AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIOTECHNOLOGY

Chimeric Fatty Acyl-Acyl Carrier Protein Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression

Marika Ziesack,[a,b] Nathan Rollins,[a,b] Aashna Shah,[a] Brendon Dusel,[a*] Gordon Webster,[a*] Pamela A. Silver,[a,b] Jeffrey C. Way[a]

[a]Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts, USA
[b]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

ABSTRACT Medium-chain fatty acids are commodity chemicals. Increasing and modifying the activity of thioesterases (TEs) on medium-chain fatty acyl-acyl carrier protein (acyl-ACP) esters may enable a high-yield microbial production of these molecules. The plant *Cuphea palustris* harbors two distinct TEs: *C. palustris* FatB1 (*Cp*FatB1) (C_8 specificity, lower activity) and *Cp*FatB2 (C_{14} specificity, higher activity) with 78% sequence identity. We combined structural features from these two enzymes to create several chimeric TEs, some of which showed nonnatural fatty acid production as measured by an enzymatic assay and gas chromatography-mass spectrometry (GC-MS). Notably, chimera 4 exhibited an increased C_8 fatty acid production in correlation with improved microbial expression. This chimera led us to identify *Cp*FatB2-specific amino acids between positions 219 and 272 that lead to higher protein levels. Chimera 7 produced a broad range of fatty acids and appeared to combine a fatty acid binding pocket with long-chain specificity and an ACP interaction site that may activate fatty acid extrusion. Using homology modeling and *in silico* docking with ACP, we identified a "positive patch" within amino acids 162 to 218, which may direct the ACP interaction and regulate access to short-chain fatty acids. On the basis of this modeling, we transplanted putative ACP interaction sequences from *Cp*FatB1 into *Cp*FatB2 and created a chimeric thioesterase that produced medium-chain as well as long-chain fatty acids. Thus, the engineering of chimeric enzymes and characterizing their microbial activity and chain-length specificity suggested mechanistic insights into TE functions and also generated thioesterases with potentially useful properties. These observations may inform a rational engineering of TEs to allow alkyl chain length control.

IMPORTANCE Medium-chain fatty acids are important commodity chemicals. These molecules are used as plastic precursors and in shampoos and other detergents and could be used as biofuel precursors if production economics were favorable. Hydrocarbon-based liquid fuels must be optimized to have a desired boiling point, low freezing point, low viscosity, and other physical characteristics. Similarly, the solubility and harshness of detergents and the flexibility of plastic polymers can be modulated. The length and distribution of the carbon chains in the hydrophobic tails determine these properties. The biological synthesis of cell membranes and fatty acids produces chains of primarily 16 to 18 carbons, which give rise to current biofuels. The ultimate goal of the work presented here is to engineer metabolic pathways to produce designer molecules with the correct number of carbons in a chain, so that such molecules could be used directly as specialty commodity chemicals or as fuels after minimal processing.

KEYWORDS protein engineering, chimeric enzymes, octanoic acid

Fatty acids are commodity chemicals with applications as detergents, plastic precursors, and biofuel precursors (1–6). Normally, bacterial fatty acids reach lengths of 16 or 18 carbons and are incorporated as membrane components. Fatty acids of these chain lengths are typical in hard-water-insoluble soap; improved lathering can be

Received 24 December 2017 **Accepted** 12 March 2018

Accepted manuscript posted online 16 March 2018

Citation Ziesack M, Rollins N, Shah A, Dusel B, Webster G, Silver PA, Way JC. 2018. Chimeric fatty acyl-acyl carrier protein thioesterases provide mechanistic insight into enzyme specificity and expression. Appl Environ Microbiol 84:e02868-17. https://doi.org/10.1128/AEM.02868-17.

Editor Volker Müller, Goethe University Frankfurt am Main

Address correspondence to Marika Ziesack, marika.ziesack@googlemail.com, or Jeffrey C. Way, jeff.way@wyss.harvard.edu.

* Present address: Brendon Dusel, Conagen Inc., Bedford, Massachusetts, USA; Gordon Webster, Amber Biology LLC, Cambridge, Massachusetts, USA.

M.Z. and N.R. contributed equally to this work.

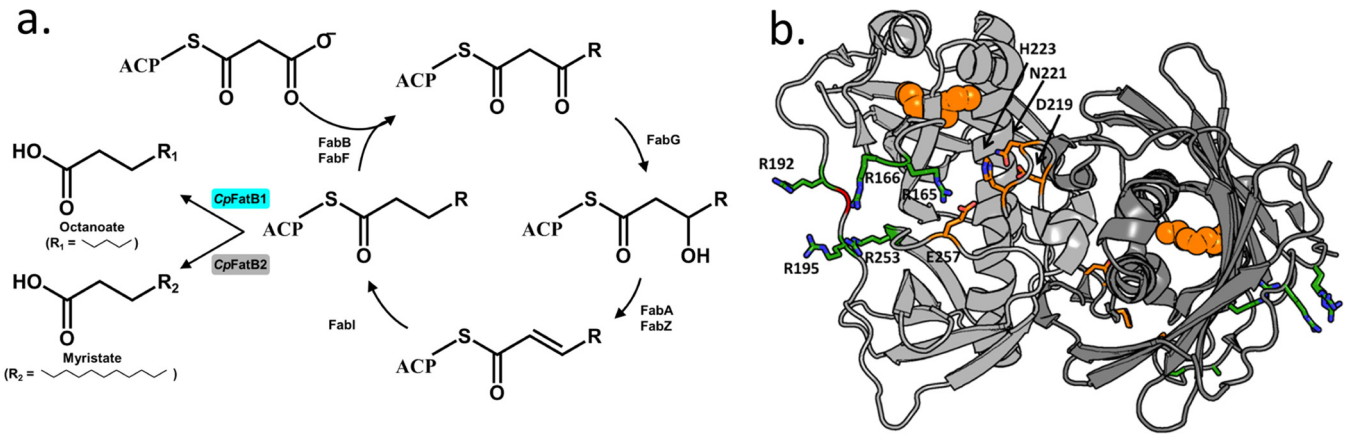


FIG 1 FatB-type thioesterase (TE) biochemistry. (a) Synthetic biochemical pathway integrating plant-derived TEs that have different chain length specificities with bacterial metabolism. During *E. coli* fatty acid biosynthesis, the hydrocarbon chains are tethered to an acyl carrier protein (ACP). Fab enzymes elongate the fatty acid chain, through sequential malonyl-ACP condensation, reduction, dehydration, and further reduction. Heterologous TEs (*Cp*FatB1 and *Cp*FatB2) interrupt this elongation cycle by hydrolyzing fatty acids of different chain lengths from the ACP. (b) Ribbon model of TE dimer (PDB accession no. 2OWN). Numbering of residues corresponds to that in Fig. 2, based on an alignment of 2OWN, *Cp*FatB1, and *Cp*FatB2 (see Fig. S1 in the supplemental material). Each monomer (light and dark gray) exhibits the characteristic hotdog fold of two α-helices partially wrapped in a β-sheet. The proposed fatty acid pocket is indicated by an electron density that was identified in the crystal structure (space filling, orange). Side chains are shown for the putative catalytic residues Asp219, Asn221, His223, and Glu257 (orange side chains) (13), as are the key positive patch residues Arg165, Arg166, Arg192, Arg195, and Arg253 (green), which may be involved in ACP binding (18). The *Cp*FatB2-specific amino acid deletion (red) is also highlighted.

achieved through the addition of shorter-chain fatty acids. Similarly, the pliability of plastic polymers can be modulated by the incorporation of fatty acids with various chain lengths. Finally, medium-chain alkanes have desirable fuel properties, such as volatility and low viscosity, and C_8 to C_{12} fatty acids would not require "cracking" to serve as cost-effective fuel precursors, for example, by the conversion to alkanes by a fatty acid reductase and an aldehyde decarbonylase (6, 7). Medium-chain fatty acids can be produced by the microbial expression of plant-derived thioesterases (TEs) alone or in combination with other heterologous engineering strategies (8–10).

TEs generate free fatty acids (FFAs) by hydrolyzing the thioester bond that connects the acyl chain to the acyl carrier protein (ACP) (Fig. 1a). TEs of the FatB class are proteins from plant seeds that are ~46 kDa in size and form homodimers (11, 12). Despite the differences in activity and specificity, they all share the highly conserved hotdog motif (11, 13). It consists of a β-sheet holding two parallel α-helices (Fig. 1b). Several hypotheses about active residues in TEs have been proposed (14–17). Structures of microbial homologs have been determined. (During the preparation of the manuscript, the first structure of a plant-derived FatB enzyme with C_{12} specificity was determined and found to be quite similar to bacterial homologues [18].) These structures suggest a potential acyl-ACP cleaving site that includes the conserved positions Asp173, Asn175, His177, and Glu211 (19) (PDB accession no. 2OWN and 2ESS, Joint Center for Structural Genomics; accession no. 4GAK, Midwest Center for Genomics) (numbering according to Fig. S1 in the supplemental material).

The determinants of chain length specificity of FatB TEs are largely undefined. Previous work has demonstrated that mutations of single amino acids in the proposed N-terminal substrate binding domain changed the specificity of *Umbellularia californica* FatB1 from C_{12} to C_{14} (18, 20). In another study, the exchange of the N terminus of *Arabidopsis thaliana* FatA and FatB indicated that this part of the protein is largely responsible for specificity despite not containing the active site residues (21). However, these studies focused on changing a short-chain-specific TE into a long-chain-specific enzyme. To our knowledge, no attempt has been made to achieve the reverse.

The interaction dynamic between ACP and fatty acid biosynthesis enzymes may be an important factor in chain length determination (19–22). ACP itself may provide specificity in differentiating between shorter and longer fatty acid tails. Bacterial and plant chloroplast ACPs have pockets that shelter part or all of the fatty acid and the

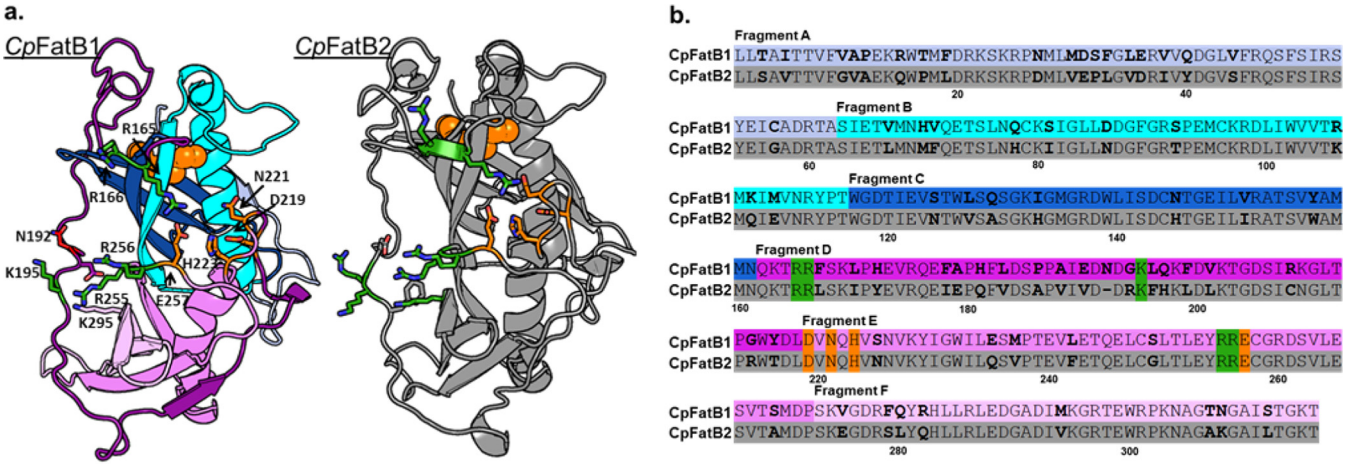


FIG 2 Homology model and sequence alignment of TEs *Cp*FatB1 and *Cp*FatB2, used as parent sequences in this study to generate chimeric enzymes. (a) Models of monomers were generated with PDB accession no. 2ESS as the template using SWISS-MODEL and Rosetta (see Materials and Methods). Fragments A to F, which were used to generate chimeras, are labeled and color-highlighted in *Cp*FatB1 (left). The electron density from structure 2OWN representing a putative alkyl chain was inserted by structure alignment and is shown as an orange space-filling model. Catalytic residue and key positive patch residue side chains are colored as in Fig. 1b. (b) Alignment of the amino acid sequences of *Cp*FatB1 and *Cp*FatB2. The fragments A to F, which were used to generate chimeric enzymes, are indicated above the sequences and color-highlighted. Nonconserved amino acids are highlighted in bold.

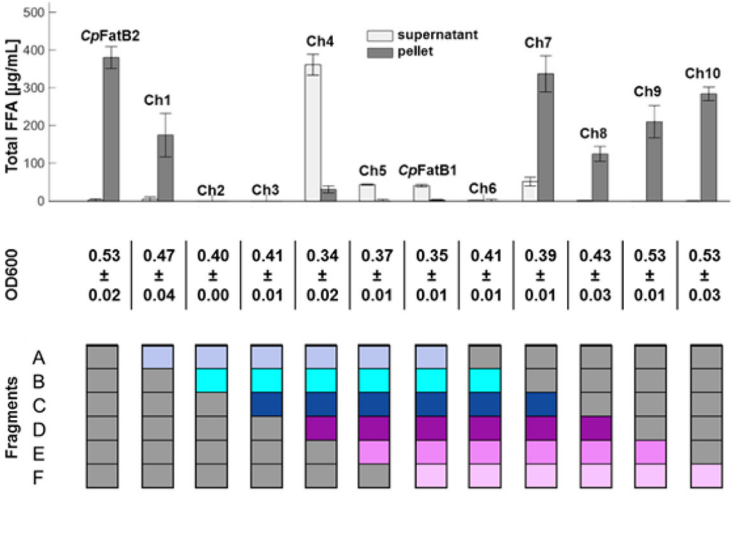
pantetheine group from the aqueous phase (23). In molecular dynamics simulations of acyl-ACP starting with an "out" conformation, longer fatty acid tails return inside the ACP less frequently, suggesting that the ratio of the in to the out states decreases with increasing fatty acid length (22). The interaction between the ACP and fatty acid-metabolizing enzymes is fleeting, but Nguyen et al. (24) solved the structure of an acyl-ACP/FabA-trapped intermediate and found an interaction with arginine side chains in a "positive patch" on FabA that might change the conformation of the ACP and induce fatty acid extrusion from within the ACP. Thus, the in versus out state of the fatty acid on the ACP is dynamic and may be regulated by the interaction with an enzymatic partner (25).

Fatty acid synthesis has many steps, and several enzymes in this process have acyl chain length specificities designed to produce chains of approximately 16 carbons, which constitute the bulk of fatty acids in membrane lipids. Several groups have engineered proteins of fatty acid synthesis to have a shortened chain length specificity (8–10, 26–28). Ultimately, it may be possible to tune the entire fatty acid synthesis pathway to produce designer molecules of practical interest. TEs will be a key enzyme in any such artificial pathway, and so here we sought to understand the features in these enzymes that control chain length specificity to enable metabolic engineers to efficiently create fatty acid products.

RESULTS

Homology modeling: homodimer structure prediction of *Cp*FatB1 and *Cp*FatB2. To understand structure-function relationships within the *Cuphea palustris* FatB TEs, we constructed homology models of *C. palustris* FatB1 (*Cp*FatB1) and FatB2 and the chimeric enzymes described below (Fig. 2a). We based the models on the structures of known hotdog fold TEs and then resampled and refined the backbone and side chain conformations using Rosetta (see Materials and Methods). The *Bacteroides thetaiotaomicron* acyl-ACP TE (Protein Data Bank [PDB] accession no. 2ESS) produced the best predictions, and the structures of the *Lactobacillus plantarum* putative oleoyl TE (accession no. 2OWN) and *Spirosoma linguale* acyl-ACP TE (accession no. 4GAK) were also closely related. The general validity of the gross predicted structures is supported by the fact that the highly conserved residues that are not part of the putative active site play key roles in the formation of the hydrophobic core and dimer interface in the models and by the similarity of our models with the C_{12} thioesterase structure reported

a.



b.

Enzyme	Total FFA [µg/mL]	Species proportion in total FFA [%]					
		C_6	C_8	C_{10}	C_{12}	C_{14}	C_{16}
None	10.8 ± 0.4	0	0	40	56	0	4
CpFatB2	372.4 ± 8.7	0	0	1	8	75	16
Ch1	158.0 ± 9.3	0	0	3	26	47	25
Ch2	12.5 ± 0.1	0	0	36	50	6	8
Ch3	11.5 ± 0.0	0	0	38	53	4	6
Ch4	394.1 ± 6.4	4	90	2	2	2	1
Ch5	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CpFatB1	84.9 ± 3.5	0	78	6	7	0	9
Ch6	25.0 ± 1.2	0	38	17	23	0	23
Ch7	321.0 ± 6.9	0	11	8	52	21	7
Ch8	116.6 ± 27.3	0	3	5	18	47	27
Ch9	275.6 ± 13.3	0	0	2	4	84	11
Ch10	293.5 ± 4.5	0	0	1	4	78	16

c.

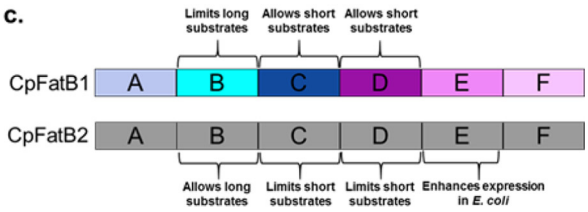


FIG 3 Typical experiment showing fatty acid production by *E. coli* expressing chimeric TEs. (a) Titers of total free fatty acids in the pellet and supernatant of *E. coli* (DE3) Δ*fadD* transformed with pBAD-HisA expressing TEs obtained by performing an enzyme-based colorimetric assay (see Materials and Methods). Densities of the corresponding cultures are given as OD_{600} values. Fragments A to F of chimeras and wild-type (WT) enzymes are shown as colored bars. Measurements were obtained from three biological replicates for each TE, with error bars indicating standard deviations. (b) The same samples were subsequently analyzed for their free fatty acid species using GC-MS (chimera 5 was omitted here but generated a fatty acid profile similar to that of *Cp*FatB1 in other experiments) (see Table S2). (c) From these data, we inferred specific functions for distinct segments of the protein.

by Feng et al. (PDB identifier [ID] 5X04 [18]), which was published while this article was in preparation.

Design and construction of chimeric TEs. The fatty acid specificity and sequences of many FatB-type TEs are known, including *Cp*FatB1 (C_8) and *Cp*FatB2 (C_{14}) of *C. palustris* (Fig. 2) (11, 29). Even though these two enzymes show distinct specificity and microbial activity levels, they share 78% sequence identity. Therefore, we hypothesized that only a small number of amino acids determine the activity and specificity. We generated chimeric proteins constructed from six 50- to 60-amino-acid-long fragments of *Cp*FatB1 and *Cp*FatB2 (Fig. 2, fragments A to F). The breakpoints for the fragments were placed at sequence regions with high identity between the two enzymes. We created the chimeric genes by systematically assembling and cloning segments into the arabinose-inducible expression vector pBAD-HisA (Thermo Fisher Scientific). The expression of hybrid TEs gave a variety of fatty acid synthesis phenotypes (Fig. 3). To initially analyze the free fatty acid (FFA) production of chimeric TEs when expressed in *Escherichia coli*, we used the free fatty acids half micro test (Roche). The final densities of the tested cultures were similar (Fig. 3a), indicating that variations in fatty acid production are most likely due to differences in enzyme expression and/or activity.

The shorter hydrocarbon chains are amphiphilic and therefore cross membranes more easily than long-chain fatty acids. C_8 fatty acids are expected to occur in the supernatant, whereas C_{14} fatty acids should be retained in the cell pellet and membranes. We detected free fatty acids in both the supernatant and the pellet and then quantified their abundance using standard curves with known concentrations of octanoate and myristate, respectively. As expected, cells expressing *Cp*FatB1 yielded fatty acids only in the supernatant and cells expressing *Cp*FatB2 yielded compounds only in the pellet (Fig. 3a). The chimeras showed a variety of fatty acid production characteristics. Enzymes that resulted in fatty acid production exhibited *Cp*FatB1-like (chimeras 4 to 6), *Cp*FatB2-like (chimeras 1 and 8 to 10), or intermediate (chimera 7) specificities. Some chimeras did not show any fatty acid production (chimeras 2 and 3). These initial results were validated by the quantitation of specific free fatty acids in

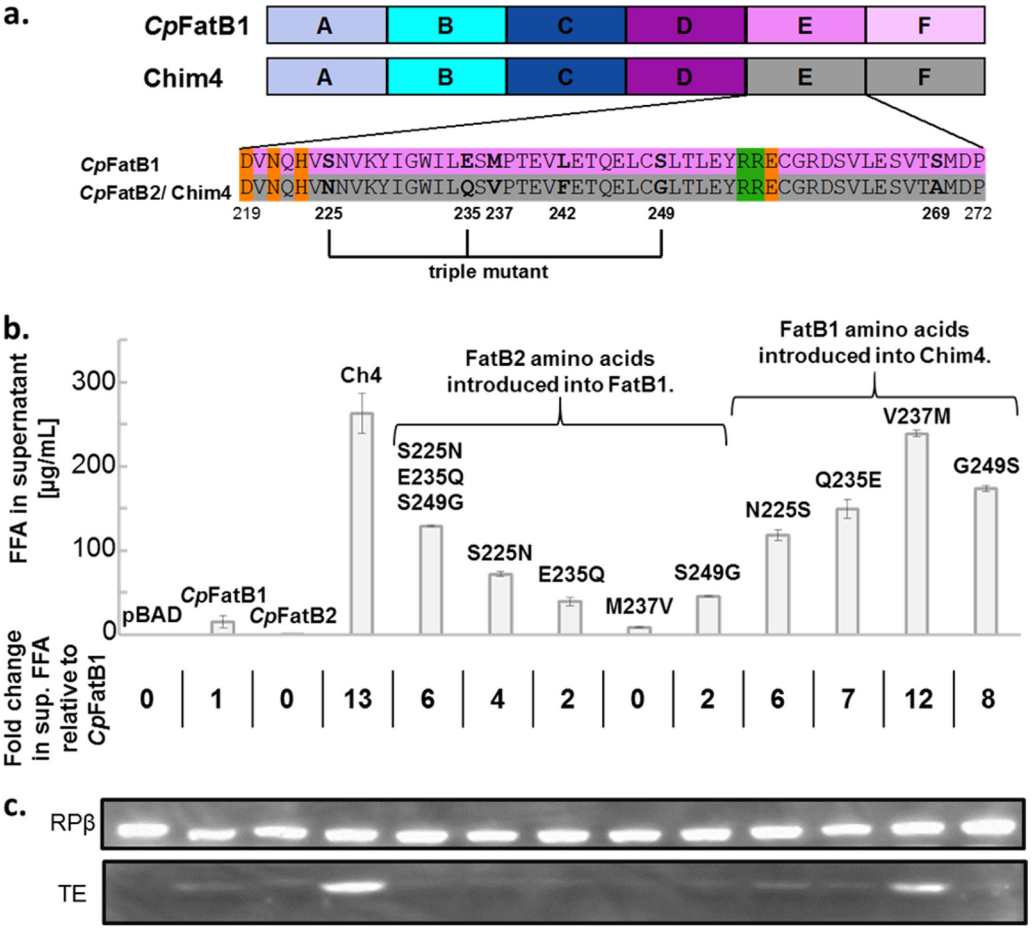


FIG 4 Dissection of fragment E's effect on microbial FFA production. (a) Chimera 4 contains fragments E and F from *Cp*FatB2. Differences in fatty acid production levels are due to fragment E (Fig. 3a). There are six amino acid substitutions between *Cp*FatB1 and *Cp*FatB2 in fragment E. (b) Amino acid substitutions were introduced into either *Cp*FatB1 (converting to the *Cp*FatB2 sequence) or into chimera 4 (converting to the *Cp*FatB1 sequence). Constructs were expressed, and FFAs in supernatants were measured using an enzymatic assay. Shown are the averages from three biological replicates; error bars indicate standard deviations. (Substitutions of L/F242 and S/A269 had no effect [see Fig. S3].) (c) From the same cultures as in panel b, TEs were visualized by Western blotting using an anti-His6 antibody (for full blot, see Fig. S4). As a loading reference, the blot was also probed with an anti-RNA polymerase beta antibody (RPβ).

whole cultures as determined by gas chromatography-mass spectrometry (GC-MS) (Fig. 3b).

Increased microbial activity of chimera 4. Chimera 4 maintained the specificity from *Cp*FatB1 but had a much higher microbial activity (Fig. 3), which may make this construct useful in the engineered synthesis of octanoic acid. Conversely, chimera 5 showed the specificity and activity of *Cp*FatB1. The only difference between the two chimeras is fragment E (amino acids 220 to 272), which also contains the proposed active site loop and adjacent helices and loops (Fig. 2). We inferred that this fragment contains the polymorphisms responsible for the microbial activity difference between *Cp*FatB1 and chimera 4 (Chim4). To quantify the composition of the fatty acids produced, we used GC-MS analysis. The octanoate production of chimera 4 exceeded *Cp*FatB1's production by approximately 6- to 13-fold (Fig. 3).

The protein expression levels were dramatically enhanced by *Cp*FatB2 genetic polymorphisms between amino acids 219 to 272, which likely explains the enhanced fatty acid production by chimera 4 (Fig. 4). This region has only six FatB1/FatB2 polymorphisms: S225N, E235Q, M237V, L242F, S249G, and S269A. To understand how these polymorphisms contribute to enhanced fatty acid production, we introduced the six *Cp*FatB2 (favorable) residues into *Cp*FatB1 and the six *Cp*FatB1 (unfavorable) resi-

dues into chimera 4 and measured free fatty acid production and protein expression levels via Western blotting (Fig. 4). Three of the polymorphisms had detectable effects within *Cp*FatB1 on fatty acid production, namely, S225N, E235Q, and S249G (Fig. 4b), while the other three did not (Fig. 4b; see also Fig. S3 in the supplemental material). For example, the introduction of the change S225 to N increased fatty acid production approximately 4-fold, while the reciprocal change of N225 to S in chimera 4 reduced fatty acid production approximately 2-fold. Exchanges of E235Q and S249G had analogous but less dramatic effects. Combining these three mutations within a *Cp*FatB1 protein background had an additive effect, with microbial activity approaching that of chimera 4. A Western blot analysis indicated that the most active proteins, chimera 4 and chimera 4 withV237M, were clearly produced at higher levels than the less active proteins (Fig. 4c). These results suggest that the folding of these plant-derived TEs may not be optimal in *E. coli*, with unfolded proteins being degraded, and that chimera 4 fortuitously has superior folding properties.

Seeking structural explanations for the phenotypes of each chimera, we constructed and examined three-dimensional homology models of *Cp*FatB1, *Cp*FatB2, and chimera 4. We generated the initial structures from SWISS-MODEL and refined them using Rosetta. An inspection of the structure models of *Cp*FatB1, *Cp*FatB2, and chimera 4 suggests that the most helpful mutation, S225N, may enhance the formation or stability of an α-helix that extends from Asn226 to Met237. In addition, Ser/Asn225 is at the $+2$ position relative to the putative active site histidine, and this substitution may cause a slight conformational change in the active site and could influence enzyme activity. The reasons why the polymorphisms at positions E235Q and S249G affect fatty acid production are not apparent from an inspection of the structure; these variations may affect the formation of incorrectly folded states that are degraded.

Permissive and restrictive determinants of chain length specificity. Chimera 7 showed a novel fatty acid production phenotype that was distinct from those of both *Cp*FatB1 and *Cp*FatB2. This chimeric protein produced a range of fatty acids from C_8 to C_{16}, with C_{12} being the most abundant species. Chimera 7 differs by 10 amino acids in fragment B (amino acids 63 to 116) from chimera 6, which produced primarily C_8 fatty acids. These results suggest that the *Cp*FatB1 polymorphisms in this region together prevent longer chain fatty acids from entering the proposed binding pocket, while the *Cp*FatB2 polymorphisms are more permissive to longer chains. (We also constructed numerous variant proteins with single amino acid substitutions in this region but were unable to find single point mutations that conferred a change in specificity [see Table S1].) The polymorphic amino acids are in the proposed fatty acid pocket, suggesting that these polymorphisms may collectively influence the binding pocket size.

Molecular modeling and an examination of homologous structures support the notion that this region of the protein defines a fatty acid binding pocket. The related structures of PDB accession no. 2ESS and 2OWN show electron densities in the putative binding pockets, consistent with this identification. Specifically, the structure 2ESS has two molecules of the crystallization agent 2-methyl-2,4-pentanediol in the pocket, and 2OWN has an unidentified electron density, modeled as a 9-carbon alkyl chain, in the same position (Fig. 1b). All together, these results suggest that a pocket defined by amino acids Gln72, Cys79, Asn86, Val104, Arg137, Trp102, and Met135, as well as Tyr/Trp157, Val 219, and Asn224 (in *Cp*FatB2), binds to the alkyl chain of the fatty acid. This interpretation is supported by mutational and structural analyses of other related structures, including the recently solved C_{12} TE (16, 18).

Fragment C (amino acids 117 to 161) appears to define a segment of *Cp*FatB2 that functions to restrict the substrate specificity of this protein to longer chain fatty acyl-ACPs. Chimera 7 has a relaxed specificity and generated a range of fatty acid products, while the distribution of products from chimera 8 is much more skewed toward longer products (Fig. 3b), with C_{14} the main product and very little C_8. Chimera 9, in which fragment D (amino acids 162 to 218) was also derived from FatB2, produced a distribution of fatty acids even more skewed toward C_{14} and with no C_8, essentially

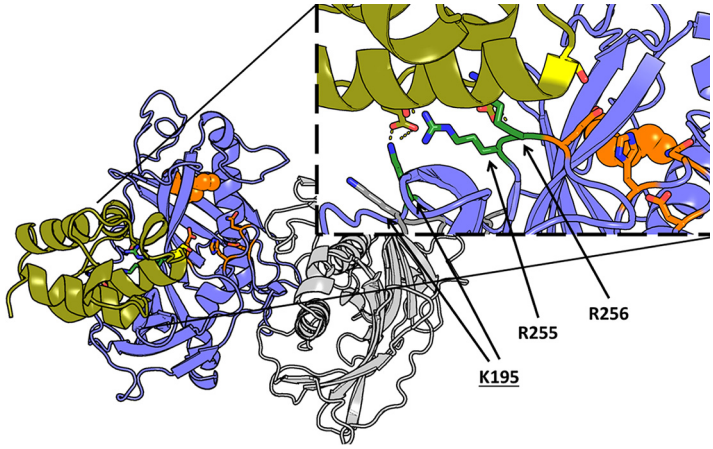


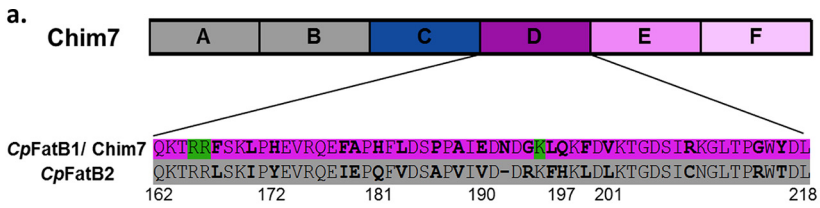
FIG 5 ACP docking to *Cp*FatB1 and *Cp*FatB2. *In silico* docking of ACP (green) with modeled *Cp*FatB1 (blue) was done using Rosetta's docking protocol (RosettaDock). We applied several solved structures of ACP (PDB IDs 1T8K, 2FAD, 2FAE, and 3EFB) and sampling was unbiased. Ten thousand arrangements were generated and scored, and the best 200 of each pair were clustered to identify 10 representative binding orientations. The proposed fatty acid pocket and active site residues (orange) are in close proximity to the ACP binding site. The upper right expanded view (rotated relative to the depicted dimer) shows direct interactions between "positive patch" residues K195, R255, and R256 and ACP residues. The predicted position of K195 in *Cp*FatB2 is superimposed in gray.

identical to that of *Cp*FatB2 itself. Figure 3c summarizes these observations. We note that according to the scheme in Fig. 3c, chimera 2, which is the converse of chimera 7, should contain segments that restrict both long- and short-chain fatty acyl substrates and was predicted to be inactive, which it is. However, there are other reasons why this chimera might be inactive, such as improper folding. We hypothesize that features of fragments C and D may promote ACP binding in a way that, in *Cp*FatB1, promotes the extrusion of the fatty acid group, particularly for shorter chains.

Docking of ACP with *Cp*FatB1 and *Cp*FatB2 models. We generated 10,000 docking conformations between *E. coli* ACPs and the models of *Cp*FatB1, *Cp*FatB1, and chimera 4 (see Materials and Methods). Of the most energetically favorable conformations, we clustered similar structures to identify the dominant binding modes (Fig. 5). An inspection revealed that the top sampled binding modes primarily occurred between the ACP and the "positive patch," including Arg255 and Arg256 as well as Arg165, Arg166, and Lys195. The binding modes that correctly oriented the pantetheine-carrying Ser (modeled without the pantetheine modification) were found exclusively in the sampling for apo-ACP with *Cp*FatB1 and chimera 4. Because sampling is not exhaustive, we also aligned this binding mode against the *Cp*FatB2 model to determine if it was substantiated by interactions specific to *Cp*FatB1 or *Cp*FatB2. The residues that directly interfaced with the ACP are conserved (Lys195, Arg255, and Arg256) (Fig. 5). The two arginines directly preceding the active site (Arg255 and Arg256) intercalate between two helices of the ACP, similar to in structures of FabA-docked ACP (PDB ID 4KEH) (24). A third supporting polar interaction in the apo form of the ACP-*Cp*FatB1/2 docking was provided by Lys195, a residue on the loop of fragment D. In *Cp*FatB2, this loop is significantly altered relative to that in *Cp*FatB1 (Fig. 2b), including the sole insertion/deletion, and the models place the FatB2 Lys195 side chain further from the putative ACP docking site than for *Cp*FatB1 (Fig. 5, inset).

The transplantation of amino acids 162 to 218 (fragment D) from FatB1 into FatB2 causes a relaxation of specificity, allowing medium-chain fatty acids to be produced (Fig. 6). These amino acids are not part of the alkyl chain binding pocket but do interact with the ACP in the docking models. Specifically, the production of C_8 and C_{10} fatty acids from *Cp*FatB2 was undetectable, while the production of these fatty acids from the FatB2 that contained fragment D from FatB1 [FatB2(FatB1-fragment D)] was approximately 10% of the total, and the C_{12} fatty acid was only 6% of the product from

a.



b.

Enzyme	Total FFA [μg/mL]	Species proportion in total FFA [%]				
		C_8	C_{10}	C_{12}	C_{14}	C_{16}
CpFatB1	32 ± 11.4	75	4	0	2	18
CpFatB2	190.1 ± 11.6	0	1	6	86	8
Ch7	185.0 ± 11.6	20	9	48	19	5
FatB2(FatB1-Fragment D)	238.9 ± 12.1	6	4	20	57	14

FIG 6 Region distinct from the fatty acid binding pocket, fragment D, modulates the FFA length specificity of *Cp*FatB1/B2. (a) Chimera 7 organization. Chimera 7 includes sequences from *Cp*FatB2 that allow binding of a long-chain fatty acid (fragment B) and sequences from *Cp*FatB1 corresponding to the ACP binding site inferred from modeling (fragments C and D). (b) Transplantation of fragment D from *Cp*FatB1 into *Cp*FatB2 widens the distribution of fatty acids produced. The entire fragment D from *Cp*FatB1 was introduced into *Cp*FatB2, proteins were expressed in *E. coli*, and fatty acid production profiles were analyzed via GC-MS. Shown are the averages from three biological replicates and their standard deviations.

*Cp*FatB2 but was 20% of the product from FatB2(FatB1-fragment D) (Fig. 6b). These results indicated that it is possible to modulate the chain length substrate specificity without changing the amino acids around the likely alkyl chain binding pocket. We hypothesize that amino acids 162 to 218 and also 117 to 161 contribute to an ACP contact site, and the polymorphisms found in *Cp*FatB1 enhance the extrusion of medium-chain fatty acids from the ACP. (We tested a few subsets of the 162-to-218 fragment but were unable to introduce short-chain specificity [see Fig. S5].)

DISCUSSION

Plant-derived fatty acyl TEs are important tools in the microbial production of fatty acids by metabolic engineering. However, the factors that determine the microbial activity levels and the alkyl chain length specificity of these enzymes are poorly understood. Here, we used a combination of hybrid enzyme construction, site-directed mutagenesis, and *in silico* modeling to find determinants of the expression and specificity of the C_8 and C_{14} fatty acyl TEs of *C. palustris*. These two enzymes are 78% identical at the amino acid level yet have strikingly different microbial activities and substrate specificities.

One hybrid enzyme (chimera 4) had the C_8 specificity of *Cp*FatB1 with the higher *in vivo* activity of *Cp*FatB2 (Fig. 3). This enzyme consists of amino acids 1 to 218 of *Cp*FatB1 and amino acids 219 to 316 of *Cp*FatB2. The enhanced microbial activity is likely due to the improved expression of the hybrid protein (Fig. 4), which appears to be expressed in *E. coli* at higher levels than either *Cp*FatB1 or *Cp*FatB2 under identical conditions. The polymorphisms Ser/Asn225, Glu/Gln235, and Ser/Gly249 in this segment appear to additively contribute to an improved *in vivo* activity. This hybrid enzyme may be particularly useful in the microbial production of C_8 fatty acids.

The binding site for the alkyl chain of the fatty acid is likely to be a deep pocket flanked by amino acids, including Cys79, Trp102, Met135, Arg137, and Tyr/Trp157. This conclusion is based on the observation that active chimeric enzymes with amino acids in this segment from *Cp*FatB1 produced only a C_8 fatty acid, while those with the amino acids from *Cp*FatB2 produced a C_{14} fatty acid (Fig. 3). Further evidence that this region encodes the fatty acid binding pocket comes from the structure of a bacterial TE, 2ESS, which has two solvent molecules in this pocket, and the putative TE structure 2OWN, which has an electron density consistent with an alkyl chain in this area. The site-directed mutagenesis studies that converted a plant C_{12} TE to accommodate C_{14}

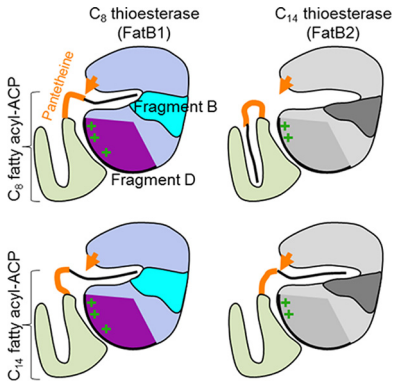


FIG 7 Model for structural mechanisms that determine chain length specificity in FatB-type TEs. According to this model, C_8 fatty acids can be cleaved by *Cp*FatB1 (blue shapes), because the positive patch (green "+" signs) on this enzyme stabilizes the required ACP state (light green shapes). *Cp*FatB2 (gray shapes) is unable to cleave C_8 fatty acids because it lacks the right conformation of the positive patch, which is mostly guided by fragment D. *Cp*FatB1 cannot cleave C_{14} fatty acids because its pocket is too short, causing the pantetheine group (orange) to stick out instead of residing near the catalytic residues (orange arrow). C_{14} fatty acids are released from the ACP more readily without the requirement of stabilization through the positive patch. *Cp*FatB2 contains a pocket that accepts larger fatty acids, which is governed by fragment B (dark gray). Hence, it is able to cleave C_{14} fatty acids.

substrates support this interpretation (18). *Cp*FatB1/2 polymorphisms in this region appear to restrict or allow access by long alkyl chains in the fatty acid (Fig. 3c).

The length specificity of *Cp*FatB1, the C_8 TE, may be explained in terms of the steric exclusion of longer alkyl chains, but the mechanism by which the *Cp*FatB2 C_{14} TE excludes shorter chains cannot be so simple. We hypothesize that *Cp*FatB2 achieves a specificity for longer chains by using the equilibrium of an ACP with longer fatty acids, which favors the protruded state. ACPs in bacteria and plants have deep pockets in which attached fatty acids preferentially reside (23). This can influence the length specificity of enzymes involved in fatty acid synthesis (8). Chan et al. (22) simulated the in-to-out and out-to-in transition rates of ACP bound to fatty acids of various lengths; they found that the longer fatty acids remain outside the ACP for a longer period of time and had fewer stabilizing hydrogen bonds between pantetheine and the protein when the fatty acid was in the "in" state. We propose that the C_8 TE may have a mechanism for encouraging the protrusion of the fatty acid from the ACP, while the C_{14} TE may rely on the natural equilibrium for the C_{14} acyl-ACP to be in the "out" state.

ACPs are negatively charged proteins, and fatty acyl-ACPs generally interact via a "positive patch" on enzymes that process fatty acids. Nguyen et al. (24) determined the structure of the *E. coli* fatty acid dehydratase (FabA) with an ACP and an attached fatty acid in the extruded state. In this structure, the ACP interacts directly with two arginines within a larger positive patch, and this interaction appears to modulate the structure of the ACP to favor fatty acid release. The C_8 and C_{14} TEs also have positive patches near the active sites: amino acids Lys97, Arg98, Arg/Lys106, Lys132, Lys163, Lys165, Arg166, Arg169, Arg175, Arg/Gly194, Lys195, Arg255, and Arg256. Most of these amino acids lie in fragments C and D, which appear to allow or block access to shorter chain substrates (Fig. 3 and 6). The entire "positive patch" is rather large, so the ACP may initially bind nonspecifically through charge-charge interactions before docking at the site that stimulates fatty acid extrusion.

Polymorphisms between *Cp*FatB1 and *Cp*FatB2 may change the ACP docking site to differentially promote fatty acid protrusion. Specifically, we hypothesize that the *Cp*FatB2-ACP interaction is less favorable than the *Cp*FatB1-ACP interaction, but this effect would be balanced by the greater intrinsic tendency of the C_{14} acyl-ACP to remain in the "out" configuration. Figure 7 summarizes this model. To address this hypothesis, we transplanted amino acids 162 to 218 (fragment D) from *Cp*FatB1 into *Cp*FatB2; this segment does not contain any amino acids around the putative alkyl chain binding pocket but does include amino acids predicted to interact with the ACP.

The resulting hybrid protein retains specificity for C_{14} substrates but has the added ability to cleave C_8-, C_{10}-, and C_{12}-length substrates (Fig. 6b). This observation suggests that *Cp*FatB TEs contain features that are permissive/restrictive for medium-chain substrates and that these features are distinct from the fatty acid binding pocket and its steric restriction of long-chain substrates.

In summary, we analyzed the structure-function relationships that control microbial activity and the length specificity of *Cuphea palustris* fatty acyl-ACP TEs. We identified sequence features of these proteins that can be modified to increase protein expression in bacteria, as well as features that may control length specificity. These observations may be useful in the production of designer chemicals by metabolic engineering.

MATERIALS AND METHODS

Strain construction. Double-stranded DNA-encoding fragments A to F of *Cp*FatB1 and *Cp*FatB2 (Fig. 2) were synthesized as gBlocks (Integrated DNA Technologies) and cloned into the pBAD-HisA (Thermo Fisher Scientific) vector by isothermal assembly (30) (all chimera sequences are listed in Fig. S2 in the supplemental material). Amino acid mutations of the *Cp*FatB1 and *Cp*FatB2 genes were incorporated by PCR amplification with mutagenesis oligonucleotides. An empty pBAD-HisA vector without a FatB gene was included as a control. The plasmids were sequence verified and transformed into *E. coli* BL21(DE3) Δ*fadD*.

Growth and induction. Fatty acid production experiments were performed in glass tubes. Individual colonies of fresh transformants were grown overnight in 1.5 ml LB with 1% glucose and 100 μg/ml carbenicillin in 15-ml culture tubes at 37°C and then diluted 1:20 into 2 ml or 5 ml M9 with 1% glycerol, grown at 30°C in a roller drum to an optical density at 600 nm (OD_{600}) of 0.4 to 0.6, and induced with 1% arabinose. The cells were harvested after 24 h for enzymatic fatty acid analysis, GC-MS analysis, and Western blotting.

OD and enzymatic FFA measurements. The OD_{595} of 100-μl cultures was measured in a 96-well flat-bottom, nontreated, sterile polystyrene plate (Corning) on a Victor 3V 1420 multilabel counter (Perkin-Elmer) using a 595/60-nm filter. Enzymatic FFA measurements were performed according to Torella et al. (10) with modifications as follows. Two hundred microliters of each culture was separated into a supernatant and a pellet by centrifugation. FFAs in 5-μl supernatants were measured directly using the enzymatic free fatty acid half micro test (Roche) according to the manufacturer's instructions, scaled down to 1/10 volume in a 96-well plate with octanoate as the standard for quantification. FFAs in the pellet were extracted into ethyl acetate as follows. To a 100-μl redissolved pellet in M9 glycerol with 1% NP-40, we added 10 μl of 10% NaCl, 10 μl glacial acetic acid, and 40 μl ethyl acetate. The solution was mixed by vortexing and centrifuged for 10 min at 16,000 rpm. Seventeen microliters of the ethyl acetate phase was then evaporated using a SpeedVac and redissolved in M9 glycerol, 5 μl of which was subjected to the enzymatic free fatty acid half micro test as described above, using a myristate standard for quantification.

Fatty acid identification and quantification using GC-MS. The identification and quantification of FFAs were essentially performed according to Torella et al. (10). In brief, we separated FFAs from the culture by acidifying 400 μl culture with 50 μl 10% (wt/vol) NaCl and 50 μl glacial acetic acid and extracting the FFAs into 200 μl ethyl acetate. FFAs in 100 μl organic phase were then esterified in 900 μl of a 30:1 mixture of ethyl alcohol (EtOH) and 37% (vol/vol) HCl by incubating them at 55°C and extracted into 500 μl hexane. Fatty ethyl esters were analyzed on an Agilent GC-MS 5975/7890 (Agilent Technologies) using an HP-5MS column (length, 30 m; diameter, either 0.25 or 0.5 mm; film, 0.25 μm). We identified chain lengths by GC retention times and mass spectra and quantified them using an internal standard (800 mg/liter pentadecanoate added to the culture before the extraction procedure). We used known concentrations of C_6, C_8, C_{10}, C_{12}, C_{14}, and C_{16} fatty acids to generate a standard curve and to quantify the production of single fatty acid species.

Computational fold prediction. The lack of structures for these TEs or chimeric proteins required that the structural analysis rely on representative homology models. To generate these models, we first assembled the initial structures using SWISS-MODEL to thread the *Cp*FatB1 and *Cp*FatB2 sequences to aligned coordinates in all structures of dimeric TEs within 25% sequence homology (31–34). We then used Rosetta to apply several consecutive rounds of random perturbations to each starting point provided by SWISS-MODEL, extending the sampling to thousands of neighboring conformations (35–37). The lowest energy perturbations of each homology model were optimized by iterative Monte Carlo adjustments to the protein backbone (38). The best scoring models were relaxed with spheroid side chain approximations, finely backbone minimized, and then side chain minimized with full-atom refinement. The final models were then clustered and selected by the Rosetta score.

***In silico* docking of ACP with modeled TEs.** We used Rosetta's docking protocol (RosettaDock) to identify the potential binding orientations of the ACP to *Cp*FatB1, *Cp*FatB2, and chimera 4 (39). To determine whether the binding modes were particular to the loading state of the ACP, we docked the solved structures of the *E. coli* ACPs, both empty and carrying 7-, 10-, and 14-carbon fatty acids (PDB IDs 1T8K, 2FAD, and 3EJB, respectively) (23, 40, 41). The sampling was unbiased, obtained by randomizing the rotational orientation of the ACP and testing random positions of the protein placed within a 45-Å radius of the TE center of mass. Ten thousand arrangements of ACPs and TE dimers were generated this way for each pairing, and side chains at the ACP-TE contacting surfaces were optimized. The overall structures were scored, and we took the best scoring of each pair. Clustering those structures to identify

the most successful binding modes, we inspected the largest clusters as representative binding orientations.

Western blot for His-tagged thioesterases. We directly measured thioesterase levels in expressing *E. coli* strains by Western blotting. Cells were grown and induced as described in "Growth and induction." Five-milliliter cultures were spun down and resuspended in 50 μl Tris-HCl 50 mM (pH 7.4), the protein was quantified using a Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Scientific), and 25 μg was loaded in a Novex WedgeWell 4 to 20% Tris-glycine gel (Invitrogen by Thermo Fisher Scientific) according to the manufacturer's instructions. The gel was run for 2.5 h at 120 V. The proteins were transferred to a nitrocellulose membrane (iBlot gel transfer stacks nitrocellulose, regular; Invitrogen by Thermo Fisher Scientific). The membrane was blocked with 4% milk in Tris-buffered saline–Tween 20 buffer (TBS-T) at 4°C overnight. The anti-RNA polymerase beta antibody (EPR18704; Abcam) was added at a 1:2,000 dilution and incubated for 30 min, followed by four washes for 5 min with TBS-T. The horseradish peroxidase (HRP)-conjugated antibodies anti-rabbit IgG (H+L) (XG-6160HRP; ProSci Incorporated) and anti-6× His tag (ab1187; Abcam) were added at 1:5,000 dilutions, incubated, and washed as with the primary antibody. The blot was imaged using SuperSignal West Dura extended duration substrate (Thermo Scientific). Western blotting was performed on whole-cell lysates rather than supernatants, because our thioesterase constructs appear to associate with membranes (see Fig. S6).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02868-17.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

ACKNOWLEDGMENTS

This work was supported by the Wyss Institute for Biologically Inspired Engineering at Harvard University, DARPA Living Foundries contract HR0011-12-C-0061, and DOE/ARPA-E collaborative agreement DE-AR0000079. We declare no competing financial interests.

M.Z. performed the experiments and N.R. performed structural modeling. B.D., G.W., and A.S. designed and built the constructs. P.A.S. and J.C.W. conceived the project. M.Z., N.R., and J.C.W. analyzed the data and wrote the manuscript.

We thank Joe Torella and Tyler Ford for experimental advice and Andy Shumaker for useful discussions.

REFERENCES

1. Lennen RM, Pfleger BF. 2013. Microbial production of fatty acid-derived fuels and chemicals. Curr Opin Biotechnol 24:1044–1053. https://doi.org/10.1016/j.copbio.2013.02.028.
2. Yu A-Q, Pratomo Juwono NK, Leong SSJ, Chang MW. 2014. Production of fatty acid-derived valuable chemicals in synthetic microbes. Front Bioeng Biotechnol 2:78. https://doi.org/10.3389/fbioe.2014.00078.
3. Zhou YJ, Buijs NA, Siewers V, Nielsen J. 2014. Fatty acid-derived biofuels and chemicals production in *Saccharomyces cerevisiae*. Front Bioeng Biotechnol 2:32. https://doi.org/10.3389/fbioe.2014.00032.
4. Sheppard MJ, Kunjapur AM, Prather KLJ. 2016. Modular and selective biosynthesis of gasoline-range alkanes. Metab Eng 33:28–40. https://doi.org/10.1016/j.ymben.2015.10.010.
5. Zhang F, Rodriguez S, Keasling JD. 2011. Metabolic engineering of microbial pathways for advanced biofuels production. Curr Opin Biotechnol 22:775–783. https://doi.org/10.1016/j.copbio.2011.04.024.
6. Sarria S, Kruyer NS, Peralta-Yahya P. 2017. Microbial synthesis of medium-chain chemicals from renewables. Nat Biotechnol 35:1158–1166. https://doi.org/10.1038/nbt.4022.
7. Howard TP, Middelhaufe S, Moore K, Edner C, Kolak DM, Taylor GN, Parker DA, Lee R, Smirnoff N, Aves SJ, Love J. 2013. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. Proc Natl Acad Sci U S A 110:7636–7641. https://doi.org/10.1073/pnas.1215966110.
8. Liu X, Hicks WM, Silver PA, Way JC. 2016. Engineering acyl carrier protein to enhance production of shortened fatty acids. Biotechnol Biofuels 9:24. https://doi.org/10.1186/s13068-016-0430-4.
9. Ford TJ, Way JC. 2015. Enhancement of *E. coli* acyl-CoA synthetase FadD activity on medium chain fatty acids. PeerJ 3:e1040. https://doi.org/10.7717/peerj.1040.
10. Torella JP, Ford TJ, Kim SN, Chen AM, Way JC, Silver PA. 2013. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. Proc Natl Acad Sci U S A 110:11290–11295. https://doi.org/10.1073/pnas.1307129110.
11. Jing F, Cantu DC, Tvaruzkova J, Chipman JP, Nikolau BJ, Yandeau-Nelson MD, Reilly PJ. 2011. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochem 12:44. https://doi.org/10.1186/1471-2091-12-44.
12. Dillon SC, Bateman A. 2004. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. BMC Bioinformatics 5:109. https://doi.org/10.1186/1471-2105-5-109.
13. Cantu DC, Chen Y, Reilly PJ. 2010. Thioesterases: a new perspective based on their primary and tertiary structures. Protein Sci 19:1281–1295. https://doi.org/10.1002/pro.417.
14. Cantu DC, Rovira C, Reilly PJ. 2014. Molecular mechanism of a hotdog-fold acyl-CoA thioesterase. Chemistry 20:9045–9051. https://doi.org/10.1002/chem.201304228.
15. Mayer KM, Shanklin J. 2005. A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. J Biol Chem 280:3621–3627. https://doi.org/10.1074/jbc.M411351200.
16. Mayer KM, Shanklin J. 2007. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. BMC Plant Biol 7:1. https://doi.org/10.1186/1471-2229-7-1.
17. Ritchie MK, Johnson LC, Clodfelter JE, Pemble CW, Fulp BE, Furdui CM, Kridel SJ, Lowther WT. 2016. Crystal structure and substrate specificity of human thioesterase 2: insights into the molecular basis for the modulation of fatty acid synthase. J Biol Chem 291:3520–3530. https://doi.org/10.1074/jbc.M115.702597.
18. Feng Y, Wang Y, Liu J, Liu Y, Cao X, Xue S. 2017. Structural insight into

acyl-ACP thioesterase toward substrate specificity design. ACS Chem Biol 12:2830–2836. https://doi.org/10.1021/acschembio.7b00641.
19. Jing F. 2013. Characterization of acyl-ACP thioesterases for the purpose of diversifying fatty acid synthesis pathway. PhD dissertation. Iowa State University, Ames, IA.
20. Yuan L, Voelker TA, Hawkins DJ. 1995. Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proc Natl Acad Sci U S A 92:10639–10643.
21. Salas JJ, Ohlrogge JB. 2002. Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Arch Biochem Biophys 403:25–34. https://doi.org/10.1016/S0003-9861(02)00017-6.
22. Chan DI, Stockner T, Tieleman DP, Vogel HJ. 2008. Molecular dynamics simulations of the apo-, holo-, and acyl-forms of *Escherichia coli* acyl carrier protein. J Biol Chem 283:33620–33629. https://doi.org/10.1074/jbc.M805323200.
23. Roujeinikova A, Simon WJ, Gilroy J, Rice DW, Rafferty JB, Slabas AR. 2007. Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. J Mol Biol 365:135–145. https://doi.org/10.1016/j.jmb.2006.09.049.
24. Nguyen C, Haushalter RW, Lee DJ, Markwick PR, Bruegger J, Caldara-Festin G, Finzel K, Jackson DR, Ishikawa F, O'Dowd B, McCammon JA, Opella SJ, Tsai SC, Burkart MD. 2014. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. Nature 505:427–431. https://doi.org/10.1038/nature12810.
25. Crosby J, Crump MP. 2012. The structural role of the carrier protein—active controller or passive carrier. Nat Prod Rep 29:1111–1137. https://doi.org/10.1039/c2np20062g.
26. Val D, Banu G, Seshadri K, Lindqvist Y, Dehesh K. 2000. Re-engineering ketoacyl synthase specificity. Structure 8:565–566. https://doi.org/10.1016/S0969-2126(00)00146-5.
27. Gajewski J, Pavlovic R, Fischer M, Boles E, Grininger M. 2017. Engineering fungal *de novo* fatty acid synthesis for short chain fatty acid production. Nat Commun 8:14650. https://doi.org/10.1038/ncomms14650.
28. Zhu Z, Zhou YJ, Krivoruchko A, Grininger M, Zhao ZK, Nielsen J. 2017. Expanding the product portfolio of fungal type I fatty acid synthases. Nat Chem Biol 13:360–362. https://doi.org/10.1038/nchembio.2301.
29. Dehesh K, Edwards P, Hayes T, Cranmer AM, Fillatti J. 1996. Two nove1 thioesterases are key determinants of the bimodal distribution of acyl chain length of *Cuphea palustvis* seed oil. Plant Physiol 110:203–210. https://doi.org/10.1104/pp.110.1.203.
30. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345. https://doi.org/10.1038/nmeth.1318.
31. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340.
32. Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. https://doi.org/10.1093/bioinformatics/bti770.
33. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL repository and associated resources. Nucleic Acids Res 37:D387–D392. https://doi.org/10.1093/nar/gkn750.
34. Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30:S162–S173. https://doi.org/10.1002/elps.200900140.
35. Friedland GD, Lakomek NA, Griesinger C, Meiler J, Kortemme T. 2009. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. PLoS Comput Biol 5:e1000393. https://doi.org/10.1371/journal.pcbi.1000393.
36. Friedland GD, Linares AJ, Smith CA, Kortemme T. 2008. A simple model of backbone flexibility improves modeling of side-chain conformational variability. J Mol Biol 380:757–774. https://doi.org/10.1016/j.jmb.2008.05.006.
37. Smith CA, Kortemme T. 2008. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. J Mol Biol 380:742–756. https://doi.org/10.1016/j.jmb.2008.05.023.
38. Leaver-fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, Thompson J, Davis IW, Pache RA, Gray JJ, Kortemme T, Richardson JS, Havranek JJ, Snoeyink J, Baker D, Kuhlman B. 2013. Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol 523:109–143. https://doi.org/10.1016/B978-0-12-394292-0.00006-0.
39. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. 2003. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 331:281–299. https://doi.org/10.1016/S0022-2836(03)00670-3.
40. Cryle MJ, Schlichting I. 2008. Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. Proc Natl Acad Sci U S A 105:15696–15701. https://doi.org/10.1073/pnas.0805983105.
41. Qiu X, Janson C. 2004. Structure of apo acyl carrier protein and a proposal to engineer protein crystallization through metal ions. Acta Crystallogr D Biol Crystallogr 60:1545–1554. https://doi.org/10.1107/S0907444904015422.