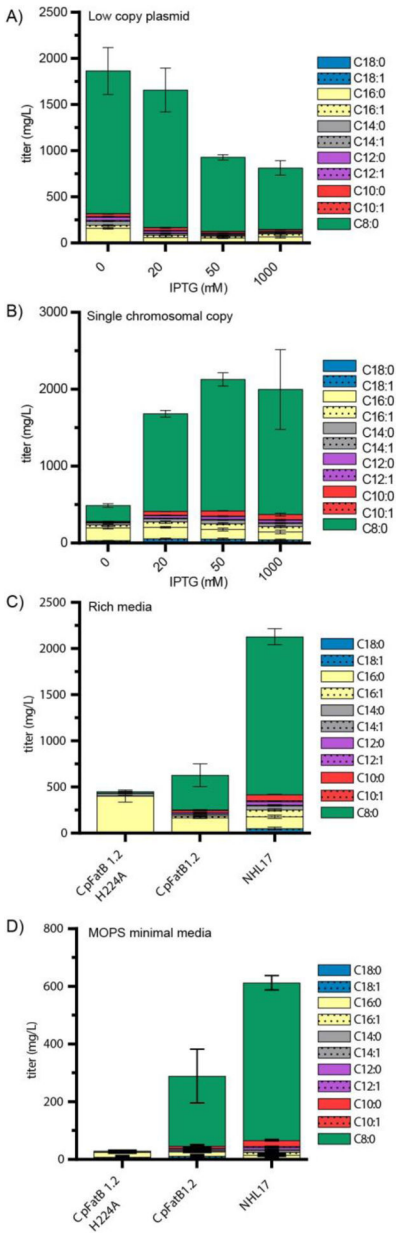


Figure 4:

Optimization of expression. A) CpFatB1.2-M4–287 was cloned into a pBTRCK plasmid with strong RBS test maximum production with plans to move the CpFatB1.2-M4–287 gene into the chromosome. After transforming into RL08ara (*ΔfadD*) strain and grown in Mops media enriched[22] with tryptone and yeast extract (see methods section) and different inductions of IPTG we found that maximum production was achieved without induction with IPTG, suggesting that sufficient expression could be achieved from the chromosome. B) Strain NHL17 (*Escherichia coli* K12 MG1655 ΔaraBAD ΔfadD::trc-CpFatB1.2-M4–287) was created (see methods) and tested in the same conditions as (A). We found that NHL17 is capable of making highest titers achieved from plasmids with a single copy of CpFatB1.2-M4–287 in the chromosome. C) Summary of highest titer achieved with CpFatB1.2 in high

copy plasmid at maximum induction versus CpFatB1.2-M4–287 from a single copy in the chromosome in Mops media enriched[22] with tryptone and yeast extract. D) Summary of highest titer achieved with CpFatB1.2 in high copy plasmid at maximum induction versus CpFatB1.2-M4–287 from a single copy in the chromosome in Mops minimal media.

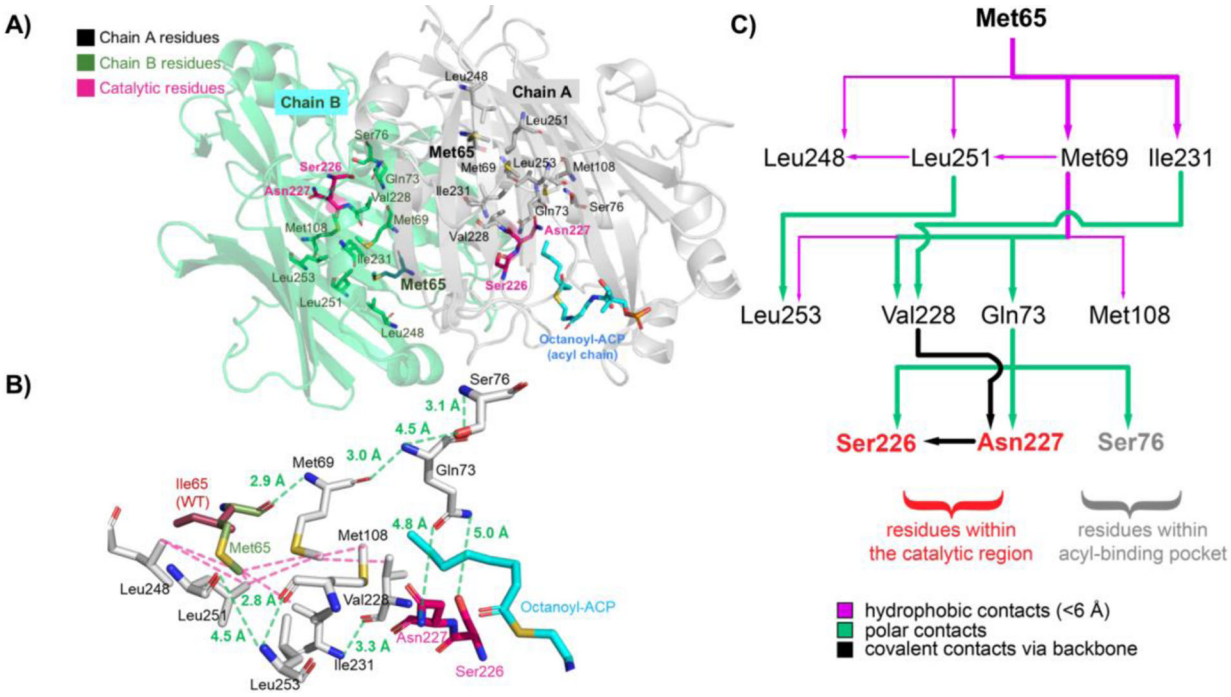


Figure 5. Effect of I65M mutation on CpFatB1.2-M4–287 mutant activity.

(A) Structural overview of CpFatB1.2-M4–287 mutant. The mutated residues are shown in bold. The catalytic residues are marked in pink. The docked configuration of the substrate (blue) has been shown in one of the two (chain A) identical binding pockets present in two chains. **(B)** Key residues that connect residue 65 to the acyl binding pocket have been shown. All the hydrophobic (pink dashes) and polar contacts (green dashes) have been overlaid on the CpFatB1.2-M4–287 mutant model. The polar distances have been labeled in green. **(C)** Contact map showing the trace of hydrophobic, polar, or covalent contacts from residue 65 to the catalytic region. The map terminates upon reaching one or more residues from the catalytic region (red) or the acyl binding-pocket (gray). Edge thickness correlates to importance of interactions.

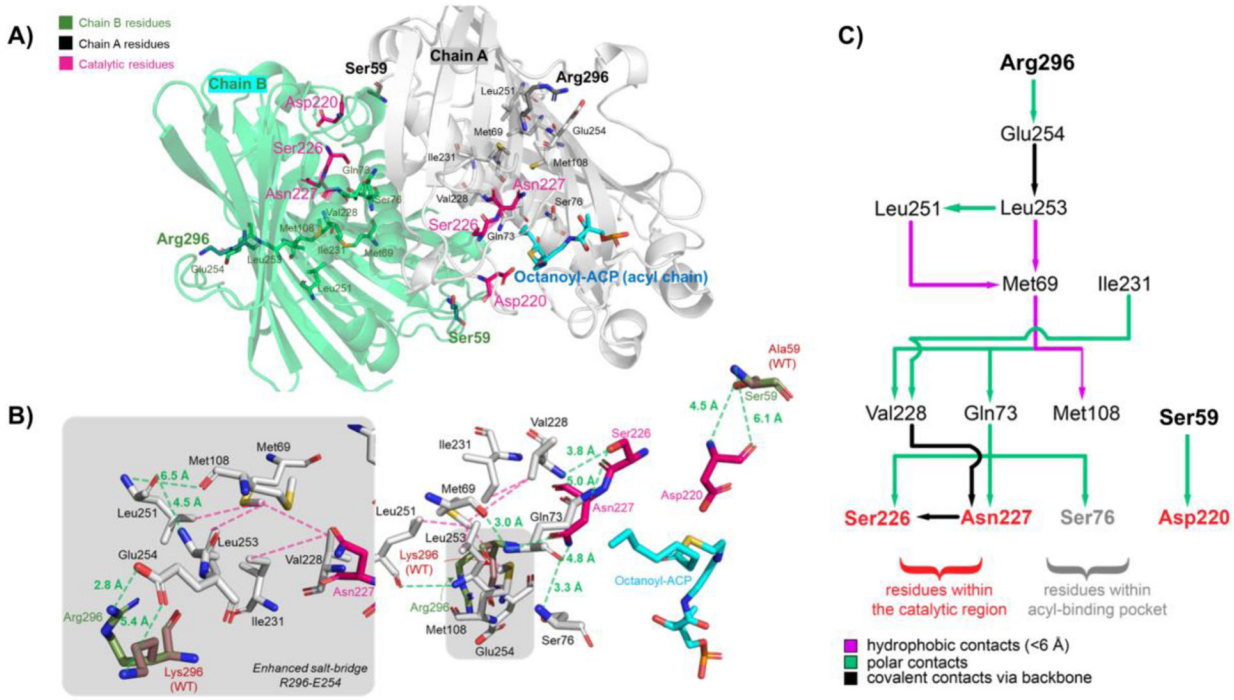


Figure 6. Effect of A59S and K296R mutations on CpFatB1.2-M3 mutant activity.

(A) Structural overview of CpFatB1.2-M3 mutant. The mutated residues are shown in bold. The catalytic residues are marked in pink. The docked configuration of the substrate (blue) has been shown in one of the two (chain A) identical binding pockets present in two chains. **(B)** Key residues that connect residues 59 and 296 to the acyl binding pocket have been shown. All the hydrophobic (pink dashes) and polar contacts (green dashes) have been overlaid on the CpFatB1.2-M3 mutant model. The polar distances have been labeled in green. The enhanced salt bridge formation between R296 and E254 has been shown inset **(C)** Contact map showing the trace of hydrophobic (pink), polar (green), or covalent contacts (black) from residues 59 and 296 to the catalytic region. Edge thickness correlates to importance of interactions.