Microbial engineering strategies for improving production of eight carbon oleochemicals

By

Néstor J. Hernández Lozada

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Chemical Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 10/22/2018

The dissertation is approved by the following members of the Final Oral Committee:

Brian F. Pfleger, Professor, Chemical and Biological Engineering
Eric V. Shusta, Professor, Chemical and Biological Engineering
John Yin, Professor, Chemical and Biological Engineering
Michael G. Thomas, Professor, Bacteriology
Philip A. Romero, Assistant Professor, Biochemistry
Christopher Stowers, Bioprocessing R&D Leader, Dow AgroSciences

To Mom

Acknowledgments

First, I would like to thank my mom, for all the unconditional love, hard work and sacrifices that you have done to get me this far. Thank you for raising me as an independent and curious individual and always support me in everything that is good. To Nésmary, for being a loving sister and my life long conscientiousness role model. To Nestito, for being a wonderful brother and a role model.

I would also like to thank my advisor, Dr. Brian Frederick Pfleger, for being a great boss and teaching me so many things over the years. I appreciate the opportunity to be a part of your laboratory and I am convinced that joining jour lab was one of the best career defining decisions I have made. Even though I never got that window office for my plants, I appreciate just about everything else that you have supported me with.

I would also like to thank my friends, especially Carlos, for being an exceptional friend and always being there in good and though times. I'm still convinced that I will see you in the UK! I would like to thank Joe for being a great classmate, roommate and friend. I really enjoyed our times hanging out, running, teaching me way too much about bikes, and boy can you keep a kitchen spotless. I would like to thank Prashant for being a great friend and collaborator, it's always a pleasure to spend time discussing serious and not so serious matters with you. To Xavier Martinez, for helping me with my scientific fair posters in middle school and turning a routine science project into an adventure. Going to the woods to find peronía seeds and tiny scorpions are some of the best memories I have from that time in my life.

During my graduate school years, I had several undergrads that helped me along the way. It was always a pleasure to have such a hardworking and smart students to always count on. I would like to thank Trevor R. Simmons for sticking with me for more than three years and helping me so much across multiple projects. To Ke Xu, thank you for helping me so much over the last year. I had a great time working with you. I know that both of you will do awesome in grad school. I would also like to thank my other undergrads that have helped me through grad school, Kelsey A. Thomas, Christian Breckner, Haley Schoenberger, Morgan Kitzerow and Vicky Alber.

I also had many mentors in my undergrad and graduate school years that contributed to my professional development. Especially, I would like to thank Dr. Alvin Negron, Dr. Eduardo Vazquez, Prof. David Suleiman, and Prof. Carlos Rinaldi for, in one way or another, giving me guidance. I would also like thank Dr. Christopher Stowers for all the advice you gave me at Dow Agrosciences as well as during graduate school.

Finally, I would like to thank the rest of the Pfleger lab for being great co-workers and always having good suggestions. In particular, I would like to thank Mark Politz, Chris Mehrer, Rung-Yi Lai, Mike Jindra, Travis Korosh, Andrew Markley, Jeff Cameron, Ryan Clark and Austin Comer. Finally, I would like to thank Dr. Daniel Mendez for teaching me molecular biology in my first year and answering so many of my questions.

Table of Contents

List of Figures

List of Tables

1. Medium chain oleochemicals[1]

1.1. Introduction to oleochemicals

The current unsustainable use of petroleum derived hydrocarbons as a source of chemicals has driven many researchers to look for sustainable and environmentally friendly alternatives[1]. One of the alternatives currently being explored is the use of different types of biomass. Biomass can be grown in large quantities and has the potential to serve as a feedstock for chemical and biological upgrading to a myriad of useful chemicals and fuels. There are three general types of biomass: starches, oils, and lignocellulosic biomass. Of these types of biomass, oils and starches are of the most commonly used in current industry. Biomass oils are obtained primarily from harvesting oil crops but also some animal fats (**Figure 1.1**). Plant and animal derived oils can be harvested and chemically transformed into various oleochemicals for use in a range of applications, e.g., food oils, surfactants, cosmetics, biodiesel, polymers[2–5].

[1] Sections of this chapter were adapted from work originally published as:

Mehrer C.R., Hernández Lozada N.J., Lai RY., Pfleger B.F. (2017) Production of Fatty Acids and Derivatives by Metabolic Engineering of Bacteria. In: Lee S. (eds) Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals. Handbook of Hydrocarbon and Lipid Microbiology. Springer, Cham.

N.H.L contribution involved the sections on fatty acid biosynthesis pathway and thioesterases.

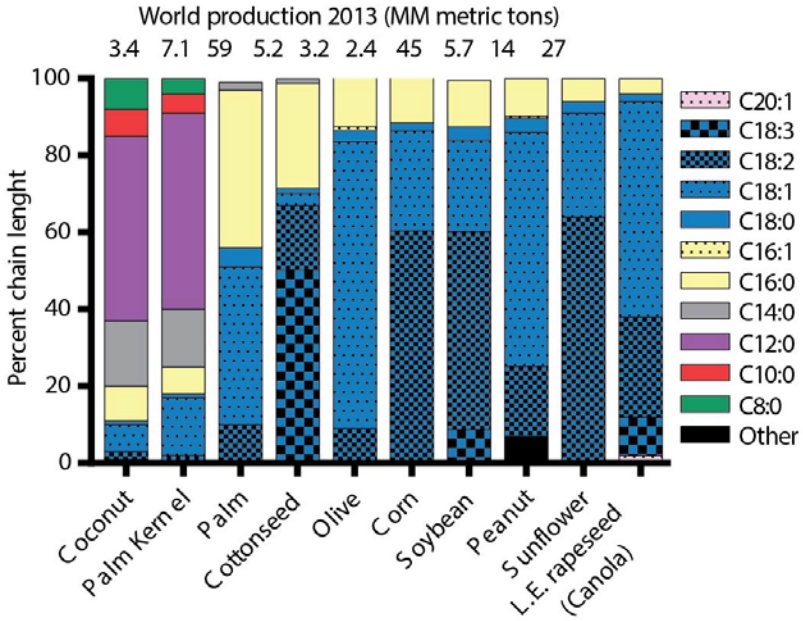


Figure 1.1. Chain length composition of various plant oils. Figure adapted from Hernandez Lozada, et al. 2018 ACS Synthetic Biology.

One challenge that the oleochemical industry has been unable to address is the limited availability of medium chain length fatty acid sources. Specifically, medium chain oleochemicals (C_8-C_{12} carbon chain lengths) are scarce, C_8 and C_{10} being only available as minor components of coconut palm oil and palm kernel oil (**Figure 1.1** & **1.2**). However, it is precisely these sources of oils that are some the most environmentally impactful. Growing these crops requires heavy deforestation in tropical rainforest regions of Asian countries such as Indonesia, Malaysia and Philippines[6]. Moreover, production of these sources has increased several fold over the last four decades (**Figure 1.2**). Taken together, this puts pressure on industry to find alternative routes for these medium chain lengths outside of both plant and petroleum derived oils. One possible solution is the microbial conversion

of lignocellulosic biomass derived sugars C_8-C_{12} oleochemicals instead of other established chemicals (e.g. ethanol).

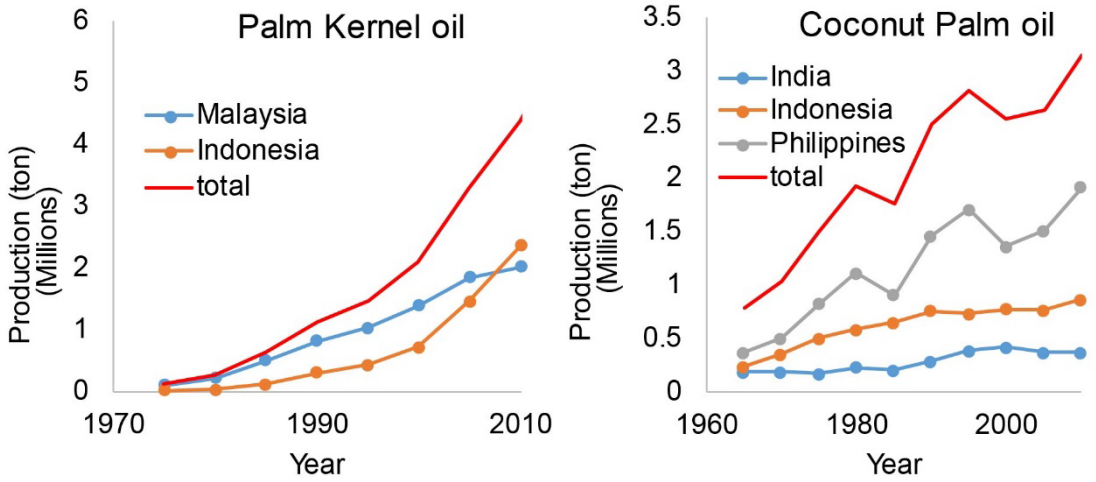


Figure 1.2: Historical production of palm kernel oil and coconut palm oil in major producing countries.

Lignocellulosic biomass is a more desirable source for the sustainable production of chemicals and fuels because lignocellulosic biomass is readily available as a by-product of existing crops and not edible, unlike starches and oils that are widely used for Calories. Lignocellulosic biomass is comprised of lignin, hemicellulose and cellulose, the latter comprising about 40 to 50% of the total dry weight[1]. The cellulose in biomass is a polysaccharide that can be hydrolyzed to release free glucose. Hydrolysis technologies, both chemical and biological, typically follow pre-treatment of the biomass with chemicals that remove barriers[1]. Biomass-derived glucose can be used as the primary source of carbon for microbial fermentation and other chemical production methods[7]. In particular, *Escherichia coli* (*E.* coli) has proven to be a robust industrial host to engineer

because it can be grown in large bio-fermenters, it can be easily genetically modified, and it has well understood regulation, metabolism and physiology[8]. Therefore *E. coli* is well suited for engineering the conversion of biomass derived sugars to important chemicals in a sustainable way. In particular, free fatty acid (FFA) production in *E. coli* has been subject of interest to find alternative routes for the oleochemical industry since it can provide with the same platform molecules that are currently obtained from plant oils (e.g. fatty alcohols, polyhydroxy alkanoates (PHA), methyl ketones, methyl esters)[2,9,10].

Significant progress has been made to produce long chain (C_{14}-C_{18}; lengths found in membrane lipids and storage granules) oleochemicals. Producing medium chains, on the other hand, remains a difficult challenge. There are multiple reasons for this such as higher product toxicity, the nature of the pathway to make long chains, lack of termination enzymes with high activity and specificity, among others[2,10]. In particular, over the past 7 years, our lab has focused on producing lauric acid (C_{12} FFA)[11,12] and some derivatives such as dodecanol[13] (C_{12} OH), C_{12} poly-hydroxy alkanoates[14] (C_{12} PHA). The focus of this thesis is identifying strategies and the required biological catalysts for producing octanoic acid (C_8 FFA; Chapter 2 and 3) and its derived oleochemicals: 1-octanol (Chapter 4), poly(3-hydroxyoctanoic acid) (Chapter 5), and 2-heptanone (Chapter 5). Of the several routes to these molecules, I elected to study routes based on the native fatty acid biosynthesis pathway as I shall discuss in the subsequent sections.

1.2. Fatty acid biosynthesis pathway

Fatty acid biosynthesis (**Figure 1.3**) is comprised of enzymes cells use to build hydrophobic acyl chains which are integral components of phospholipids in cell membranes and its composition a vital component of any healthy bacterial cell[15,16]. In *E. coli*, the cell uses the acyl-acyl carrier protein (ACP) as a functional handle for the enzymes involved in the pathway with the exception of the first two steps of the pathway, which use coenzyme A (CoA) in an analogous manner. The pathway starts with the carboxylation of acetyl-CoA to malonyl-CoA through an irreversible reaction catalyzed by acetyl-CoA carboxylase (ACC) (**1**) in which one ATP is hydrolyzed as a driving force. This step is followed by the FabD mediated transacylation of malonyl-CoA to produce malonyl-ACP (**2**) for use in both initiation and elongation reactions. For the initiation, malonyl-ACP goes through a Claissen condensation with a second acetyl-CoA molecule catalyzed by β-ketoacyl-ACP synthase III (KASIII, FabH in *E. coli*) to get acetoacetyl-ACP (**3**), the four carbon β-keto acyl-ACP precursor that initially enters elongation stage.

In most bacteria, the cycle of fatty acyl-chain elongation and reduction (**Figure 1.3**) begins with acetoacetyl-ACP and ends with the synthesis of C_{16} or C_{18} acyl-chains that comprise membrane phospholipids[16,15]. In each cycle, the β-keto position of a β-ketoacyl-ACP is reduced to a hydroxyl group by β-ketoacyl-ACP reductase (KR, e.g. FabG in *E. coli*) consuming one NADPH (**4**). Next, the β-hydroxy acyl-ACP is dehydrated by β-hydroxyacyl-ACP dehydratase (DH, e.g. FabZ and FabA in *E. coli*), creating a *trans* double bond between the α and β carbons (**5**). The resulting *trans*-2-enoyl-ACP is reduced to a saturated acyl-ACP

by enoyl-ACP reductase (ER, e.g. FabI in *E. coli*), consuming one NADPH (**6**). The next cycle begins when the saturated acyl-ACP is condensed with a malonyl-ACP molecule by β-ketoacyl-ACP synthases I and II (KAS, e.g. FabB and FabF in *E. coli*), releasing carbon dioxide and producing a new β-ketoacyl-ACP two carbons longer than the previous cycle (**7**). In addition, bacteria makes unsaturated fatty acids (UFA) to regulate the membrane fluidity by adjusting the saturated:unsaturated fatty acid ratio in phospholipids. Unsaturated fatty acid biosynthesis in *E. coli* branches out of the saturated fatty acid synthesis when the chain length of the enoyl-ACP reaches ten carbons in length. Two enzymes, FabA and FabB catalyze the initiation and elongation of UFA on the cell, respectively. Specifically, FabA enzyme isomerizes trans-2-decenoyl-ACP to cis-3-decenoyl-ACP (**8**) and FabB does the subsequent elongation of the unsaturated acyl-ACP produced (**9**). Both, the activation of acetyl-CoA to malonyl-CoA (**1**) and subsequent decarboxylation of malonyl-ACP during the condensation (**7**) provides a major thermodynamic driving force for this reaction and provides a “push” for acyl-chain synthesis through the FAB pathway[17,18]. In sum, each cycle consumes one acetyl-CoA, 1 ATP, and two reducing equivalents (e.g. NADPH). The pathway has also been reconstituted *in vitro* and the kinetic properties of each enzyme have been characterized with at least one substrate. Pathway activity was shown to be maximized with equal molar amounts of FabA, FabB, FabD, FabF, FabG, FabH, and elevated amounts of FabI (10X), FabZ (10X), and holo-ACP (30X) [19].

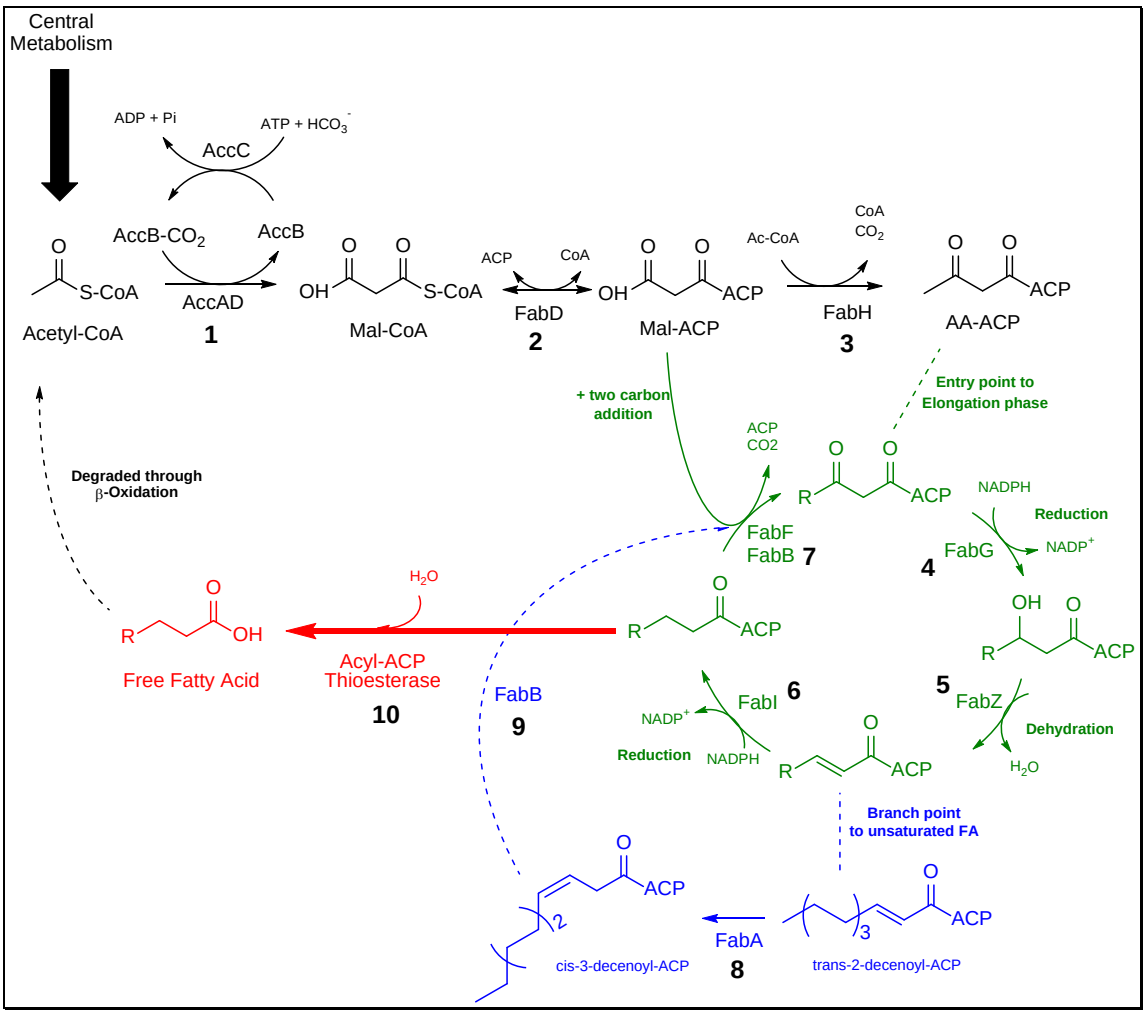


Figure 1.3. Fatty acid biosynthesis pathway. (black) initiation stage of the fatty acid biosynthesis pathway where acetyl-CoA is converted to acetoacetyl-ACP. (green) elongation phase of the FAB pathway where the acyl chain grows by two carbons per turn. (blue) unsaturated fatty acid branch point where trans-2-decenoyl-ACP is isomerized. (red) hydrolysis of acyl-ACP to make free fatty acids.

Regulation of FAB is best understood in a few model bacteria *E. coli* and *Bacillus subtilis*, where regulation occurs both transcriptionally and allosterically[20]. Because FAB is an energetically expensive process, bacteria use transcriptional regulation to strictly control flux through FAB with regulator proteins, such as FadR in *E. coli*, DesT in *Pseudomonas aeruginosa*, and FapR in *B. subtilis* [21]. From an

engineering perspective, overexpression of FadR in *E. coli*, which is known to transcriptionally activate FAB genes *fabA* and *fabB*, leads to increased fatty acid production when expressed in combination with a thioesterase[22]. The unsaturated fatty acid biosynthesis is also regulated by the transcriptional factor protein FabR which allosterically senses the concentration of long chain. Genes *fabA* and *fabB* are both downregulated by a transcription factor, FabR which binds the spacer between the -35 and -10 promoter sequence of both genes[23].

In *E. coli* when long-chain acyl-ACPs accumulate they are known to allosterically inhibit ACC, FabH, and FabI and thus slowing down overall lipid synthesis[16]. Expression of acyl-ACP thioesterases produces a pool of free fatty acids (FFA), thereby depleting long-chain acyl-ACPs (**Figure 1.3, 10**). This “pull” strategy removes the allosteric inhibition and increases flux through the FAB pathway. When coupled with the elimination of fatty acid catabolism (e.g. by deleting FadE or FadD in *E. coli*), thioesterase expression can lead to dramatically increased FFA titers in *E. coli* [9].

1.3. Thioesterase expression for the production of octanoic acid

Heterologous expression of thioesterases (TE), typically from bacteria or plants, is the most important strategy to engineer medium and long chain FFA (C_8-C_{18}) production in bacteria. In addition to catalyzing the termination step of the pathway, the expression of a thioesterase in *E. coli* also creates a “product sink”

from the elongation and reduction cycle, resulting in the deregulation of FAB in *E. coli*. A strategy for producing FFAs is completed by blocking the β-oxidation from consuming the FFAs produced by the thioesterase and sometimes overexpressing acetyl-CoA carboxylase to produce malonyl-CoA, an intermediate dedicated to FAB. The combination of these strategies has allowed for the successful production of different chain-length fatty acids[9].

Naturally occurring thioesterases have a high degree of diversity in specificity towards particular chain lengths. In plants, fatty acyl-chains are synthesized in the chloroplast and must be transported into the cytoplasm for incorporation in to lipids and oils. To enable transport across the chloroplast membrane, acyl-chains are cleaved by thioesterases and re-activated by acyl-CoA synthethases in the cytosol for incorporation into lipids [24,25]. Given the diversity in composition of plant oils, a wide range of plant thioesterases [26] have been expressed in *E. coli* for production of FFAs (for a literature survey of numerous thioesterases and their specificity profiles refer to **Appendix III)**. Although many thioesterases have a broad chain-length specificity, most of the success using the FAB pathway for the production of FFAs has come through the expression thioesterases with narrow product specificity to produce medium- to long-chain fatty acids[9].

While many acyl-ACP thioesterases have been characterized *in vivo* and their main substrate specificities reported (see **Appendix III**), the *E. coli* thioesterase ‘TesA and the plant thioesterase from *Umbellularia californica,* BTE, are among the most commonly used. A modified version of the native periplasmic

E. coli thioesterase, ‘TesA (a TesA variant lacking the native N-terminal secretion tag), has been successfully used to produce C14:0 fatty acids[27,28]. The plant thioesterase BTE from *Umbellularia californica* has been heterologously expressed in *E. coli* to produce C12:0 fatty acids [11,29]. A good example of a very specific thioesterase comes from the tree *Umbellularia californica* (BTE)[11,29]. This thioesterase has specificity of about 70% for dodecanoic acid (C_{12}) and 30% to other products (mainly C_{14}, tetradecanoic acid and C_{10}, decanoic acid). However, despite the efforts to find native enzymes with high substrate specificity, the other thioesterases found so far are less specific than BTE.

Unfortunately, in addition to having a limited pool of naturally occurring thioesterases that are highly specific, many plant thioesterases that might be highly specific, have limited activity as well. The limited activity is sometimes associated with truncation of N-terminal localization and membrane insertion sequences which can lead to solubility issues. Low thioesterase activity can also be due to thioesterase expression levels, which must be tuned in order to maximize production without severely impacting cell fitness[11,27]. However, there is evidence that points that many thioesterases simply may not have high specific activity[30]. In Chapter 3 we will address this point while engineering a low activity plant thioesterase from *Cuphea palustris*.

Therefore, being this step so important in the resulting product profile, it would be desirable to have thioesterases that are both, more active and specific for a particular chain length. The possibility of engineer thioesterases rationally suffers two main hurdles; (1) at the beginning of this thesis, only one crystal

structure of a thioesterase was known: broad specific *E. coli* thioesterase I (TesA); and (2) no methods had been elucidated yet for screening FFA production in a high throughput manner (making random mutation approaches unavailable). Despite of these hurdles, rational design of thioesterases to change the substrate specificity towards a desired chain length is feasible because the only difference between the substrates of interest is size. We will address this rational approach in Chapter 2 by engineering *E. coli* TesA.

In some cases, chain-length specificity can be improved by increasing the level of its preferred substrates in the elongation and reduction cycle. For example, the production of octanoic acid via a heterologous plant acyl-ACP thioesterase was increased in *E. coli* by increasing the abundance of octanoyl-ACP[31]. To do this, FabF (KASII) was replaced with an enzyme variant incapable of elongating acyl-ACP species past C_8, thereby making long-chain lipid synthesis dependent on FabB (KASI). To produce octanoic acid, FabB levels were decreased through proteolytic degradation, which was mediated by a degradation tag fused to FabB and an inducible chaperone. Degradation of the enzyme was showed via western blot. Once cells were induced, elongation of medium-chain acyl-ACP was prevented, leading to increased flux to octanoic acid[31].

1.4. Upgrading octanoic acid to other oleochemicals

Octanoic acid, like other free fatty acids, is a versatile molecule that can be transformed into several related chemical species (**Figure 1.4**). In this thesis I will focus the attention on three of these products: 1-octanol, poly(3-hydroxyoctanoic acid) and 2-heptanone. We will do this by first reactivating octanoic acid into octanoyl-CoA via acyl-CoA synthetase and then employing termination enzymes within the β-oxidation pathway. In the next sub-sections, I will give an overview of each of these products in more detail.

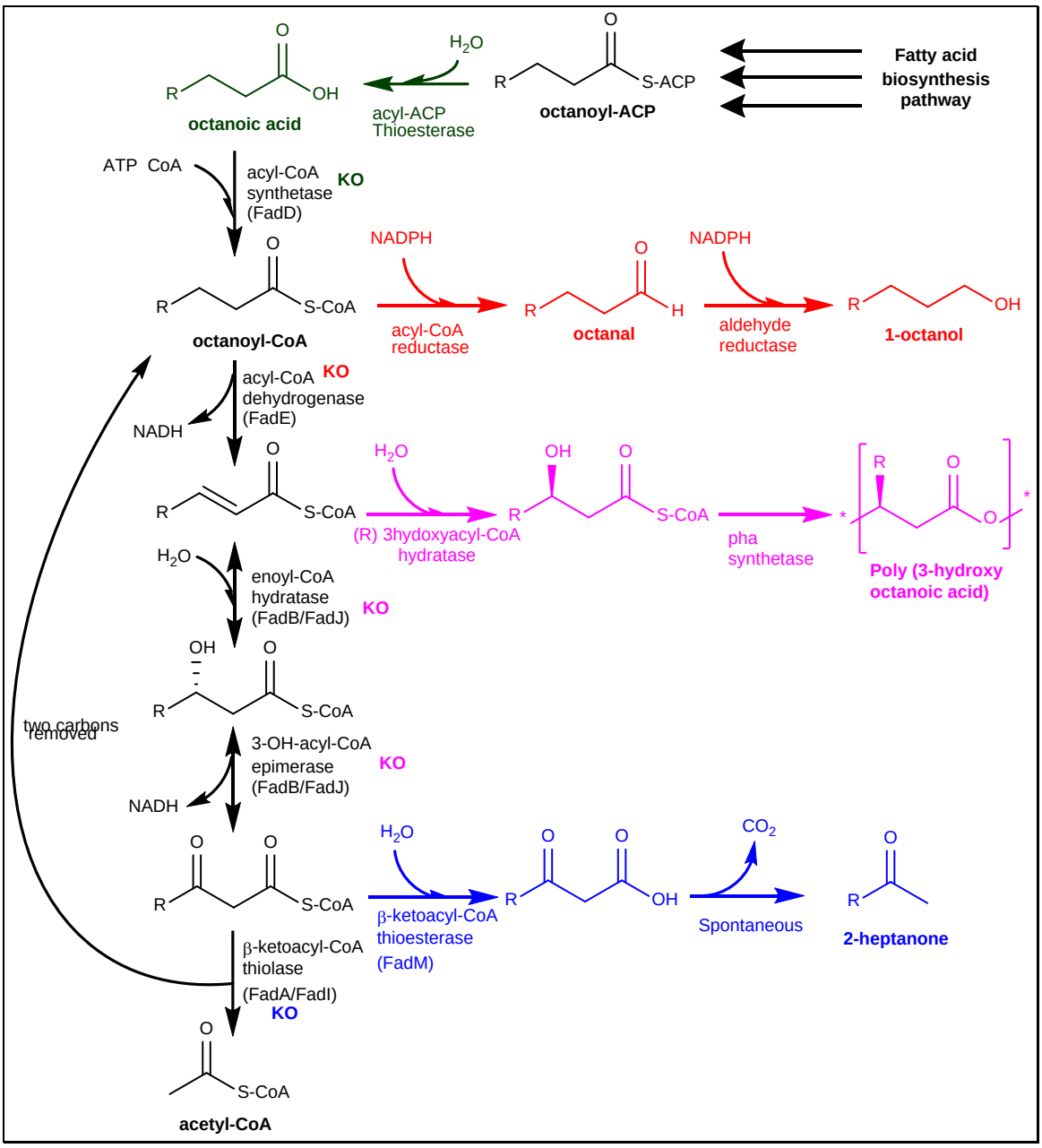


Figure 1.4. Thioesterase route to 1-octanol, Poly (3-hydroxyoctanoic acid) and 2-heptanone. Colored reactions represent the termination enzymes to produce various C8 oleochemicals. (green) thioesterase reaction to produce octanoic acid. (red) reductase reactions to convert octanoyl-CoA to 1-octanol. (magenta) stereospecific hydratase and polymerase reactions to produce Poly (3-Hydroxyoctanoic acid) and (blue) thioesterase reaction to convert b-ketooctanoyl-CoA to 2-heptanone. Necessary gene knockouts for each strategy are also highlighted in the corresponding colors.

1.4.1. **1-Octanol**

The most common method of producing fatty alcohols is through the use of an alcohol-forming acyl-CoA reductase, which catalyzes both the reaction from acyl-CoA to a fatty aldehyde and then to fatty alcohols. In this category includes AdhE2 from *C. acetobutylicum* [32] and Ma-ACR (Maqu_2507) from *Marinobacter aquaeolei* [33,34]. AdhE2, native to the Clostridial butanol pathway, has also been used to produce butanol in *E. coli* at high titers. When used in combination with a β-reduction pathway AdhE2 enabled production of hexanol and octanol [35]. The enzyme MAACR has activity toward longer-chain products [34] and has been used to produce fatty alcohols from both FAB and β-reduction pathways. Using a β-reduction pathway, MAACR enabled production of a mixture of C_6-C_{10} alcohols [36]. To obtain a narrow product distribution using FAB, the thioesterase BTE, an acyl-CoA synthetase (*E. coli* FadD), and MAACR were expressed in a strain lacking β-oxidation (Δ*fadE*), leading to production of largely C_{12} and C_{14} fatty alcohols [13]. A similar study using ‘TesA, FadD, and MAACR in *E. coli* demonstrated production of fatty alcohols distributed between C_{12} and C_{18} [37].

Fatty alcohols have also been produced using separate enzymes for the aldehyde production and the subsequent reduction to an alcohol. The acyl-CoA reductase Acr1 from *Acinetobacter calcoaceticus* produces fatty aldehydes[38] and has been used to produce longer-chain alcohols (≥C12) in combination with aldehyde reductases (AHR) in *E. coli* [39]. The carboxylic acid reductase (CAR) from *Mycobacterium marinum* has been used in combination with an AHR from

Synechocystis sp. PCC 6803 for the production of C_{12}-C_{18} fatty alcohols in *E. coli* [40]. An acyl-ACP reductase (AAR) from *Synechococcus elongatus* has also been shown to reduce acyl-ACPs directly to aldehydes; this method was used to produce C_{12}-C_{18} fatty alcohols in *E. coli*, where the native aldehyde reductase AdhP was shown to reduce the aldehydes to alcohols [41].

More recently, the β-reduction (also known as the reverse β-oxidation) pathway[42] has been engineered for C_4-C_{16} alcohol production. This pathway relies in the replacement of the acyl-CoA dehydrogenase in the β-oxidation (FadE, see **Figure 1.4**) with a trans-2-enoyl-CoA reductase which has activity in the reductive direction to form an acyl-CoA. Since the other enzymes in the β-oxidation pathway are reversible (FadB and FadA) the pathway is able to operate in the synthetic direction. This pathway has higher theoretical yields, as it does not rely on ACC carboxylation of acetyl-CoA to malonyl-CoA and can be run anaerobically[43]. A main drawback for this pathway is the lack of enzymes with high specificity, leading to heterogeneous product profiles.

1.4.2. **Poly (3-Hydroxyoctanoic acid)**

Poly(3-hydroxyalkanoates) (PHAs) are a class of biopolymers that have similar properties to commercial plastics, such as polypropylene, but are synthesized from intermediates of the β-oxidation/reduction pathway. Generally, PHA can be classified as short chain length (C_3-C_5) (scl-PHA) or medium chain (C_6-C_{14}) (mcl-PHA), based on the chain length composition of the 3-hydroxyacyl-

CoA monomers by which the PHA polyester is synthesized. scl-PHA, such as poly(3-hydroxybutyrate) (PHB) is a brittle and rigid plastic, with T_m of 180 °C whereas mcl-PHA have more elasticity and lower T_m (40-60 °C)[44]. In addition, scl-PHA and mcl-PHA have also different biosynthetic routes in the cell. The most studied scl-PHA, poly(3-hydroxybutyrate) (PHB), is produced by several species of bacteria, such as *Cupriavidus necator*. Pathways for PHB production are dedicated toward that product and proceed as a partial β-reduction pathway from acetyl-CoA via a β-keto thiolase (phaA) producing the intermediate acetoacetyl-CoA which can then be reduced enzymatically with a stereo specific acetoacetyl-CoA reductase (phaB) to (R)-3-hydroxybutyryl-CoA followed by polymerization with a PHA synthase (phaC)[45]. On the other hand, mcl-PHA are synthesized from one of three routes, (1) the transesterification route: medium chain 3-hydroxyacyl-ACP from the fatty acid biosynthesis pathway can be trans esterified to (R)-3-hydroxyacyl-CoA by phaG followed by polymerization by PHA synthase (phaC); (2) the thioesterase route: medium chain acyl-ACP from the fatty acid biosynthesis pathway can be hydrolyzed to produce medium chain free fatty acid via a thioesterase, followed by reactivation into acyl-CoA by an acyl-CoA synthethase (ACS), then the acyl chain is oxidized by acyl-CoA dehydrogenase (fadE) and hydrated by an stereo specific enoyl-CoA hydratase (phaJ) into (R)-3-hydroxyacyl-CoA which is finally polymerized with a PHA synthase (phaC) (**Figure 1.4**). This route also requires other β-oxidation genes, *fadBA* and *fadIJ* to be knocked out in order to prevent the cell from metabolizing the enoyl-CoA back into acetyl-CoA. The major advantage of this second strategy is the ability to control the polymer

chain length with the use of a specific thioesterase [14]. The third route follows as (2) from fed free fatty acids.

A β-reduction pathway has also been demonstrated to synthesize short and medium chain-length-PHAs by direct synthesis of the *trans*-2-enoyl-CoA. The medium-chain-length *trans*-2-enoyl-CoA can be hydrated with PhaJ1 and polymerized with PhaC2 from *Pseudomonas aeruginosa.* A heteropolymer of short- and medium-chain PHAs can be synthesized by the expression of the PHB pathway from *C. necator* along with PhaC2 from *Pseudomonas stutzeri* [46]. As we have seen with other pathways, the enzymatic diversity allows for a variety of strategies to be used for the production of PHAs.

1.4.3. **2-Heptanone**

Methyl ketones can be generated by thioesterase catalyzed hydrolysis of a β-keto-acyl-thioester and spontaneous decarboxylation of the β-keto fatty acid (**Figure 9**). Methyl ketones have been produced from fatty acid producing strains by the expression of a CoA ligase (e.g. FadD) followed by β-oxidation to a β-keto-acyl-CoA via the activity of FadE and FadB. When β-oxidation is blocked by deleting thiolases (Δ*fadA*), the cell accumulates β-ketoacyl-CoAs which can be cleaved by thioesterases such as FadM. Using this method, C_{11}-C_{15} methyl ketones have been produced to high titers in fed-batch fermentation [47,48]. The methyl ketone 2-pentanone has been produced using a β-reduction pathway similar to the butanol fermentation pathway in *C. acetobutylicium*. In this study, β-

ketohexanoate was generated though using the 3-oxoadipate CoA-succinyl transferase PcaIJ from *Pseudomonas putida*. Instead of relying on spontaneous decarboxylation, this study used acetoacetate decarboxylase, Adc from *C. acetobutylicum*, to catalyze the decarboxylation of β-ketohexanoate to produce 2-pentanone [49].

Although production of 2-heptanone has not been achieved in *E.coli*, its known to be produced naturally in the spores of the organism *Penicillium roqueforti*[50]*.* This is an organism that grows in blue cheese and 2-heptanone is known be key component of the flavor of this cheese. There has been multiple reports of high 2-heptanone production from fed octanoic acid[51,52] in *Penicillium roqueforti* but production of 2-heptanone from unrelated sugars has not been reported yet[10]. In *E. coli*, 2-heptanone production can be achieved in a similar manner to 1-octanol and PHA by expressing and octanoyl-ACP thioesterase to produce octanoic acid followed by its reactivation to octanoyl-CoA via acyl-CoA synthetase and using enzymes in the β-oxidation pathway; acyl-CoA dehydrogenase (FadE), enoyl-CoA hydratase/epimerase (FadB) to produce β-ketooctanoyl-CoA. Finally, 2-heptanone can be synthesized from the spontaneous decarboxylation of β-ketooctanoic acid which can be derived from the hydrolysis of β-ketooctanoyl-CoA species via a thioesterase reaction (**Figure 1.4**).

2. Computational redesign of acyl-ACP thioesterase with improved selectivity towards medium-chain fatty acids[2]

Abstract

Enzyme and metabolic engineering offer the potential to develop biocatalysts for converting natural resources into a wide range of chemicals. To broaden the scope of potential products beyond natural metabolites, methods of engineering enzymes to accept alternative substrates and/or perform novel chemistries must be developed. DNA synthesis can create large libraries of enzyme-coding sequences, but most biochemistries lack a simple assay to screen for promising enzyme variants. Our solution to this challenge is structure-guided mutagenesis in which optimization algorithms select the best sequences from libraries based on specified criteria (i.e. binding selectivity). Here, we demonstrate this approach by identifying medium-chain (C_6-C_{12}) acyl-ACP thioesterases through structure-guided mutagenesis. Medium-chain fatty acids, products of thioesterase-catalyzed hydrolysis, are limited in natural abundance compared to long-chain fatty acids; the limited supply leads to high costs of C_6-C_{10}

[2] This Chapter was originally published as:

Grisewood, M. J.*, Hernandez-Lozada, N. J.*, Thoden, J. B. 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

*These authors contributed equally to this work.

oleochemicals such as fatty alcohols, amines, and esters. Here, we applied computational tools to tune substrate binding to the highly-active ‘TesA thioesterase in *Escherichia coli*. We used the IPRO algorithm to design thioesterase variants with enhanced C_{12}- or C_8-specificity while maintaining high activity. After four rounds of structure-guided mutagenesis, we identified three thioesterases with enhanced production of dodecanoic acid (C_{12}) and twenty-seven thioesterases with enhanced production of octanoic acid (C_8). The top variants reached up to 49% C_{12} and 50% C_8 while exceeding native levels of total free fatty acids. A comparably sized library created by random mutagenesis failed to identify promising mutants. The chain length-preference of ‘TesA and the best mutant were confirmed *in vitro* using acyl-CoA substrates. Molecular dynamics simulations, confirmed by resolved crystal structures, of ‘TesA variants suggest that hydrophobic forces govern ‘TesA substrate specificity. We expect that the design rules we uncovered and the thioesterase variants identified will be useful to metabolic engineering projects aimed at sustainable production of medium-chain oleochemicals.

2.1. Introduction

Free fatty acids (FFAs) are energy-rich precursors of membrane lipids, natural oils, liquid transportation fuels (a.k.a. biodiesel), and high-value oleochemicals (e.g. fatty alcohols, aldehydes, olefins, and waxes)[2,11,53,54]. Oleochemical properties such as energy content, melting point, and volatility are

dictated by the chain length, degree of saturation, and branching pattern of the acyl-chain[55]. Fuels and oleochemicals derived from microbially produced FFAs could displace current, unsustainable plant feedstocks and reduce carbon footprints relative to petrochemical alternatives[56,57]. Unfortunately, natural sources of medium-chain length FFAs and lipids are significantly less abundant than longer chain compounds. The limited supply and costly petrochemical synthesis alternative leads to higher selling prices for medium-chain oleochemicals (e.g., 1-octanol costs approximately twice as much as 1-hexadecanol per pound)[2]. These economic drivers make bioproduction of medium-chain length FFA and oleochemicals an attractive opportunity if biosynthesis pathways with high yield and selectivity can be assembled and optimized.

While many oleochemical pathways have been demonstrated in model hosts, chain length selectivity remains an unsolved challenge. The product distribution of most metabolic engineering efforts has been restricted to the chain-length of the most abundant acyl-thioester in the cell or the distribution created by expression of a thioesterase. Acyl chains that comprise FFAs, lipids, and oleochemicals are made by an iterative series of elongation, keto-reduction, dehydration and enoyl-reduction reactions acting on acyl-coenzyme A (acyl-CoA) or acyl-acyl-carrier protein (acyl-ACP) thioesters[4]. An acyl-chain is elongated by two carbon atoms per cycle until it is trans-esterified (into phospholipids, waxes, or esters), reduced (to a fatty aldehyde or alcohol), or hydrolyzed (yielding a FFA)[53]. The chain length distribution of these terminal products is controlled by the relative kinetics of elongation, transesterification, reduction, and/or hydrolysis. In

Escherichia coli, the activity of phospholipid synthases (PlsB, PlsC) and fatty acid synthases (FabB, FabF, FabH) constrain the lipid composition to mostly C_{16} or C_{18} acyl-chains with little FFA content[58]. In contrast, *E. coli* can produce high titers of FFA with a wide range of chain-length distributions by the upregulation of native thioesterases (TesA and TesB) or heterologous expression of plant and bacterial thioesterases (see **Appendix III**). In the absence of pathways for catabolizing FFA and/or acyl-thioesters (i.e. beta-oxidation), the specificity of the acyl-ACP thioesterase controls the chain-length distribution and the chemical properties of downstream oleochemicals (see **Figure 2.1**). Collectively, thioesterases exhibit a wide range of substrate specificities[30,29,54,59–62]. that has been further diversified through protein engineering and evolution[63–65]. Even with this known diversity, very few thioesterases are specific towards a single aliphatic chain length[10]. Worse, heterologous expression of thioesterases frequently begets unexpected product distributions[4, 10], poor FFA yields[66,67], or both. Of the many studied thioesterases, ‘TesA (a cytosolic *E. coli* TesA lacking the N-terminal signal peptide) has been used extensively in metabolic engineering studies and has a crystallographically resolved structure[68]. However, ‘TesA has broad substrate specificity with relatively low medium-chain content[54,59]. For these reasons, 'TesA is an attractive system for applying enzyme engineering to improve thioesterase selectivity towards medium chain lengths.

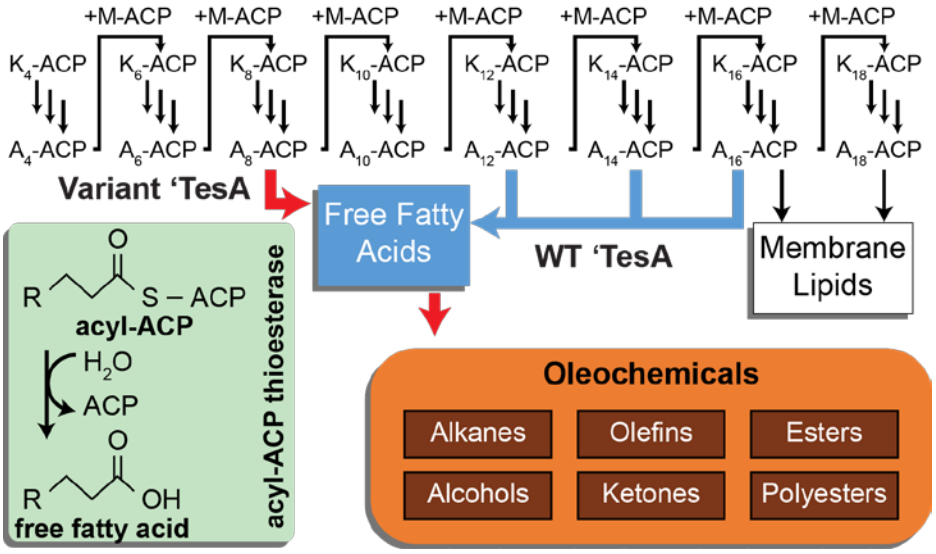


Figure 2.1: Overview of fatty acid biosynthesis and role of acyl-ACP thioesterases. In each cycle of fatty acid biosynthesis, two carbons are added from a malonyl-ACP (M-ACP) yielding a β-ketoacyl-ACP (K_x-ACP where x is the number of carbons). Three reactions (three vertical arrows) reduce the K_x-ACP to a saturated acyl-ACP (A_x-ACP). In *E. coli*, A_{16}-ACP and A_{18}-ACP are incorporated into membrane lipids. Thioesterases produce free fatty acids by hydrolyzing the acyl-thioester bond (green inset). Tailoring the specificity of the acyl-ACP thioesterase ('TesA) dictates free fatty acid and downstream oleochemical chain lengths (red arrows).

Enzyme engineering is generally pursued using directed evolution approaches that rely on high-throughput screening of large mutant libraries[69]. These large libraries are constructed using various mechanisms for diversifying the gene pool, including homologous (e.g., DNA shuffling) or non-homologous (e.g., overlap extension PCR) recombination, random mutagenesis (e.g., error-prone PCR), or combinations thereof (see Ref.[70] for review). Large library sizes rely on high-throughput screening that takes advantage of optical properties, such as fluorescent or colorimetric assays. Currently, no high-throughput screen that can discriminate between different FFA chain lengths has been developed. When optical screens are unsuitable, more laborious experiments (e.g., mass

spectrometry or NMR spectroscopy) can be used in low-throughput screens that mandate small, more focused libraries. Focused libraries can be generated through site saturation mutagenesis (e.g., degenerate oligonucleotide-primed PCR) but this approach can only include a small number of sites (three saturated sites would yield 8,000 variants) with a high percentage of inactive mutants. Smaller library sizes can be formed through site-directed mutagenesis[71] (i.e., rational design) but so far it has been very difficult to forecast the effect of multiple mutations. When a protein structure is known, molecular modeling tools can suggest a handful of promising mutations while considering sequence-structure relationships as well as ther approximations.

Structure-based protein redesign procedures such as the Iterative Protein Redesign and Optimization (IPRO) method[72] used here, offer several advantages over tools that simply suggest “hot spot” residues[73–78], as they can capture the simultaneous effect of multiple mutations. Another class of computational protein redesign tools uses mostly sequence information to suggest crossover locations for generating combinatorial libraries[79–85]. However, these combinatorial libraries are limited by the parental sequence space and do not take full advantage of available structural information. *De novo* enzyme designs[86–90] are usually less active than native ones (without the aid of directed evolution)[86,87,89]. Thus, structure-based protein redesign can harness natural protein performance and suggest directed modifications using structural insight to meet a single or multiple design objectives. IPRO differs from other structure-based protein redesign procedures in that it employs a mixed-integer linear program to guarantee a global

minimum for a given protein backbone structure, it can handle multiple decision criterion simultaneously, and distance restraints can be easily imposed to keep catalytic machinery intact[72,91,92]. More extensive reviews of existing computational protein engineering procedures have been provided by Pantazes *et al.*[93], Samish *et al*.[94], and Huang *et al*.[88].

In this study, we applied the Iterative Protein Redesign and Optimization (IPRO) method[72] to guide 'TesA mutagenesis in search of variants that both improve medium-chain FFA specificity and maintain high thioesterase activity. Specifically, we engineered 'TesA to yield additional dodecanoic acid (C12:0) and octanoic acid (C8:0) at the expense of the natively preferred tetradecanoic acid (C14:0). The employed redesign procedure involved recursively predicting *in silico* 'TesA mutants with enhanced binding capabilities, analyzing the *in vivo* FFA composition, and modifying the computational explorations in a Design-Build-Test-Learn cycle[95]. We identified three separate 'TesA mutants that exhibited a statistically significant ($p<0.05$) improvement in C_{12} composition over wild-type (WT) and twenty-two mutants with a statistically significant improvement in C_8 composition. In comparison to previously studied thioesterases (see **Appendix III**), our computationally predicted mutants include four of the ten most C_{12}-specific and one of the ten most C_8-specific thioesterases. While computational enzyme redesign that reaches industrially relevant performance metrics has so far remained elusive[93], the results presented here demonstrate the potential of the adopted Design-Build-Test-Learn paradigm to pinpoint promising enzyme mutants[93,95,96].

2.2. Materials and methods

2.2.1. TesA model construction

The structure of ‘TesA was derived from PDB 1U8U, where it is in complex with octanoic acid[68]. The acyl-ACP structures were derived from PDB 2FAE, where decanoyl-ACP is held in an internal binding cavity[97]. Other acyl-ACP structures include hexanoyl-ACP from PDB 2FAC and heptanoyl-ACP from PDB 2FAD[97] are similar to decanoyl-ACP as demonstrated by respective all-atom RMSDs of 1.1 Å and 1.1 Å. In order to dock octanoyl-ACP with ‘TesA, the acyl chain was systematically rotated about the phosphopantetheine linker and superimposed with the bound octanoic acid in 1U8U. The rotation that led to the lowest root-mean-square deviation was energy-minimized within CHARMM34[98]. Acyl-ACPs with different chain lengths were adapted from this initial complex by either deleting atoms or adding atoms using CHARMM’s internal coordinate system. Lazaridis-Karplus solvation files and CHARMM input files were constructed using published parameters for lipids and proteins. The constructed topology and parameter files were in close agreement with CGenFF-derived parameters[99,100].

2.2.2. Structure-based redesign and analysis

All computationally-predicted mutants were identified using multiple IPRO trajectories that each ran for 1000 iterations without ensemble structure refinements[72]. The primary constraint of each trajectory was to eliminate binding

to a larger acyl-ACPs (C_{14} for Rounds 1-3, C_{12} for Round 4) with a secondary constraint to improve binding to the shorter acyl-ACPs (C_{12} for Rounds 1-3, C_8 for Round 4). Round 4 C_8/C_{14} and C_8/C_{12} fraction ratios correlate with $r = 0.88 \pm 0.02$, alluding to a similar repulsive force for C_{12} and C_{14} binding (**Figure 2.2**). Design position selection is described in the "Design Position Selection" subsection. Restraints were imposed to ensure that the intermolecular catalytic distances (± 0.2 Å) were maintained (i.e., S10, G44, N73, D154 and H157). All other IPRO parameters were set to their standard values, and calculations were run on the Lion-XF system at Penn State University. Error propagation was performed manually and replicated using the Python uncertainties module[101]. Statistical differences were calculated using Welch's t-test between the WT and mutant FFA profiles. For a given enzyme-FFA complex, the interaction energy is found using $IE=G_{Enz\text{-}FFA,min}-G_{Enz}-G_{FFA}$. Mutants were sorted by the interaction energy difference between the short-chain FFA and C_{14} ($\Delta IE=IE_{C12,C8}-IE_{C14,C12}$). Therefore, for a given round, Mutant 1 (smallest ΔIE) would be expected to show the biggest change in specificity.

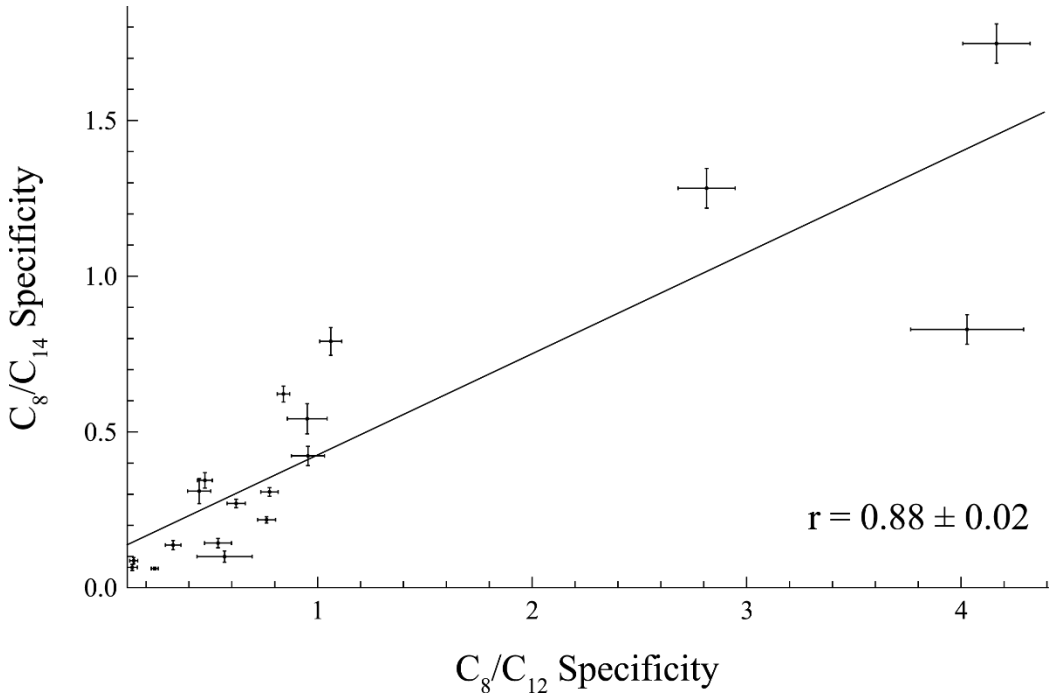


Figure 2.2. Correlation between R4 C_8/C_{12} and C_8/C_{14} molar composition. The strong correlation (r=0.88±0.02) between these values supports the concept that reduced binding site volume for C_{12} binding furthermore eliminates binding to C_{14} due to its longer aliphatic chain. The slope for the linear regression is less than one, which could imply an alternative binding mode for C_{14} in addition to the conformation assumed for C_{12} binding. Error bars are standard deviations in mole fractions propagated from Appendix III.

2.2.3. **Design Position Selection**

For R1, 'TesA residues were sorted by distance to carbon atoms in the acyl group of the 'TesA:tetradecanoyl-ACP complex (i.e., C-1 through C-14). The minimum interatomic distance between the residue's heavy atoms (i.e., not hydrogen atoms) and the acyl carbon atoms was used for sorting. In R1, residues constituting the catalytic triad (i.e., S10, D154, H157) and the oxyanion hole (i.e., S10, G44, N73) were not considered during design position selection. At the time design positions for R1 were designated, IPRO was unable to handle mutations

from proline so these residues were not considered during the selection of design positions[72]. L109 was considered to be important for ‘TesA functionality and was also removed from consideration as a design position[68]. Residues that were near the undesired end of the acyl group (i.e., near C-1 instead of ω-1) were no longer regarded as potential design positions. Residues considered “near C-1” were those that contained a heavy atom within 4.5 Å of the thioester sulfur atom. A final set of residues that were strongly oriented away from the ω-1 terminus of the acyl moiety were also no longer considered as potential design positions. Residues “strongly oriented away from the ω-1 terminus” were those whose Cα atom was over 0.75 Å closer to the ω-1 atom than the Cβ atom (glycines not considered). The sorted residues are provided in **Appendix VI** with any exceptions annotated. The eight nearest residues that were not filtered out were selected as design positions. The set of design positions used for R1 were L11, G72, L76, I107, R108, A111, F139, and Y145.

For R2, ‘TesA residues were once again sorted by distance to carbon atoms in the acyl group of the ‘TesA:tetradecanoyl-ACP complex. Residues constituting the catalytic triad and oxyanionic hole, proline residues, residues near the C-1 terminus of the acyl moiety, and L109 were once again removed from consideration, as described for the R1 design position selection procedure. In addition, residues belonging to the three flexible loops (i.e., $loop_{31-35}$, $loop_{75-80}$, $loop_{111-120}$) were no longer recognized as prospective design positions for the same reason that L109 was removed from consideration[68]. Finally, I107 and R108 were removed as possible design positions because these positions invariably mutated

to lysines in R1 and all instances of I107K and R108K abolished catalytic activity (see **Figure 2.3**, **Appendix IV**). The set of design positions (i.e., the eight nearest residues that were not filtered out) employed for R2 was L11, G72, F139, M141, E142, Y145, G155, and I156.

For R3 and R4, a very different approach was used when compared to R1 and R2. Instead of sorting residues by distance to carbon atoms in the acyl group of 'TesA:tetradecanoyl-ACP, residues were sorted by distance between the residue's Cβ atom and the ω-1 atom of dodecanoyl-ACP from the 'TesA: dodecanoyl-ACP complex. In lieu of the Cβ atom, Cα was used to calculate interatomic distances for glycine residues, and Cγ was used for H180 because Cα and Cβ were not part of the solved crystal structure[68]. Unlike R1 and R2, only residues that were aligned to gaps or had ≥40% sequence conservation were not considered as candidate design positions. Sequence alignment was performed using the conserved domain database, where 81 members (including 'TesA) of the lysophospholipase L1-like subgroup from the SGNH-hydrolase superfamily were found[102]. The sorted residues are provided in **Appendix VI**. Residues aligned to gaps or conserved residues are also noted in **Appendix VI**. The final set of design positions (i.e., the eight nearest residues that were not filtered out) used for R3 and R4 was I107, R108, L109, S122, M141, E142, Y145, and L146.

2.2.4. **Scoring function re-weighting**

A new set of weights for the IPRO scoring function was found that approximately doubles native rotamer recovery relative to the unmodified scoring function. The scoring function was modified using a symmetric logistic regression within Weka, a collection of machine-learning algorithms[103]. A dataset of native and non-native rotamers was collected from the top8000 database, which is a dataset of 8000 high-resolution (< 2 Å), quality-filtered (< 2.0 MolProbity score[67]), nonhomologous (<70% identity) protein structures[104]. Of these structures, 50 were randomly selected for use with the machine-learning training set. A separate set of 80 structures was randomly selected to validate the results. From these 130 structures, the native rotamer was found by finding the rotamer (of the same amino acid type) with the lowest root-mean-square deviation to the crystallized side chain. Then the van der Waals, electrostatic, and Lazaridis-Karplus solvation energies were calculated for each rotamer (regardless of amino acid type) as well as a binary indicator as to whether this was the native rotamer at the position or not. The data was separated for residues at the protein surface (≤ 20 C^{β} atoms within 10 Å) and within the core of the protein (> 20 C^{β} atoms within 10 Å) using a distance-based metric developed by Kuhlman and Baker[105]. Finally, since the number of non-native rotamers heavily outweighed the number of native rotamers, non-native rotamers were randomly removed until there was approximately a 60:40 split of non-native:native rotamers. The rotamer data was used to determine the set of weights that can optimally classify a rotamer as native or non-native. The updated energy equation for the core residues is $E = 0.04\ E_{VDW} + 0.02\ E_{Elec} + 0.16$

E_{LK}. For the surface residues, the updated energy equation is $E = 0.03\ E_{VDW} + 0.01\ E_{Elec} + 0.09\ E_{LK}$. The inaccuracy of the original scoring function that used equally-weighted energy terms (i.e., $E = 0.01\ E_{VDW} + 0.01\ E_{Elec} + 0.01\ E_{LK}$) stemmed mostly from the Lazaridis-Karplus implicit solvation energy term.

Using the modified scoring function weights, the classifier was able to correctly classify 78.4% of the core rotamers and 68.1% of the surface rotamers as either native or non-native from the validation set. Revisiting the complete validation set (before modifying to the 60:40 ratio), the full set of rotamers for each position was sorted from lowest energy to highest. The native rotamer was found to be the lowest-energy rotamer for 8.48% of the total dataset for the modified scoring function, up from 3.46%. Furthermore, the native rotamer was in the top 1.5%, 3.0%, and 6.0% of the energy-sorted rotamers in 19.09%, 31.64%, and 45.89% of the time. This was up from 9.09%, 16.86%, and 30.36%, respectively, for the original scoring function. Similar improvements were demonstrated on two different validation sets: one used to train the Rosetta scoring function[69] (native: 8.38% (3.57%), 1.5%: 18.56% (8.79%), 3.0%: 13.01% (16.30%), 6.0%: 45.07% (29.20%)) and one on high-quality antibody structures (native: 8.02% (4.00%), 1.5%: 16.68% (9.05%), 3.0%: 28.31% (16.46%), 6.0%: 43.10% (30.19%)). The success of the modified scoring function on diverse sets of protein structures that do not include ‘TesA implies that the new scoring function is equally accurate across most (if not all) systems.

2.2.5. DNA synthesis and ‘TesA variant construction

All mutants were created starting with WT ‘*tesA* gene cloned into a pBAD18 plasmid[106] to link ‘TesA expression to the presence of L-arabinose. Round 1 mutants were constructed using Agilent technologies QuickChange II site directed mutagenesis kit following the given protocol. For Rounds 2-4, all mutants were constructed using Gibson assembly strategies using primers containing the desired mutations in the 5' tails. All cloning was performed in *E. coli* DH5α strain.

‘TesA in pBad18 was randomly mutagenized using Gene Morph II random mutagenesis kit following the kit protocol to make a library of 61 mutants with a mutation rate of 1.8 amino acids per gene (**Appendix IV**). Primers were designed to include the start and stop codons to ensure keeping those positions of the mutants invariant. See **Appendix I-II** for a list of strains and plasmids used in this study.

2.2.6. Bacterial culturing and fatty acid production

FFA production was assayed from small batch cultures (5-50 mL) of *E. coli* strain RL08ara (K-12 MG1655 Δ*fadD* Δ*araBAD* ΔaraFGH Φ(Δ*araEp* P_{CP18}-*araE*))[11] harboring each thioesterase expression vector. Three single colonies of each mutant were grown overnight on LB media containing 100 mg/L of ampicillin. Overnight cultures of each strain were diluted 1:100 into 25 mL of LB media containing 100 mg/L of ampicillin and 0.4% w/v glycerol in a 250 mL baffled shake flask and grown at 37°C and 250 rpm. When the OD_{600} reached 0.2-0.3, cultures

were induced with 0.2% w/v L-arabinose and shaken for 24 h. All mutants were tested in triplicate and error bars represent the standard error of the measurements.

2.2.7. **Lipid extraction and quantification**

After 24 h post induction, 2.5 mL culture samples were collected in 10 mL glass centrifuge tubes, and 5 µL of 10 g/L heptadecanoic acid in ethanol solution was added as an internal standard. For fatty acid extraction into a chloroform layer, 100 mL of glacial acetic acid was added, followed by 5 mL of a 1:1 v/v solution of chloroform and methanol. Samples were vortexed and centrifuged for 10 min at 1000g to separate the layers. The chloroform extract was dried using a SpeedVac SC250EXP concentrator at no heat setting for 75 minutes and 1.0 torr. Samples were further dried for 30 min in a lyophilizer to remove any residual liquid. To methylate the dried extract, 0.5 mL of 1.25 M HCl in methanol was added and left overnight at 50°C. Finally, 5 mL of a 100 g/L sodium bicarbonate solution was added and fatty acid methyl esters were then extracted twice with 0.5 mL hexane for GC-FID quantification. Samples were collected and analyzed using a GC-FID model Shimadzu GC-2010 equipped with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The GC temperature protocol was 100°C for 2 min, ramp to 150°C (at 80°C/min), hold for 4 min, ramp to 218°C (at 4°C/min), ramp to 250°C (at 8°C/min), and hold for 2.5 min.

2.2.8. **Protein expression and purification of WT ‘TesA**

The WT ‘*tesA* gene was cloned into pET28t, a vector previously modified to contain a TEV protease site rather than a thrombin cleavage site between ‘*tesA* and a N-terminal polyhistidine tag[107]. When translated, the modified enzyme was fused to the following peptide: MGSSHHHHHHSSENLYFQGGGG. The pET28t-*tesA* plasmid was used to transform *E. coli* Rosetta2(DE3) cells (Novagen). Cultures were grown at 37°C with shaking in lysogeny broth supplemented with 50 mg/L kanamycin and 50 mg/L chloramphenicol until the OD_{600} reached 0.8. Flasks were cooled in an ice bath, induced with 1 mM isopropyl-β-D-thiogalactopyranoside, and incubated overnight at 21°C. Cells were harvested by centrifugation and frozen as pellets in liquid nitrogen. The frozen cell pellets were sonicated on ice in a lysis buffer composed of 50 mM sodium phosphate, 20 mM imidazole, 10% glycerol, and 300 mM NaCl (pH 8.0). The lysate was cleared by centrifugation, and ‘TesA was purified at 4°C utilizing Ni-nitrilotriacetic acid resin (Qiagen) according to the manufacturer’s instructions. TEV protease was added in a 1:20 molar ratio to the pooled protein solution and subsequently dialyzed against 50 mM sodium phosphate, 300 mM NaCl, and 20 imidazole (pH 8.0) at 4°C for 36 h. Both the TEV protease and the uncleaved protein were removed by passage over a Ni-nitrilotriacetic acid resin. The cleaved protein was collected and dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl and concentrated to ~35 g/L based on an extinction coefficient of 0.62 $(g/L)^{-1}cm^{-1}$.

2.2.9. **Crystallization and structural analysis of WT ‘TesA**

Crystallization conditions for ‘TesA were surveyed by the hanging drop method of vapor diffusion using a laboratory-based sparse matrix screen. The enzyme was initially tested either in the presence or absence of 3 mM C_8 FFA. Crystals were subsequently grown from 22-26% poly(ethylene glycol) (PEG) 5000 with 100 mM Homo-PIPES buffer (pH 5.0). The protein solution used contained 3 mM C_8 FFA. Crystals belonged to the monoclinic space group $P2_1$ with unit cell dimensions of *a* = 40.9 Å, *b* = 82.1 Å, *c* = 53.9 Å, and β = 90.4°. The asymmetric unit was comprised of two monomers. Prior to X-ray data collection at 100K, the crystals were transferred to a cryoprotectant solution composed of 30% PEG 5000, 250 mM NaCl, 3 mM C_8 FFA, and 13% ethylene glycol with Homo-PIPES buffer (pH 5.0). An X-ray data set was collected with a Bruker AXS Platinum-135 CCD detector using the PROTEUM software suite (Bruker AXS Inc.) The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. Data were processed with SAINT and scaled with SADABS (Bruker AXS Inc.). X-ray data collection statistics are listed in **Appendix VII**. The structure of WT ‘TesA was solved via molecular replacement using the software package PHASER[108] and the PDB 1U8U as the search model[68]. Model refinement with REFMAC[109] and manual model building with COOT[110,111] reduced the overall *R*-factor to 18.8% at 1.65 Å resolution. Refinement statistics are presented in **Appendix VIII.**

2.2.10. Crystallization and structural analysis of R3.M4

The R3.M4 ‘tesA gene was subcloned, expressed, and purified as described for WT. Crystals were obtained at both pH 5.0 and pH 7.5. Those obtained at pH 5.0 were grown from 20-25% PEG 5000 with 100 mM Homo-PIPES buffer (protein solution contained 3 mM C_8 FFA). The crystals were cryoprotected as described for WT ‘TesA. Crystals belonged to the monoclinic space group $P2_1$ with unit cell dimensions of a = 40.7 Å, b = 55.2 Å, c = 42.3 Å, and β = 105.2^o with a single monomer in the asymmetric unit. R3.M4 crystals obtained at pH 7.5 were grown from 24-28% PEG with 100 mM HEPES buffer (pH 7.5). Again the protein solution contained 3 mM C_8 FFA. These crystals were isomorphous to those obtained at pH 5.0 and were cryoprotected with a solution composed of 32% PEG, 250 mM NaCl, 3 mM C_8 FFA, and 13% ethylene glycol with 100 mM HEPES buffer (pH 7.5). Given the concern that at pH 5.0 the C_8 FFA would most likely not bind at full occupancy, these crystals were subsequently moved in a final experiment to solutions buffered at pH 7.5 that contained an additional 3 mM C_8 FFA. These “soaked” crystals were cryoprotected in a similar manner to those grown at pH 7.5. X-ray data from R3.M4 crystals obtained at pH 5.0, pH 7.5, and from the “soaked” crystals were collected as described for WT. The structure R3.M4 at pH 5.0 was solved by molecular replacement using the WT model as the search probe whereas the structures either grown or soaked at pH 7.5 were solved via Fourier difference analyses. X-ray data collection statistics and model refinement statistics are provided in **Appendix VII and VIII**.

2.2.11. Purification for enzymatic assays of WT ‘TesA and R3.M4

The ‘tesA gene was subcloned as described for WT. Expression followed the same procedure, except the construct was transformed into BL21(DE3) cells. The frozen cell pellets were sonicated on ice in a lysis buffer composed of 50 mM sodium phosphate, 10 mM imidazole, and 300 mM NaCl (pH 8.0). The lysate was cleared by centrifugation, and ‘TesA was purified at 20°C utilizing Ni-nitrilotriacetic acid resin (Qiagen) according to the manufacturer’s instructions. TEV protease was added in a 1:20 molar ratio to the pooled protein solution and subsequently dialyzed against 10 mM Tris-HCl (pH 7.5) at 4°C for 18 h. Both the TEV protease and the uncleaved protein were removed by passage over a Ni-nitrilotriacetic acid resin. The cleaved protein was concentrated to 2.5 mL to be solvent exchanged into 50mM potassium phosphate (pH 7.0), and 30% glycerol using a PD-10 desalting column (GE) according to the manufacturer’s instructions.

2.2.12. Enzymatic assays of WT ‘TesA and R3.M4

Enzymatic thioesterase assay was performed with WT and R3.M4 ‘TesA to compare their activities on various chain lengths. The reaction conditions follow as per Shin *et al.* ([112]), except the enzyme concentration used was 40 nM and substrate concentrations ranged 0-120 µM for six saturated acyl-CoA substrates of 6-16 carbons (hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, dodecanoyl-CoA, tetradecanoyl-CoA and hexadecanoyl-CoA). The assay tracks generation of free-CoA as ‘TesA hydrolyses the thiol bond in the acyl-CoA. This hydrolysis is tracked

by the increase of absorbance at 412 nm due to the free-CoA dependent reduction of 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB) present in the reaction mixture (63). Absorbance at 412 nm was followed using a NanoDrop 2000c (Thermo Scientific) using a path length of 10 mm and measurements were taken for 2 minutes in 10 second intervals.

2.2.13. Molecular dynamics

VMD was used to solvate enzyme-FFA complexes within a 12.0 Å water box with 0.17 M NaCl and contained ≈49,000 atoms[113]. Each complex was minimized and slowly heated to 310K and 1 atm over 7 ns using Langevin dynamics. Force field parameters were identical to those used for the IPRO trajectories. Periodic boundary conditions were applied, and long-range electrostatic forces were considered using the particle mesh Ewald method. 40 ns production simulations were performed using NAMD over 30 nodes on the Lion-XF cluster at Penn State University using the NVE ensemble[114].

2.3. Results and Discussion

2.3.1. Overview of the Design-Build-Test-Learn approach

In support of efforts to alter the product profile of ‘TesA towards medium chain length FFAs, we performed four rounds of site-directed mutagenesis structured around a Design-Build-Test-Learn cycle (**Figure 2.3**). In each round, mutagenesis targets and specific amino acid substitutions were selected after analyzing simulations of enzyme-substrate binding performed with IPRO. Genes encoding the designed ‘TesA variants were constructed by Quikchange® or Gibson Assembly® of PCR products (Materials and Methods) and cloned into arabinose-inducible expression vectors. The impact of each mutation was assessed by quantifying the FFA content of a ‘TesA expressing *E. coli* culture (see **Appendix III**). After each round, the product distributions for each mutant were analyzed and used to improve to the IPRO framework via changes in the scoring function parameters or refocusing the mutagenesis targets. The following sections describe the deployment of our computational enzyme Design phase, results of the Build-Test phase, and a discussion of what we have Learned about thioesterase selectivity after each round.

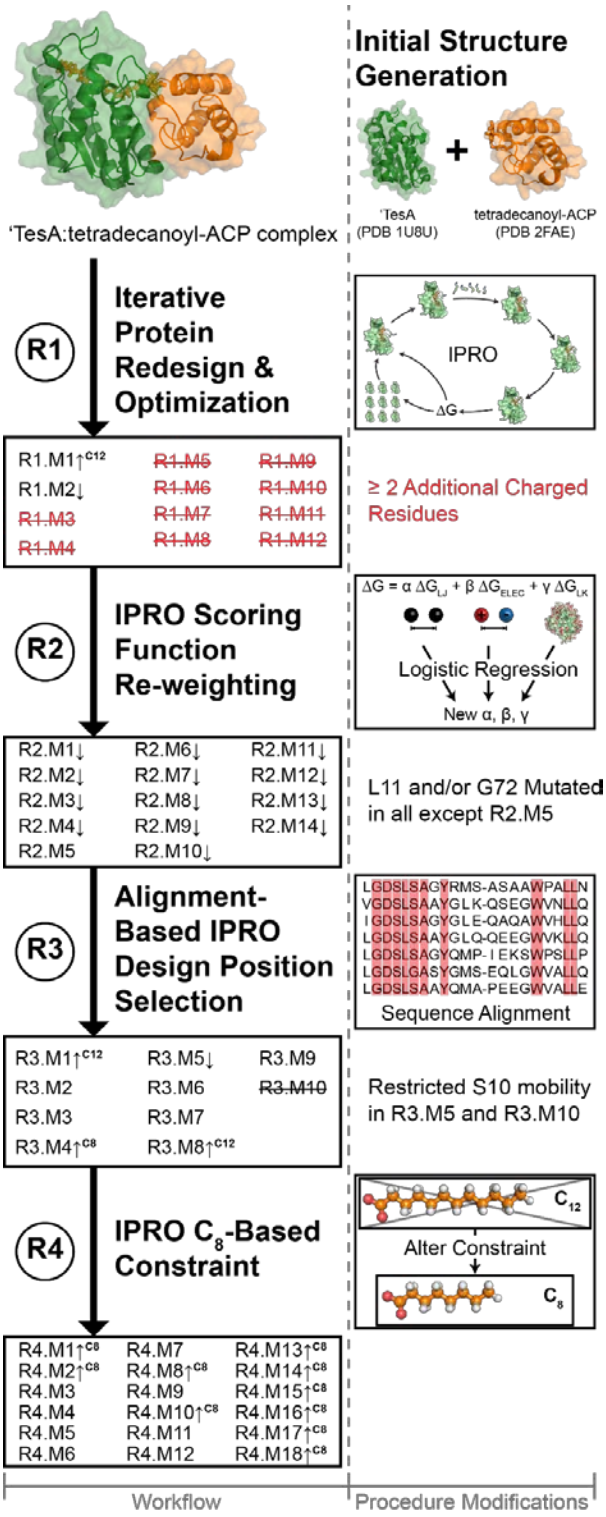


Figure 2.3. Overview of Predict-Design-Revise-Learn approach used to guide 'TesA redesign. Pictorial illustration of the steps traversed (*left column*) and mutants identified with improved specificity towards C_{12}- or C_8- acyl-ACPs. The right column denotes the changes in the computational procedure modifications in response to the

experimental results. Variants that produced significant improvements in the C_{12} ($p<0.05$) or C_8 ($p<0.005$) fraction while maintaining WT production levels are indicated with an upward arrow followed by the FFA. Major improvements in C_8 are indicated by $p<0.005$ rather than $p<0.05$ to highlight top designs Struck through variants indicate enzyme inactivity, and names followed by a downward arrow represent variants with reduced total production levels relative to WT.

2.3.2. **IPRO constraints and implementation**

IPRO enzyme redesign requires a three-dimensional model of the protein complex structure, a set of predetermined mutable residues known as design positions, and a set of constraints that quantify the desired improvements that protein variants need to reach relative to wild-type. Examples of such constraints include imposing relations that require a (i) stronger interaction energy with the new substrate and (ii) weaker interaction energy with the native substrate compared to wild-type. Here, we constructed a structural model of ‘TesA bound to an acyl-ACP with acyl chains ranging from C_8 to C_{14}. The model was assembled from published structures of ‘TesA (PDB 1U8U)[68] and decanoyl-ACP (PDB 2FAE)[97]. Docking between ‘TesA and octanoyl-ACP was modeled by systematically rotating the octanoyl-ACP structure about its phosphopantetheine linker until it aligned with bound octanoic acid in the ‘TesA crystal structure (PDB 1U8U, Materials and Methods). Other ‘TesA:acyl-ACP complexes were derived from the ‘TesA:octanoyl-ACP complex structure by adding atoms to the ω-1 carbon (i.e., the carbon furthest from ACP) of the octanoyl-ACP structure. The IPRO algorithm was used to search for mutations that led to improved binding between 'TesA and the desired substrate while simultaneously discouraging binding of

undesired substrate(s). Design positions for ‘TesA were selected based on proximity to the ω-1 carbon of bound tetradecanoyl-ACP while not considering positions vital for catalytic turnover (Materials and Methods). The fitness of each predicted variant was assessed using interaction energy as a proxy for binding energy (i.e., ΔG) thereby reducing force field dependence and requiring fewer calculations. Interaction energy is defined as $G_{Enz:FFA,min}$-G_{Enz}-G_{FFA}, where G is Gibbs free energy, “Enz” represents the enzyme (‘TesA or a variant thereof), and “min” indicates that the molecule(s) have undergone an energy minimization. In contrast, binding energy is defined as $G_{Enz:FFA,min}$-$G_{Enz,min}$-$G_{FFA,min}$. The first constraint is aimed at worsening the interaction energy between ‘TesA and the native substrate, tetradecanoyl-ACP (e.g., C_{14}). This first constraint safeguards against binding to even longer acyl-ACPs (e.g., C_{16}, C_{18}) because the repulsive interactions, which disfavor interactions with the long acyl-ACP (e.g., C_{14} and longer), are further exaggerated due to steric clashes or hydrophobicity. The second constraint requires interaction energy improvements between ‘TesA and the medium-chain acyl-ACP (e.g., C_{12}). The imposition of the second constraint attempts to enhance binding with C_{12} or even shorter acyl-ACPs (e.g., C_8, C_{10}).

IPRO operates by successively performing redesign iterations for a preset number of cycles until a variant that simultaneously satisfies all constraints and optimizes the objective function is found. IPRO has been applied previously to modify *E. coli* β-glucuronidase substrate specificity[115], alter the cofactor specificity of *Candida boidnii* xylose reductase[116], graft a calcium-binding pocket into *Thermoactinomyces vulgaris* thermitase or a copper-binding pocket into *E. coli*

thioredoxin[117], and *de novo* design antibody variable regions that target influenza hemagglutinin, HIV gp120 and Ebola GP1-GP2 viruses[90]. Each IPRO iteration begins with a local backbone perturbation nearby a randomly selected design position from the requisite input set. The second step implements a mixed integer linear program (MILP) to identify the optimal set of amino acids for the new backbone conformation. The third step of IPRO executes a local, rigid-body docking (i.e., intramolecular movements are prohibited and only relative positioning between the enzyme and ligand is considered) to reorient the ligand (i.e., the acyl-ACP) within the binding site. The fourth and fifth steps perform an energy minimization of the entire enzyme complex. The sixth step and final steps evaluate the constraints set forth (i.e., (i) reducing binding to C_{14} and (ii) increasing binding to C_{12}). The results of the iteration are retained or discarded based on the Metropolis criterion whereby worsening solutions are accepted with a decaying exponential probability as in simulated annealing (see Ref.[72] for further details). The two constraints are imposed only at the ground state of the thioester hydrolysis with no additional calculations at the transition state. Instead, simple restraints on catalytic distances were imposed during all IPRO iterations so as to preserve, but not necessarily boost, catalytic activity[50]. Multiple IPRO trajectories (~10 independent trajectories) were simulated to discover alternative routes for improving specificity. For each variant, the difference of interaction energies for the short and long acyl-ACP with ‘TesA (i.e., $\Delta IE = IE_{C12} – IE_{C14}$) was calculated. The variants were prioritized based on the extent of the energy differences, and 10-20 variants were selected to build a focused library for experimental testing.

2.3.3. **Method implementation for ‘TesA redesign**

The first round (R1) of IPRO-guided mutagenesis helped tune the parameters of the scoring function to improve prediction accuracy. Our initial objective was a modest shift in substrate preference from C_{14} to C_{12}, that makes up to 20% of the native 'TesA product profile, in order to assess the efficacy of the redesign protocol. Of twelve tested variants, one improved C_{12} composition) and maintained WT production levels (i.e., ≥ WT_{Total}, $p<0.05$; R1.M1; **Appendix III**). In contrast, ten were inactive (i.e., < $Control_{Total}$) and included more than two charged substitutions (all except R1.M1 and R1.M2). The scoring function implemented within IPRO’s MILP rotamer-residue selection algorithm was identified as the source of the charged residue bias. The scoring function energy terms were re-weighted using logistic regression on a dataset of high-quality protein structures[104]. The updated scoring function roughly doubled native rotamer recovery relative to the existing scoring function (Materials and Methods). The former scoring function was adequate for earlier systems[72,90,115–117], but the high hydrophobicity of ‘TesA (the energy term that was underemphasized in the former scoring function) made this enzyme especially susceptible to unsuccessful designs.

The modified scoring function was used to design a second round (R2) of variants with improved activity on C_{12} acyl chains. While all fourteen variants were active, thirteen mutants (all except R2.M5) produced less total FFA than WT and none improved the C_{12} fraction ($p<0.05$). All thirteen variants incorporated a mutation at either position L11 or G72. These positions were in hindsight deemed conserved (exhibiting 86.3% and 60.0% sequence conservation, respectively, across the L1-

like lysophospholipase subgroup of the SGNH-hydrolase family) as they are immediately adjacent to S10 and N73 which are part of the oxyanion hole[68]. Therefore, both L11 and G72 were eliminated from the list of design positions in subsequent rounds.

The third round (R3) of mutagenesis, based on the updated set of design positions and revised scoring function, resulted in a higher fraction of variants that maintained WT activity (80.0%) compared to that of R1 and R2 (8.3% and 7.1%, respectively). Of the ten tested R3 designs, two improved the C_{12} product composition (R3.M1, R3.M8; $p<0.05$) and one led to a major increase in the C_8 mole fraction (R3.M4; $p<0.005$). Notably, as the design constraints imposed by IPRO in rounds R1 through R3 did not preclude binding to acyl-ACPs smaller than C_{12} (i.e., C_6, C_8, C_{10}), it led to the serendipitous isolation of an octanoyl-ACP dominant variant (R3.M4). The discovery of a C_8 mutant R3.M4 and the higher commercial value of octanoic acid[2] prompted a final round of computational predictions focusing on C_8 composition only.

In the fourth round (R4), the IPRO imposed design constraints were similar to that of rounds R1 through R3 except that dodecanoyl-ACP (the “desired substrate”) was replaced with octanoyl-ACP, while tetradecanoyl-ACP (the “undesired substrate”) was replaced with dodecanoyl-ACP. These changes were made to drive more aggressive mutagenesis towards C_8 preference. Of the eighteen R4 variants tested, all were active, sixteen maintained WT FFA production levels (all except R4.M11 and R4.M12), thirteen improved the C_8 mole fraction, but none produced more octanoic acid than R3.M4. The reduced

production levels for R4.M11 and R4.M12 could be explained by the rearrangement of aromatic side chains in the binding crevice (i.e., R108F, F139, Y145, and Y145F) relative to the active R4 variants, which may have disrupted the hydrophobicity of the binding crevice.

Through four rounds of the Design-Build-Test-Learn paradigm, 54 variants were tested, 43 were active, 25 maintained WT production levels (23 in R3-R4), 3 improved the C_{12} mole fraction, and 27 raised C_8 composition (see **Appendices III** and **IV**). All three C_{12}-specific variants and all but six of the 27 C_8-specific variants (R1.M2, R2.M2, R2.M3, R2.M4, R3.M3, and R4.M12) maintained or exceeded WT production levels. The success rate of computational enzyme design varies drastically based on the procedures employed, the system studied and the ambition of the (re)design (published values can range from as low as 7% to as high as 78%), and the frequency of favorable outcomes from IPRO in this study (24% for R1-R4, 43% for R3-R4) falls in line with these previously reported values[95,96,118–121]. The variant with the highest C_{12} mole fraction (48±8%, a 1.8-fold improvement over WT), R3.M1, consisted of three mutations: S122K, Y145K, and L146K. Despite attempting to target octanoic acid production in R4, the variant with the highest C_8 mole fraction was R3.M4 (50±3%, a 10-fold improvement over WT), containing mutations M141L, Y145K, and L146K. The FFA profiles for the top C_{12}- and C_8-specific variants are summarized in **Figure 2.4**.

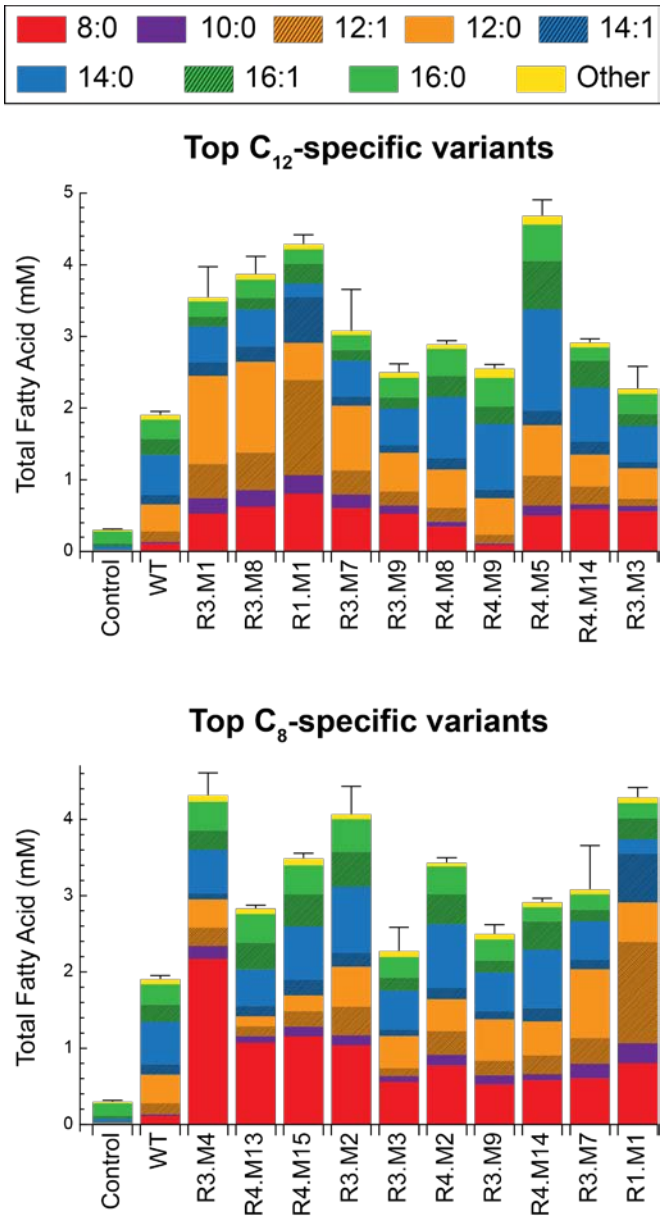


Figure 2.4. Fatty acid production profiles for the most C_{12}-specific and C_8-specific computationally designed variants. FFA titers are shown as bars, where error bars indicate total FFA standard deviation. Profiles of uninduced cells (Control) and wild-type 'TesA (WT) are provided for reference. All profiles are listed in Appendix III.

2.3.4. **Computation-guided design outperforms random mutagenesis**

Several of our best performing variants have a small number of mutations that may have been recovered from classical random mutagenesis approaches. Therefore, we created a small library of randomly mutated ‘TesA variants by error-prone PCR mutagenesis and screened for changes in product profile. The purpose of this library was to provide a negative control for testing that a library not directed by IPRO would not achieve the same level of success. The FFA profile of 61 *E. coli* cultures harboring expression vectors for unique ‘TesA variants was measured (N=1). Of the 61 random mutants (RMs) screened, 46 were active (i.e., ≥ 240 μM), and 20 maintained WT FFA production levels (i.e., ≥ 1750 μM). The best dodecanoic acid producing RM (RM.M39, 44±6%) demonstrated a comparable C_{12} composition to the top computationally predicted variant (R3.M1, 48±8%) but at the expense of a substantial reduction in total FFA titer ($p<0.05$, 80% of R3.M1). The best octanoic acid producing RM (RM.M29, 21.9%) produced a lower fraction of C_8 than R3.M4 (50±3%, 44% reduction) and displayed only 58% of the total activity of R3.M4 (see **Appendix V**). A comparison of these results shows that IPRO-guided mutagenesis generated more hits, more active mutants, and better leads than a library of similar size made through random mutagenesis.

2.3.5. **Analysis of successful ‘TesA redesigns**

The best C_{12} producing variants were dominated by three mutations: S122K, Y145K, and L146K. Mutation S122K (R3.RD3, **Appendix III**) alone was

sufficient to shift the C_{12} fraction to 35% of total FFAs, equal to the best C_{12}-producing mutant R3.M1. A non-polar mutation at the same position (S122L – R3.RD4) had a similar but less pronounced shift towards C_{12} at the expense of C_{14}, indicating that S122 is an important residue in the active site (see **Figure 2.5**). The additional mutations in R3.M1 (Y145K and L146K – equivalent to the R3.M7 mutation) also reduced the long-chain composition but produced a higher fraction of C_8 and lower total activity. Nearly all of the top C_8 producing variants contained a mutation at Y145 with lysine or phenylalanine as the dominant substituents. The best C_8 producing mutant (R3.M4) contained Y145K and L146K mutations as well as a M141L, which by itself (R3.M2) was able to dramatically increase the C_8 composition. The Y145K mutation drastically increased the fraction of unsaturated products in the C_{12} and C_{14} chain lengths.

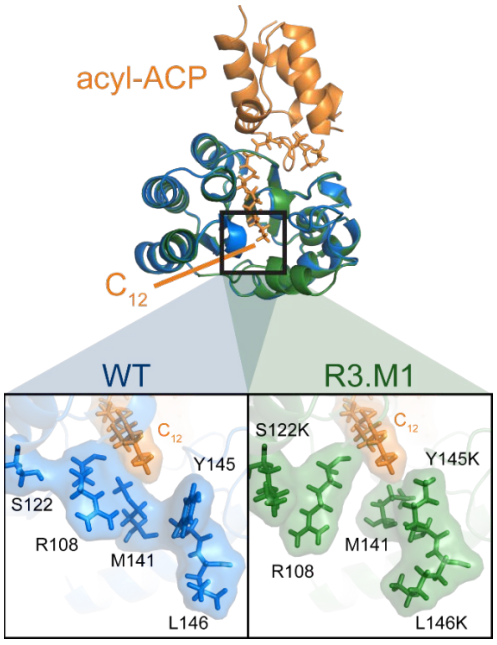


Figure 2.5. IPRO-derived structure of R3.M1 bound to dodecanoyl-ACP. The quaternary structure of the complex is shown at top with proteins shown as cartoons. Acyl-ACP is orange and the dodecanoyl moiety (C_{12}) is depicted as sticks. The inset

contrasts the binding sites of WT (blue) and R3.M1 (green). Mutated positions in R3.M1 (122, 145, and 146) are shown as sticks with a transparent surface. Two other nearby residues that show substantial movements are also formatted as sticks with transparent surfaces.

2.3.6. Crystal structures and simulations show that hydrophobic interactions govern specificity

Crystal structures of WT 'TesA and R3.M4 bound to octanoic acid were solved to confirm the structures predicted by IPRO and facilitate analysis of helpful mutations. The WT crystal structure was very similar to the previously published structure[68] with an all-atom root-mean-square deviation (RMSD) of 1.1 Å. The X-ray resolved structures corroborated the structures predicted by our computational methods, quantified by an all-atom RMSD of 1.6 Å for R3.M4:octanoyl-ACP (**Figure 2.6**). The structural differences in $loop_{111\text{-}120}$ may have affected IPRO's ability to accept or reject mutations. However, the strong structural similarity between the crystallized and modeled structure at positions M141L, Y145K and L146K indicate that the conformation of these side chains is favorable despite the movement of $loop_{111\text{-}120}$. Comparison of the WT and R3.M4 structures reveal that the largest differences (RMSD ≥ 2.5 Å) occur at (i) the mutated positions (M141L, Y145K, L146K), (ii) the flexible regions as suggested by MD-derived B factors (L11-A19, Q32-S33, G44-D45, N73, I107-R115 and D153-I156), (iii) the solvent-exposed residues adjacent to these flexible loops (i.e., H157 and R160), and (iv) the C-terminus (i.e., L177, **Figure 2.7**). Whereas structural differences in the flexible regions are possibly artifacts due to high residue mobility, the structural

differences at the mutated positions provide insight to enzyme specificity determinants. The M141L side chain extends laterally towards the ω-1 carbon of the FFA yielding an attractive dispersion force (**Figure 2.8-A**). Y145K partially occludes the binding crevice by forming a barrier between the ω-1 carbon of the FFA and solvent. L146K forms a salt bridge with E143 that may stabilize the conformation of Y145K but does not directly influence binding. Because mutant R3.M4's (and all other active variants') catalytic machinery includes an oxyanion hole that necessitates a negative charge for binding, R3.M4 in complex with octanoic acid (pK_a 4.9) was also crystallized at a higher pH to ensure binding site occupancy. The additional experiments included crystallization at pH 7.5 and crystallization at pH 5.0 followed by additional octanoic acid buffered at pH 7.5. These additional experiments revealed substantial differences (RMSD ≥ 2.5 Å) at R16 (2.534 Å), K34 (3.024 Å), A111 (4.362 Å), Y113 (9.311 Å), G114 (3.536 Å), and L177 (3.723 Å, **Figure 2.9**). These highly flexible regions indicate that their movement is highly sensitive to changes in hydrophobicity and their motion may be essential for catalytic turnover. The structure of N112 was not solved for R3.M4 at pH 5.0 because of low-resolution electron density maps.

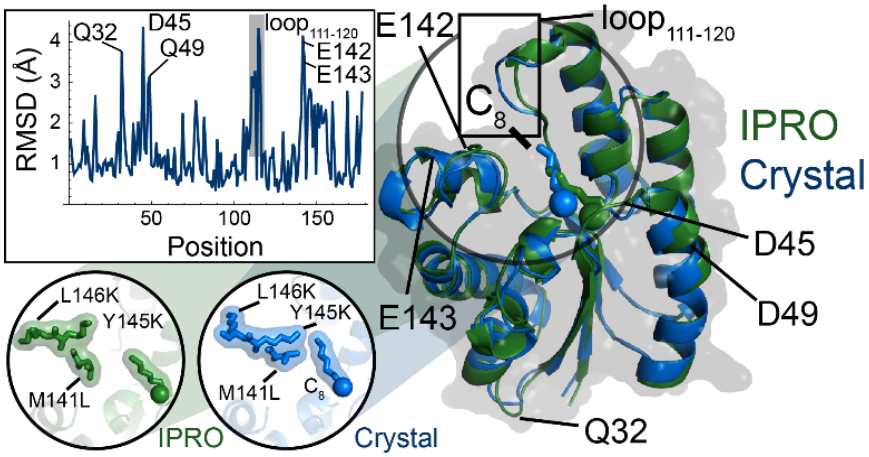


Figure 2.6. Comparison of computationally-predicted and crystallized structure of R3.M4. The IPRO-derived structure (green) and crystallized structure at pH 5.0 (blue) of R3.M4 are shown as cartoons bound to octanoyl-ACP (truncated at the thioester bond) and octanoic acid (C_8), respectively. The carbonyl oxygen of each FFA structure is shown as a sphere. Regions of relatively large structural differences (RMSD ≥ 3.0 Å) are shown in a plot of RMSD versus position and annotated. Design positions (141, 145, and 146) are shown in the bottom left inset, where each residue is shown by sticks and a transparent surface. Hydrogen atoms are excluded from the IPRO predicted structure.

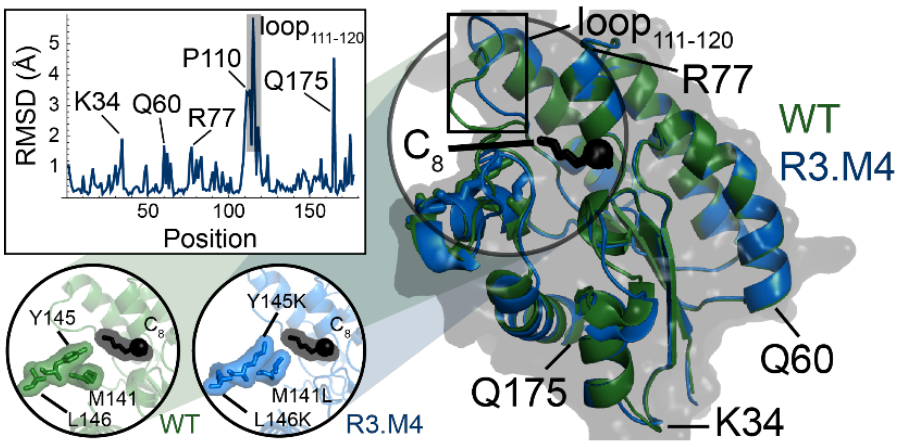


Figure 2.7. Comparison of WT and R3.M4 crystal structures. Both WT (green) and R3.M4 (blue, pH 5.0) are shown as cartoons to the right of the Figure bound to octanoic acid (C_8). The carbonyl oxygen of each FFA structure is shown as a sphere. Regions with relatively large changes (RMSD ≥ 1.5 Å) are annotated. Quantification of the structural differences is shown with the inset plot of RMSD versus sequence position. The structural differences between the residues at positions 141, 145, and 146 (the

mutated positions in R3.M4) are depicted in the bottom left inset. Here, each of the three residues, as well as octanoic acid, is represented as sticks with a transparent surface.

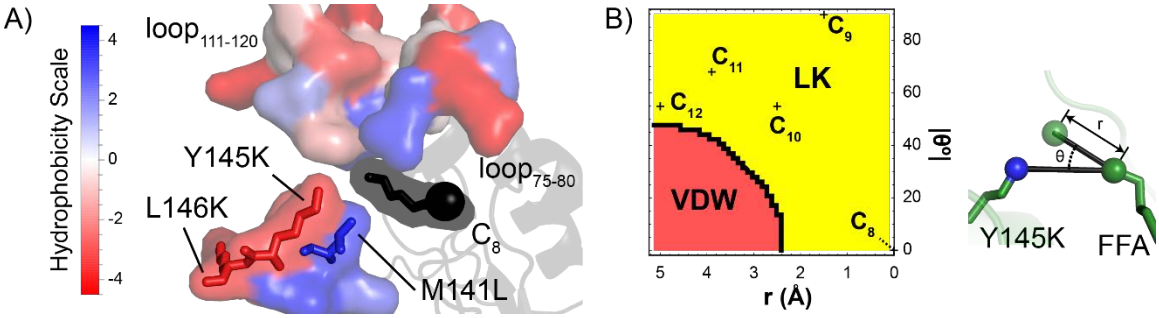


Figure 2.8. Hydrophobicity drives substrate specificity as demonstrated by (A) R3.M4 crystal structure and (B) a two-atom model for the Y145K mutation. In (A), crystallized R3.M4 (pH 5.0) is bound to octanoic acid (C_8). The two critical hydrophobic loops ($loop_{75-80}$ and $loop_{111-120}$) as well as the residues nearby the design positions are shown as molecular surfaces and colored according to their hydrophobicity score. The FFA is represented as sticks (with its carbonyl oxygen shown as a sphere) and a black molecular surface, as its hydrophobicity score is unknown. In (B), a two-atom model is shown to examine the ability of Y145K to eliminate binding to tetradecanoyl-ACP. The ε-amino nitrogen of Y145K (blue sphere) and the ω-1 carbon of FFA (green sphere) constitute the two-atom model. The model shows the effect of extending the acyl chain, which changes the hypothetical position of the ω-1 atom (transparent green sphere). r represents the interatomic distance between the C_8 ω-1 atom and the hypothetical ω-1 atom. θ represents the angle between the ε-amino nitrogen, the C_8 ω-1 atom and the hypothetical ω-1 atom. At (r=0 Å, θ=0°), the ω-1 atom occupies the position of the crystallized ω-1 C_8 atom. The energy term with the largest contribution towards the interaction energy between the ε-amino nitrogen and the hypothetical ω-1 atom at a given r and θ is provided in the contour plot (red=van der Waals energy, yellow=Lazaridis-Karplus solvation energy). The positions of the hypothetical ω-1 atoms using ideal FFA geometry (C-C bond length of 1.54 Å, 109.5° angle, 180° dihedral angle) are also labeled in the contour plot. C_{13} (r=6.4 Å, θ=62.8°) and C_{14} (r=7.5 Å, θ=54.9°) are beyond the boundaries of the contour plot. Molecular structures were generated using PyMOL.

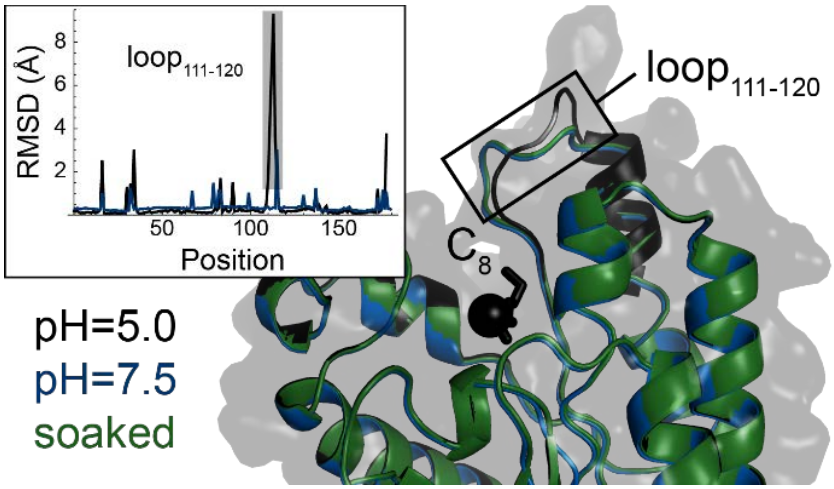


Figure 2.9. Crystallized R3.M4 at three different pH conditions. Crystallized R3.M4 under three different pH conditions is shown bound to octanoic acid (C_8). R3.M4 is shown as a cartoon and colored black (pH 5.0), blue (pH 7.5), or green (crystallized at pH 5.0 and buffered at pH 7.5). The carbonyl oxygen the FFA is shown as a sphere. The loop with the largest conformational change between the three structures is labeled ($loop_{111-120}$). The inset plots the RMSD of R3.M4 at pH 5.0 (black) and pH 7.5 (blue) to the soaked structure against the residue position number, quantifying the regions with the largest changes. Position 112 in the pH 5.0 structure is gray because it was not solved for the crystal and was instead modeled.

MD simulations were used to elucidate the mechanisms by which FFA composition is controlled. Seven total trajectories were analyzed using MD. These included WT (bound to C_{14}, C_{12}, or C_8 acyl-ACPs), R3.M1 (bound to C_{14} or C_{12}), and R3.M4 (bound to C_{12} or C_8). The MD results revealed an alternate binding mode nearby T46, S47, N73, and R77 that is only assumed for preferred chain lengths in the WT and R3.M4 trajectories. This binding mode is likely not observed for R3.M1:dodecanoyl-ACP because of insufficient conformational sampling. In addition to the alternate binding mode, MD analysis revealed that enzymes bound to acyl-ACPs beyond their preferred chain lengths have a deformed loop between G75 and Q80. With the exception of R3.M4:octanoyl-ACP and WT:octanoyl-ACP,

all other enzymes bound to an acyl-ACP equal to or just below (≤ 2 carbon atoms) its preferred chain length maintain a constant conformation of this loop (**Figure 2.10**). Finally, B factors for each of the seven simulated enzymes were estimated and exhibited six regions with increased mobility. These regions were found from L11-A19, Q32-S33, G44-D45, N73, I107-R115, and D153-I156 ($B_i \geq \overline{B_i} + \sigma$).

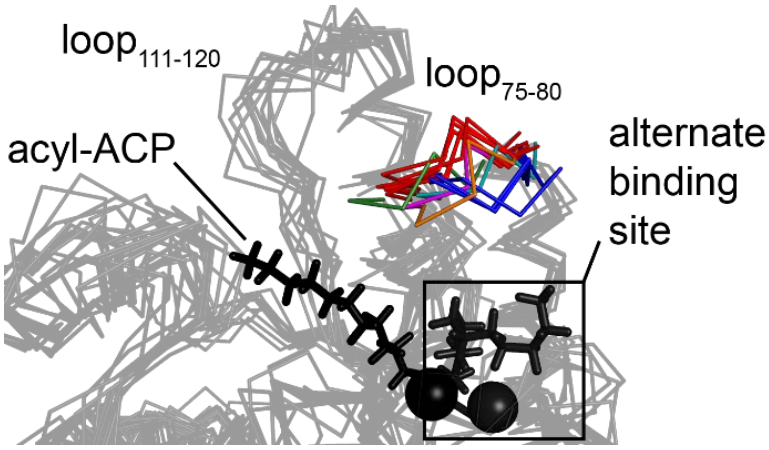


Figure 2.10. Structures of bound R3.M1, R3.M4, and WT enzymes from MD simulations. Each of the variants was shown as a black ribbon. The critical switch loop found between positions 75 and 80 was colored based on clustering of different conformations. Crystallized WT:C_8 published by Lo *et al.*, crystallized WT:C_8 from this study, crystallized R3.M4:C_8 (pH 5.0) from this study, WT:C_{12} from MD, WT:C_{14} from MD, and R3.M1:C_{12} from MD assume a similar conformation and are colored red. Unbound 'TesA, an unbound L109P variant (that abolishes switch loop movement), and a bound L109P variant from previous crystallography experiments assume a similar conformation and are colored blue (from Lo *et. al*). R3.M4:C_8 from MD, R3.M4:C_{12} from MD, R3.M1:C_{14}, and WT:C_8 are colored green, cyan, orange, and pink, respectively. The carbonyl oxygen atoms of the acyl-ACP structures are depicted as spheres.

Combining the MD and crystallography results with previously published investigations of 'TesA acyl-ACP specificity[68] support the theory that hydrophobic interactions formed between loop$_{75\text{-}80}$ (i.e., residues forming the loop between G75

and Q80), the acyl chain of acyl-ACP, and loop$_{111\text{-}120}$ govern substrate selectivity. Lo *et al*.[68] previously postulated that loop$_{75\text{-}80}$, described therein as the "switch loop" that is in one conformation (i.e., "on") when the acyl-ACP is bound and in another conformation (i.e., "off") when the substrate is not bound. The switch loop is dependent on the acyl chain length and stabilizes the enzyme:substrate complex during hydrolysis. Our MD simulations are consistent with these findings, as the switch loop is in the "on" position for variants bound to an acyl-ACP equal to or just less than (≤ 2 carbon atoms; i.e., WT bound to C_{12}) its preferred chain length. Prior studies[68,122] revealed that rigidity of the switch loop or the lack of a bound acyl-ACP force the switch loop into the "off" conformation. MD simulations for WT:octanoyl-ACP, R3.M1:tetradecanoyl-ACP, and R3.M4:dodecanoyl-ACP occupy intermediate states between the "on" and "off" conformations. The switch loop is in a completely different conformation for R3.M4:octanoyl-ACP, which is due to the increased hydrophobicity in the binding crevice towards the C-1 terminus caused by M141L. For R3.M4:octanoyl-ACP, the movement of the switch loop towards loop$_{111\text{-}120}$ allows octanoyl-ACP to slide underneath the switch loop into the alternate binding cavity, which could be important for octanoic acid release. The pH-dependent conformation of loop$_{111\text{-}120}$ for R3.M4 from the crystallography experiments suggests that a change in protonation state of a titratable residue (i.e., an acidic or basic amino acid) causes a major conformational change, demonstrating the sensitivity of the hydrophobic region to electrostatics. At pH 7.5, the hydrophobic loop$_{111\text{-}120}$ extends away from the typically hydrophobic binding crevice towards the highly hydrophilic bulk solvent (**Figure 2.9**), which is

unexpected since nonpolar substances typically aggregate (i.e., the hydrophobic effect). This movement of $loop_{111\text{-}120}$ is postulated to be due to a newly charged residue in the binding crevice, weakening the binding crevice's hydrophobic environment. This observation is consistent with octanoic acid serving as the titratable residue. While M141L increases the hydrophobicity near the C-8 of R3.M4, Y145K limits hydrolysis of longer acyl-ACPs by reducing binding crevice hydrophobicity near the ω-1 atom (i.e., C-10, C-12, C-14) thereby disrupting the conformation of the switch loop. A simple two-atom model involving the FFA ω-1 carbon and the ε-amino nitrogen (i.e., the side chain nitrogen) atom had pairwise energies calculated and demonstrates that the Lazaridis-Karplus solvation term, a computationally accessible proxy for hydrophobicity, governs unfavorable interactions with the ω-1 carbon (**Figure 2.8-B**). The two-atom model illustrates that solvation energy dominates the phase space with the exception of the van der Waals region. The ω-1 carbon would not be expected to occupy the van der Waals region because the energy is much more unfavorable (i.e., positive) at these close distances ($E \sim r^{-12}$ for van der Waals, $E \sim \exp(-r^2)$ for Lazaridis-Karplus solvation). Longer FFA chain lengths extend closer to the ε-amino nitrogen, exacerbating the repulsive solvation energy. Adding additional carbon atoms to the point where the ω-1 is further from the ε-amino nitrogen will only add to the total repulsive energy, although the marginal cost will decrease with each additional carbon. Finally, the large degree of switch loop and $loop_{111\text{-}120}$ mobility could explain the broad substrate specificity that is typically observed for 'TesA.

$\Delta\Delta G$ ($\Delta G_{Variant}$-ΔG_{WT}) values estimated from the computations exhibited good agreement with experimentally derived values (Supporting Information). A Pearson correlation coefficient of 0.6 ± 0.2 was found for R3 and 0.43 ± 0.07 for R4. These correlation coefficients are in line with earlier computational studies[72,123,124]. Notably, the regions with elevated B factors from the MD trajectories ($loop_{11-19}$, $loop_{32-33}$, $loop_{44-45}$, $loop_{73}$, $loop_{107-115}$, and $loop_{153-156}$) overlap with the regions with elevated B factors from the crystallography experiments ($loop_{30-35}$, $loop_{59-62}$, $loop_{98-100}$, $loop_{111-115}$, and $loop_{153}$) and other published[68,122] crystallographic structures ($loop_{31-35}$, $loop_{75-80}$, and $loop_{111-120}$).

2.3.7. In vitro assays of 'TesA WT and R3.M4 C_8-specific confirms in vivo results.

'TesA can catalyze hydrolysis of both acyl-CoA and acyl-ACP substrates. Given the relative availability of these substrates, we compared the kinetic activity of WT 'TesA and R3.M4 on acyl-CoAs ranging from 6-16 carbons in length. We monitored reaction progress by tracking the abundance of free CoA released by hydrolysis. The highest *in vitro* 'TesA WT activity (**Figure 2.11-A**) was observed for C_{12}-CoA, C_{14}-CoA and C_{16}-CoA, consistent with the *in vivo* data for release of FFAs from acyl-ACPs (**Figure 2.4**). In contrast, R3.M4 showed a significant increase in activity on C_8-CoA compared to WT, consistent with the observed *in vivo* production of octanoic acid, and a modest decrease in activity on C_{12}-CoA and C_{14}-CoA activity. For both enzymes, we observed an unexpected drop in activity on hexadecanoyl-CoA (C_{16}-CoA) beyond a threshold concentration.

Interestingly, in the R3.M4 mutant the inhibitory effect of C_{16}-CoA is exacerbated and C_{14}-CoA also shows inhibition (not seen on ‘TesA WT). Given the linear reaction progress curves we observed, we suspected that the enzymes were substrate inhibited. Therefore, we performed assays with both C_8-CoA and the inhibitory CoA species (**Figure 2.11-B-E**). Competitive binding assays were performed at a constant concentration of C_8-CoA (50 µM) and variable concentrations of C_{14}-CoA (**Figure 2.11-B** for WT and **2.11-C** for R3.M4) and C_{16}-CoA (**Figure 2.11-D** for WT and **2.11-E** for R3.M4). In all cases, production of free CoA was inhibited by C_{14}-CoA and C_{16}-CoA in a concentration dependent manner consistent with the original assay in **Figure 2.11-A**. A possibility for this result is that the increased affinity for the C8-CoA molecule resulted also in increased binding affinity for C14 and C16 and that the strength of this binding is prevents it from reaching transition state.

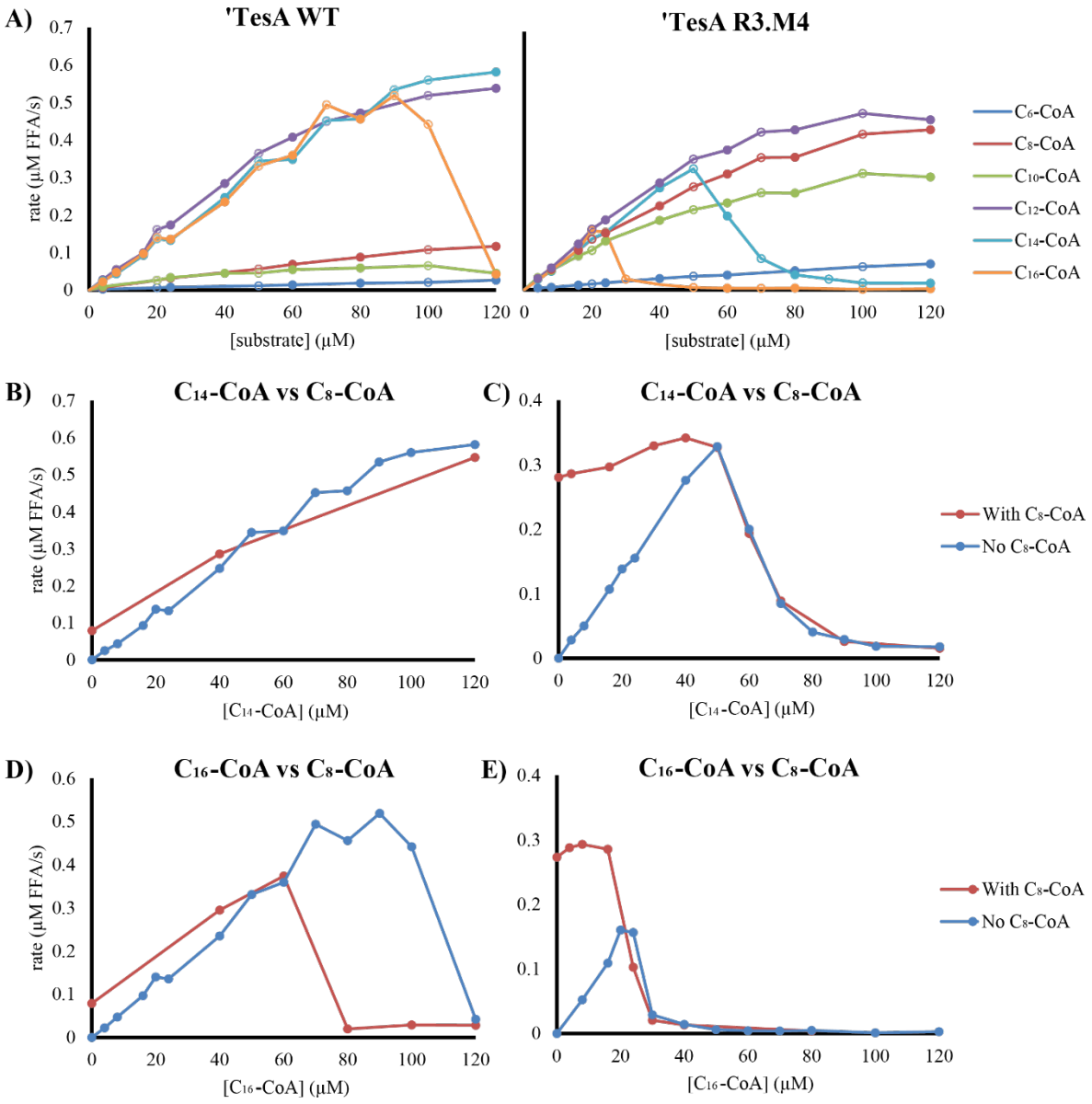


Figure 2.11. Enzymatic assays of 'TesA WT and R3.M4 (A) confirm the mutant's increase in specificity for C_8 species, and competitive binding assays (B-E) show 'TesA WT and R3.M4 activity on C_8-CoA as a function of C_{14}-CoA (B and C) and C_{16}-CoA (D and E) concentration. In (A), the activity of 'TesA WT (left) and R3.M4 (right) as a function of substrate concentration for six different acyl-CoA substrates. Open and closed circles indicate measurements taken in separate days. Competitive binding assays (B-E) were done to see the effect that an increase in C_{14}-CoA and C_{16}-CoA would have on the activity of the enzymes on C_8-CoA. Competitive binding assays were done at a constant C_8-CoA concentration of 50 μM and variable concentrations of C_{14}-CoA ((B) for WT and (C) for R3.M4) and C_{16}-CoA ((D) for WT and (E) for R3.M4). In all cases the activity of 'TesA was impacted by the C_{14}-CoA and C_{16}-CoA in a concentration dependent manner consistent with the original assay in (A).

2.4. Conclusions

The potential of the IPRO algorithm to aid in protein engineering efforts was demonstrated using a Design-Build-Test-Learn approach to alter the substrate preference of ‘TesA. Our approach leverages computational protein design procedures to achieve successful experimental redesign beyond what has been achievable so far[63,65], yielding two top variants. One, R3.M1, produced 48±8% C_{12} composition, a 1.8-fold improvement over WT, while maintaining native production levels. Despite a preference for C_{14} production in WT ‘TesA, three amino acid substitutions constitute R3.M1, which is the third most C_{12}-specific thioesterase known to date (see **Table 2.1**). Similarly, R3.M4 produced 50±3% C_8 composition, a 10-fold improvement, while maintaining WT production levels. R3.M4 is the tenth most C_8-specific thioesterase described to date. In total, the Design-Build-Test-Learn paradigm yielded three and twenty-one variants with significant ($p<0.05$) improvements in C_{12} and C_8 mole fraction, respectively, while maintaining WT production levels.

Table 2.1. Top ten C_{12}-specific and C_8-specific thioesterases published to date. Data was collected from Appendix III and sorted by dodecanoic acid (either saturated or unsaturated) and octanoic acid mole fractions in the product pool. WT compositions were provided as a reference. “Composition” refers to mole fraction of the corresponding FFA. Names match those provided in Appendix III. The most C_{12}-specific and C_8-specific ‘TesA variants found in R1-R4 correspond to the third most C_{12}-specific and tenth most C_8-specific thioesterases identified to date respectively.

	Dodecanoic Acid (C_{12})			Octanoic Acid (C_8)		
Rank	**Name**	**Composition**	**Source**	**Name**	**Composition**	**Source**
-	WT	28±1	This study	WT	5.9±0.3	This study
1	Q41635	76.9[a,b]	[13]	AAC49179	97.5±0.2	[10]
2	rTE15	49.0[a]	[16]	AAC72882	93.5[a]	[16]

3	R3.M1	49±8	This study	EEI82564	87±2	[10]
4	R3.RD3	48±10	This study	ABG82470	70±4	[10]
5	R3.M8	46±4	This study	CAD63310	68.0±0.8	[10]
6	RM.M39	44±6	This study	CvB2MT40	61.4[a]	[16]
7	R1.M1	43±2	This study	ABJ63754	55.5±0.7	[10]
8	AEM72521	40±2	[10]	AEM72522	52±6	[10]
9	R3.M7	40±11	This study	TEGm258	50.4[a]	[16]
10	R4.RD11	37.7±0.6	This study	R3.M4	50±3	This study

[a] Standard deviation not provided

[b] Data approximated using image processing software

Despite sampling a similar library size, random mutagenesis yielded fewer active mutants than the computationally guided library. In addition, the random library produced two fewer C_{12}-specific variants with native production levels and twenty fewer C_8-specific variants with at least WT productivity. Unlike random mutagenesis, variants from R3 and R4 achieved high levels of activity by directly enforcing catalytic contacts through imposed restraints and preserving conserved amino acids. Furthermore, R3 and R4 variants systematically tailored the binding crevice environment to generate hydrophobic clusters between $loop_{111\text{-}120}$, the acyl-ACP, and the switch loop ($loop_{75\text{-}80}$). MD and crystallography results from this work and elsewhere[68,122] suggest that this hydrophobic packing is essential for enzyme functionality. Whereas the *in silico* method directly accounts for hydrophobicity (albeit only approximately) through the Lazaridis-Karplus solvation

energy term, random mutagenesis techniques can only improve binding site hydrophobicity by chance alone.

The results from this work not only establish the potential of computational methods in enzyme redesign, but the lessons learned from earlier rounds of design (i.e., R1 and R2) may inform redesign work with other systems. Computational procedures are especially valuable when high throughput screening is impractical. We found that the number of inactive designs was reduced by applying a modified scoring function that alleviated bias towards senseless mutations and avoiding design positions that are highly conserved in family sequence alignments. These initial rounds thus established the essentiality of working with a correctly calibrated scoring function and carefully selecting design positions with the aid of sequence alignments. The integrated deployment of computations with experiments in a sequential manner allows for the “early on” identification of deficiencies in molecular modeling and erroneously targeted design positions, providing a tractable workflow for engineering enzymes for higher specificity and activity.

2.5. Acknowledgements

Support for this work was provided by the National Science Foundation Award CBET-0967062, CBET-1149678, the Advanced Research Projects Agency-Energy Award DE-AR0000431, and the National Institutes of Health (NIH GM115921). Nestor Hernandez-Lozada is the recipient of a NIH Chemistry-

Biology Interface Training Program fellowship (T32 GM008505) and a Graduate Engineering Research Scholars fellowship from the UW-Madison College of Engineering.

3. Highly active C8-acyl-ACP thioesterase variant isolated by a synthetic selection strategy[3]

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

[3] This Chapter was originally published as:

Hernandez-Lozada, N. J.*, Lai, R.* *et al.* Highly active C8-acyl-ACP thioesterase variant isolated by a synthetic selection strategy. *ACS Synth. Biol.* 2205–2215 (2018). doi:10.1021/acssynbio.8b00215

*These authors contributed equally to this work.

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.

3.1. Introduction

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)[3,125]. 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 1.1**). 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[29] and *Cuphea* species often contain high percentages of C_8 fatty acids[126], 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[10]. In this approach, metabolic engineering principles are used to redirect carbon flux through fatty acid biosynthesis to specific products[2,9] 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[26] 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[127,128,129,130,131], but complete selectivity and high activity remains an unmet challenge. In contrast, many plant thioesterases act on a narrow set of substrates[30]. Plants synthesize fatty acids in the chloroplast and lipids in the cytosol[25]. 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[2]. 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.

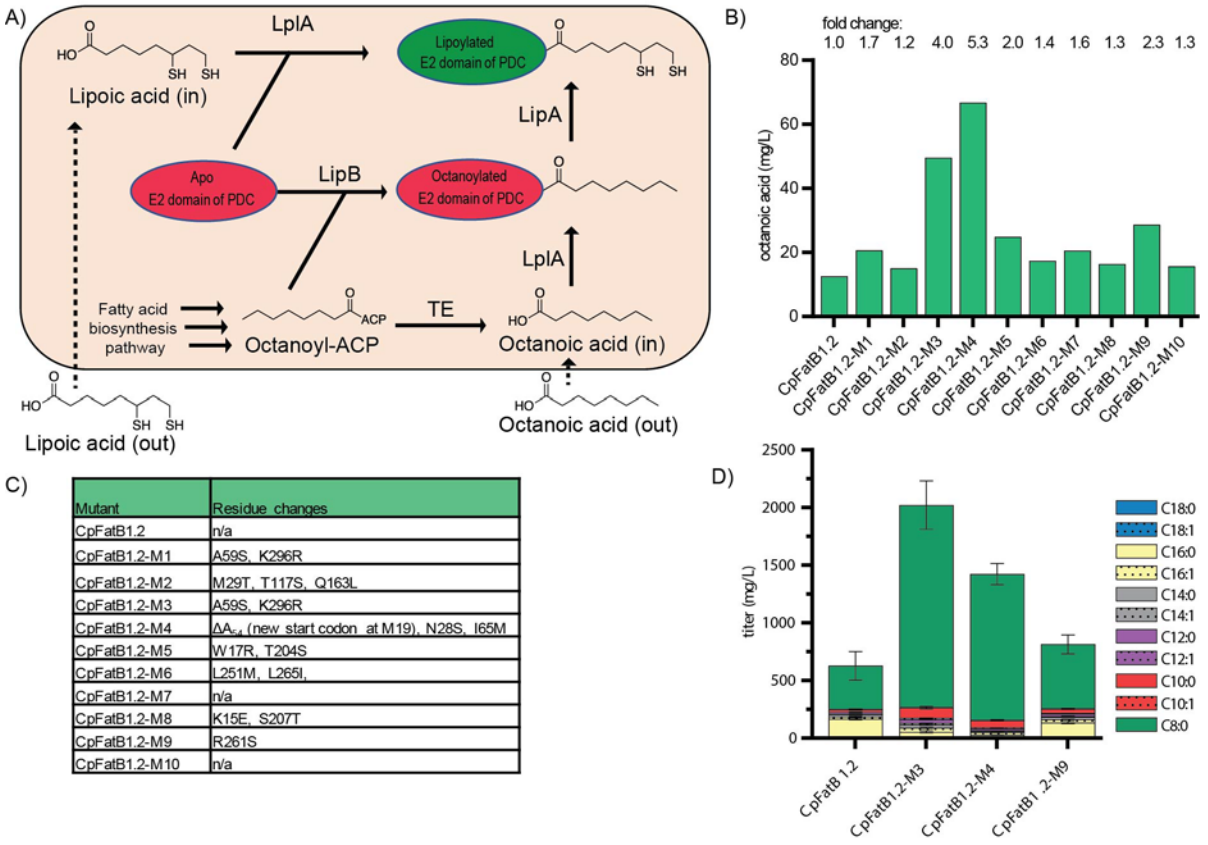
One challenge to thioesterase engineering is the lack of good screening methods to differentiate products with different chain-lengths. The analysis of fatty acid chain-length typically uses gas chromatography to separate fatty acid methyl esters (FAME) derived from biological samples. While accurate, 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[127] and domain swapping[130,131] 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[27] and G-protein coupled receptors[132,133]. 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[134]. 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 3.1-A**) 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 3.1-A**, growth of *E. coli* Δ*lipB* mutants can be rescued by supplying lipoic acid as well as octanoic acid in the media[135]. 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. Furthermore, other chain-length carboxylic acids were not able to rescue growth of the mutant. 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 Chapter, 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[136]. 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.



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_{2-4} (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 3.1. Overview of the selection platform and first round of screening. 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 3.6-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[36] 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.

3.2. Materials and Methods

3.2.1. 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).

3.2.2. 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.,[137] to delete the *lipB* protein coding sequence. Standard lambda red recombination[138] 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* (**Appendix X**).

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 *Ptrc-CpFatB1.2-M4-287* between 500 base pairs of homology upstream and downstream of *fadD* protein coding sequence (**Appendix X**).

All plasmids made were constructed using Gibson Assembly of PCR products[139]. See **Appendix I-II** for a list of strains and plasmids used in this study.

3.2.3. **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.

3.2.4. **Lipoic acid selection**

In order to find suitable conditions for the Δ*lipB*-based selection method, purified plasmid pBTRK- 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.

3.2.5. **Fatty acid production**

Plasmid based expression of thioestereases was performed in *E. coli* RL08ara[140] 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 3.1-B**, **Figure 3.5-A**, **Appendix IX**) as well as the truncations of CpFatB1 (**Figure 3.2**) and truncations in CpFatB1.2-M4 (**Figure 3.6-B**), 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 3.1-B**). Mutants M3, M4, and M11-M90 (**Figure 3.5**), truncations of CpFatB1 (**Figure 3.2**) and truncations in CpFatB1.2-M4 (**Figure 3.6-B**) were induced with 1 mM IPTG.

For experiments designed to test for high octanoic acid production (**Figures 3.1-D, 3.6-C-D, 3.9-A-B)** overnight cultures as described above were used to inoculate 50 mL of medium described in Kim, et al.[36] 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[141] containing 1% glucose and 0.240 mM K_2HPO_4 in order to create phosphate limiting conditions[142].

3.2.6. 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[127].

3.2.7. 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 x 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.

3.2.8. Protein expression and purification of Vibrio harveyi AasS and Bacillus subtilis SfP, CpFatB1.2 and CpFatB1.2-M4-287

Vibrio Harveyi AasS, *Bacillus subtilis* SfP, CpFatB1.2 and CpFatB1.2-M4-287 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^{-}1cm-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-M4-287.

3.2.9. Protein expression, purification and ESI Mass analysis of CpFatB1.2-M4-CHis

RL08ara containing pTRC99a-CpFatB1.2-M4-CHis were grown in LB+0.4% glycerol with carbenicillin as the antibiotic. The culture was incubated at 37 °C with 250 rpm shaking. When OD600 of the culture reached 0.2, IPTG was added to make the final concentration as 1 mM to induce CupTE2-M4-C-His overexpression. Then the culture was incubated at 30 °C with 250 rpm shaking for 24 hours. The cells were harvested by centrifugation. Finally, the protein was purified by Ni-NTA column following the

manufacture's instruction (GE Healthcare Life Sciences). The fractions containing CupTE2-M4-C-His were combined and concentrated. The protein solution was desalted using PD-10 column (GE Healthcare Life Sciences) equilibrated with desalting buffer (100 mM Tris, 100 mM NaCl, pH 7.5). Finally, the solution was frozen in liquid nitrogen and stored at -80 °C for future studies. Samples were analyzed by ESI Mass analysis at the UW-Madison Biotechnology Center (**Figure 3.7**).

3.2.10. **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[143]. 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 sample purity was confirmed via HPLC (see **Figure 3.8** and section 3.2.11) and its concentration was quantified using BCA assay.

3.2.11. 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 X 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.

3.2.12. In-vitro analysis of CpFatB1.2-M4-287

Octanoyl-ACP thioesterase activity of CpFatB1.2, CpFatB1.2-M4-287 and TesA-R3.M4[127] 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 triplicates.

3.2.13. Structural Modeling of CpFatB1.2-M3 and CpFatB1.2-M4-287

The CpFatB1.2 model was created using homology modeling of the CpFatB1.2 sequence and BTE (PDB:5x04 structure) as the template structure. Subsequently, the amino acid changes to create energy-minimized structures of the CpFatB1.2 mutants were made using Mutator[72]. The catalytic residues Asp220, Asn222, His224, Glu258, and Cys259 were identified using the *Umbellularia californica* thioesterase UaFatB1 (BTE) structure (PDB:5x04) as a guide where the analogous residues have been reported to be Asp281, Asn283, His285, Glu319, and Cys320[129]. 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[129]. 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.*[144]. 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 CpFatB1.2 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.

3.3. Results and Discussion

3.3.1. Establishment of a baseline thioesterase

Cuphea palustris FatB1 thioesterase (CpFatB1) is highly selective for C8:0-ACP when expressed in *E. coli*[30] albeit with lower activity relative to *E. coli* 'TesA and other commonly used thioesterases[9,26,11,54,67,42,145]. 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[31] and sequence alignment-CpFatB1.2, CpFatB1.3, and CpFatB1.4. Each was cloned into a high copy plasmid, pTRC99a (**Figure 3.2, Appendices I & II**) 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.

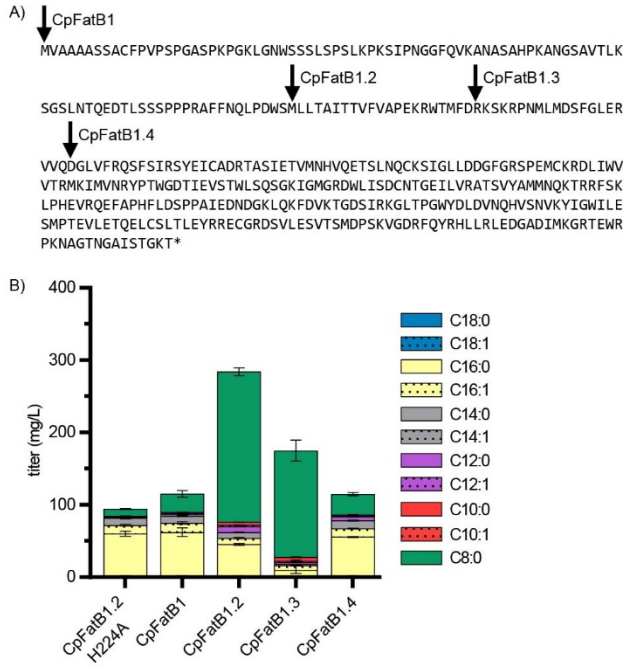


Figure 3.2. Establishing a baseline thioesterase via truncations. A) CpFatB1 truncation sites with arrows indicating the position M1 for each new truncation. B) Each truncation was cloned into pTRC99a plasmid Isolated plasmids were transformed into RL08ara (Δ*fadD*) strain and grown in LB 0.4% glycerol and 1 mM IPTG. CpFatB1.2 was chosen as baseline thioesterase for mutagenesis and selection in Δ*lipB strain.*

3.3.2. **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[146]. 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[141] containing 0.2% arabinose to induce expression (**Figure 3.3**).

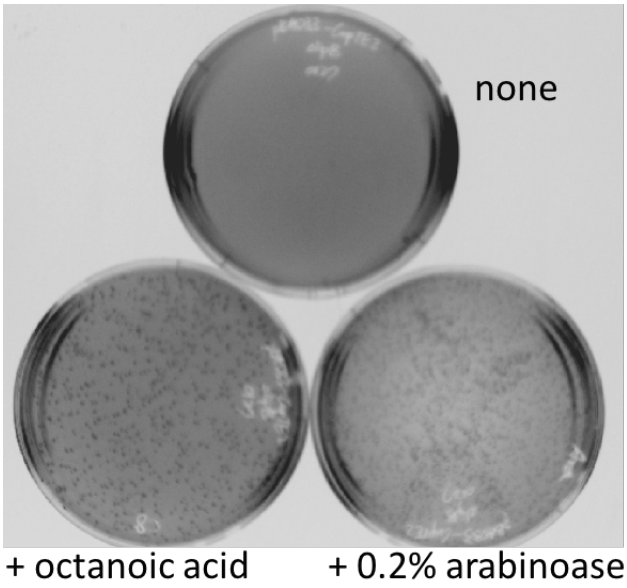
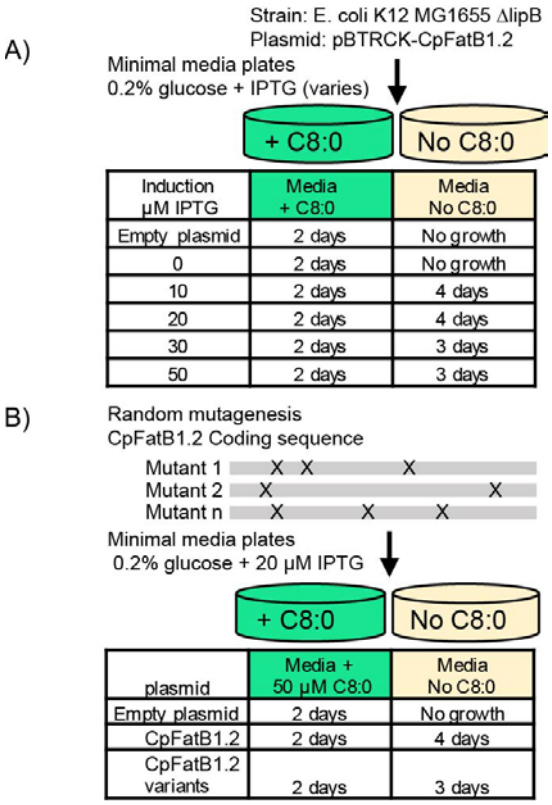


Figure 3.3. Rescue of the *E. coli lipB* mutant via CpFatB1.2 thioesterase.*E. coli lipB* null mutant containing pBAD33-CpFatB1.2 can grow on MOPS minimal media-agarose containing 0.2% arabinose compared without the addition of the pBAD33-CpFatB1.2. The positive control was the plate with the addition of 50 μM octanoic acid. This shows as a proof of concept that it is possible to use a thioesterase as an alternative to LipB in minimal media.

In a second attempt to reduce activity, we cloned the original pTrc-CpFatB1.2 expression cassette onto a low-copy plasmid, pBTRK (pBBR1 origin), and used a series of low IPTG concentrations to vary expression (**Figure 3.4A**). 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 3.4A**). 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 pBTRK plasmid and 20 μM IPTG in our mutagenesis study.



Induction μM IPTG	Media + C8:0	Media No C8:0
Empty plasmid	2 days	No growth
0	2 days	No growth
10	2 days	4 days
20	2 days	4 days
30	2 days	3 days
50	2 days	3 days

plasmid	Media + 50 μM C8:0	Media No C8:0
Empty plasmid	2 days	No growth
CpFatB1.2	2 days	4 days
CpFatB1.2 variants	2 days	3 days

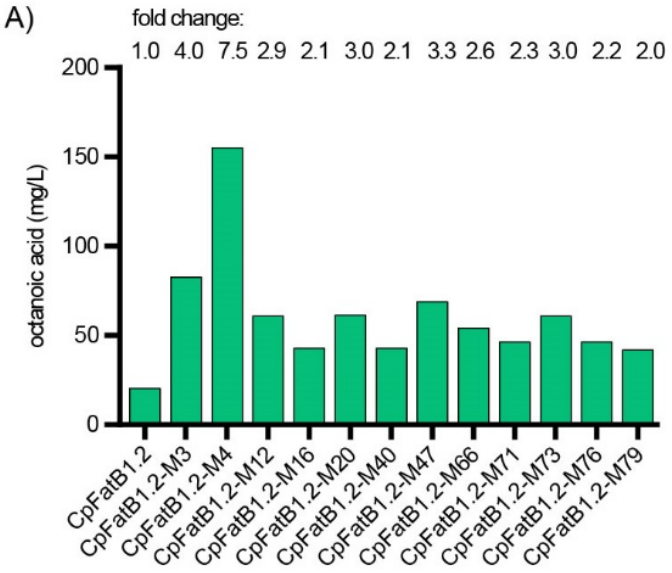
Figure 3.4. Selection of optimal conditions for lipB screen in CupTE.A) A low copy plasmid, pBTRK-CpFatB1.2 with CpFatB1.2 under an IPTG inducible promoter was transformed into *E. coli* ΔlipB strain and plated into MOPS minimal media plates containing different concentrations of IPTG. Growth in these plates was monitored for 4 days. Plates with 20 μM IPTG was selected for selection of mutants because *E. coli* ΔlipB carrying CpFatB1.2 took an extra day to grow (4 days), giving room for improved mutants to be selected the day before (3 days) B) Screening for improved mutants of CpFatB1.2. A library of pBTRK-CpFatB1.2 mutants was created by random mutagenesis of the coding sequence CpFatB1.2 at a low mutation rate. Next, this library was transformed into *E. coli* ΔlipB strain and plated into MOPS minimal media plates

containing 20 µM IPTG. 90 putative mutant colonies were selected after 3 days for further characterization.

3.3.3. **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 pBTRK, plasmids were transformed into *E. coli* Δ*lipB*, 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 3.4B**). To validate each hit, plasmids were isolated from each colony, retransformed into fresh *E. coli* Δ*lipB* 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 3.1B**). 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 3.5**, **Appendix IX**).

A)



B)

mutant	Residue changes
CpFatB1.2	n/a
CpFatB1.2-M3	A59S, K296R
CpFatB1.2-M4	ΔA_{54} (new start codon at M19), N28S, I65M
CpFatB1.2-M12	M136V
CpFatB1.2-M16	K296R
CpFatB1.2-M20	W17STOP, D293V
CpFatB1.2-M40	D293V
CpFatB1.2-M47	N146K, D293V, N309D
CpFatB1.2-M66	D293V
CpFatB1.2-M71	M136I
CpFatB1.2-M73	V268I, R279H, D293V
CpFatB1.2-M76	K23E, L86Q
CpFatB1.2-M79	V9M, K15E, E236A

Figure 3.5. Summary of highest performing mutants from the 90 selected colonies (see Fig 1B and Appendix IX). Interestingly, a mutation D293V appeared independently in five of these mutants, M20, M40, M47, M66, and M73.

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[36] 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 3.1D**). 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.

3.3.4. **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 3.6A, Appendix X**). 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 pBTRK vector to position the new start adjacent to the original RBS. Plasmids harboring each variant were individually transformed into RL08ara (Δ*fadD*) 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 3.6B**). 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 Ni-NTA 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 3.7**). 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 3.6-B, Appendix IX**), 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 (Δ*fadD*), and cultured cells in MOPS media enriched[36] with tryptone and yeast extract (see methods) and different concentrations of IPTG to optimize expression (**Figure 3.6C**). 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 3.6C**). 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[11] or induction[147]. Impaired growth was connected to reduced membrane fitness and increased cell permeability caused by the production of dodecanoic acid[148].

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 3.6C**. 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 3.6D**). 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 3.6B**).

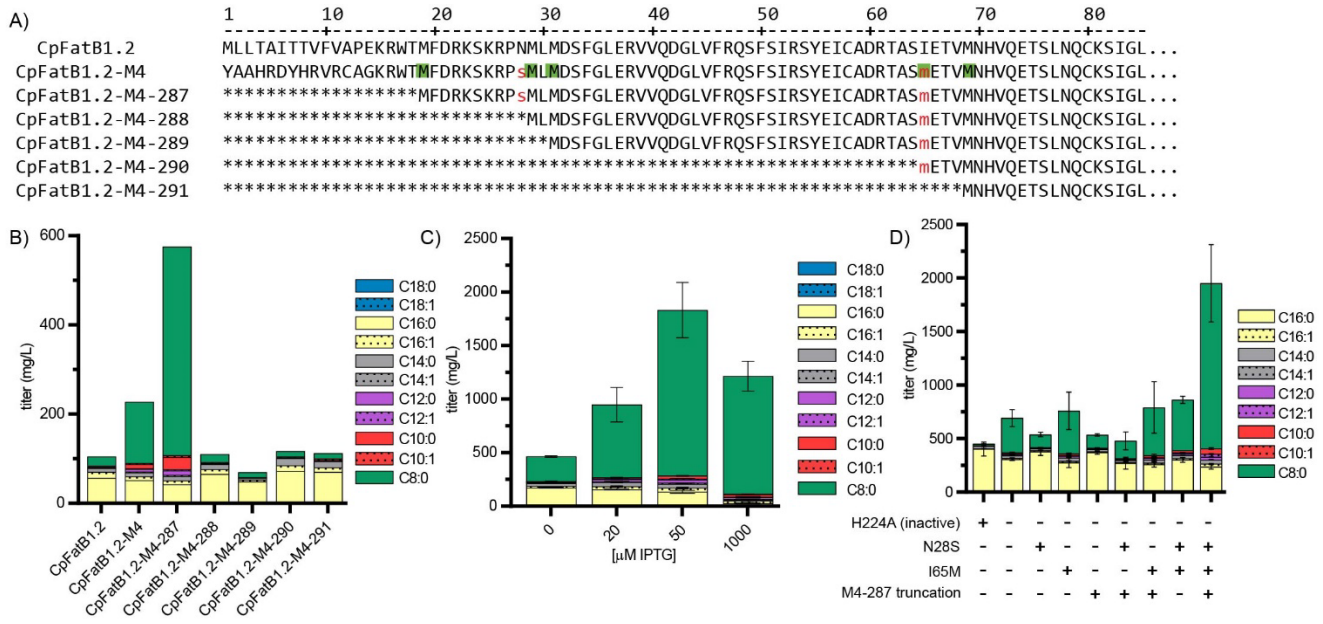
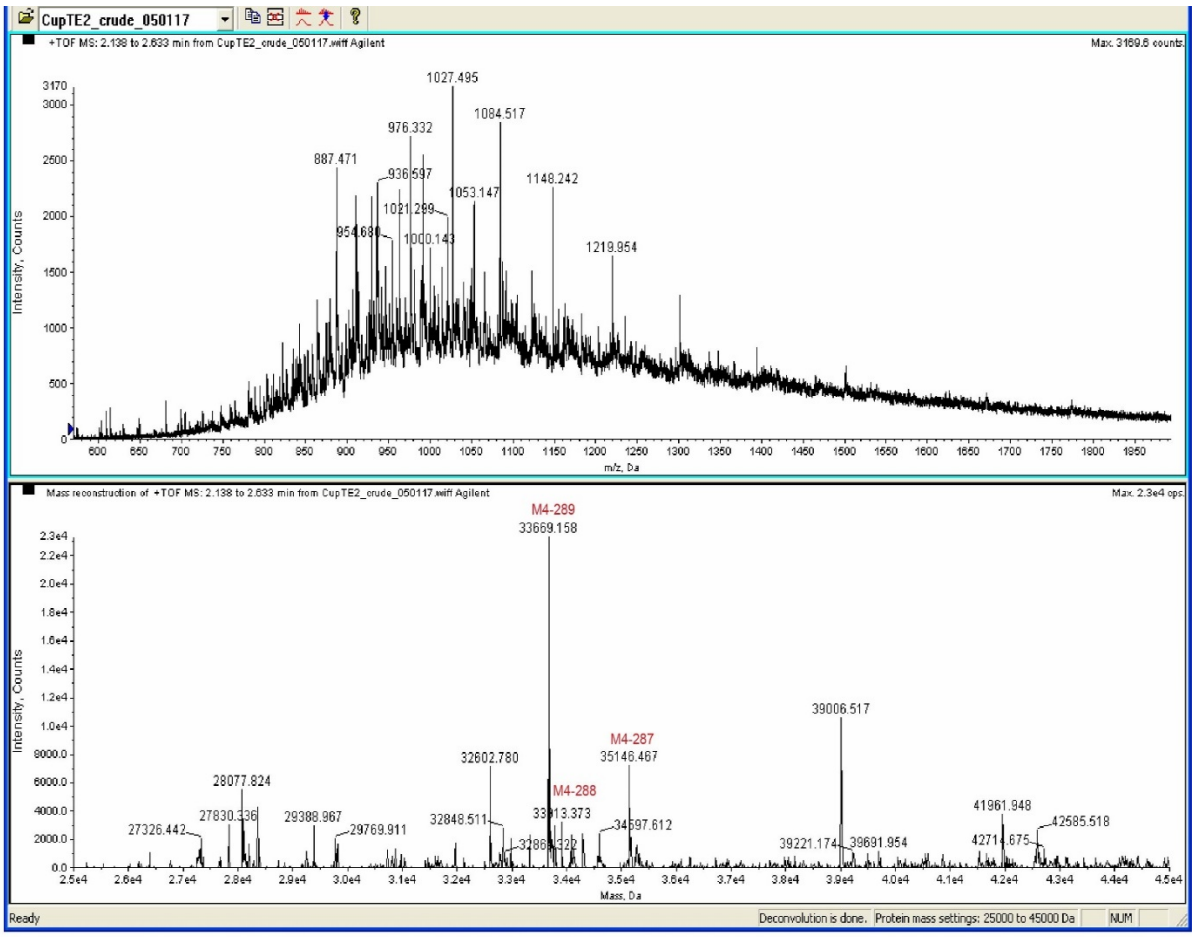


Figure 3.6. Characterization of CpFatB1.2-M4. A) Alignment of CpFatB1.2 with CpFatB1.2-M4 (ΔA54, N28S, I65M) shows a frame shift which that rose from nucleotide A_{54} deletion, producing a stop codon. Given that the enzyme is active, we proposed that it is being translated from a different in-frame 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[36] 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.

A)



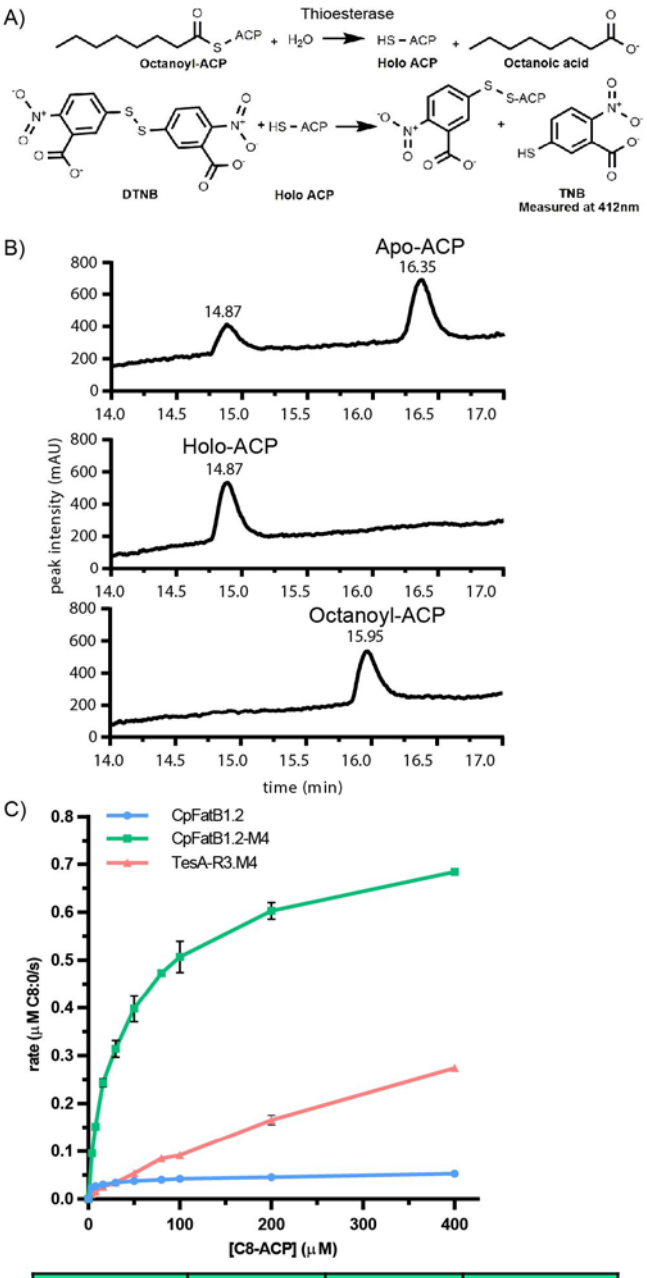
B)

Enzyme variant	Molecular weight (Da)
CpFatB1.2	37218
CpFatB1.2-M4-287	35149
CpFatB1.2-M4-288	33916
CpFatB1.2-M4-289	33672
CpFatB1.2-M4-290	29806
CpFatB1.2-M4-291	29345

Figure 3.7. Electrospray Ionization analysis of purified CpFatB1.2-M4-CHis. A) ESI showed the presence of truncation M4-289 as the major component with M4-287 and M4-288 as minor components. B) Predicted molecular weight for the C-Terminus Histidine tagged proteins that would result from each truncation.

3.1.1. Characterization of M4-287 variant *in vitro*

The *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 3.8A**). 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[143]. Complete synthesis of octanoyl-ACP was confirmed by HPLC (**Figure 3.8B**). 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 Chapter 2[127] 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 3.8C-D**). Interestingly, we found that a TesA-R3.M4 variant developed in prior work[127] 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.



D)

Mutant	kcat (min^{-1})	Km (μM)	kcat/Km (μM^{-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 3.8. Enzymatic assay to characterize our most active thioesterase. 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-M4-287 and TesA-R3.M4[127] 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.

3.3.5. **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, pBTRK-CpFatB1.2-M4-287, with an optimized RBS and tested FFA production under various induction levels (**Figure 3.9A**). 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 3.10**). 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 3.9B**, **Appendix IX**). As can be seen from **Figure 3.9B**, 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 3.10**), a phenomenon seen before[11]. 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 pBTRK-CpFatB1.2-M4-287 plasmid and 50 µM IPTG in NHL17) (**Figure 3.9A-B,**). 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 3.9C**) and minimal media (**Figure 3.9D**).

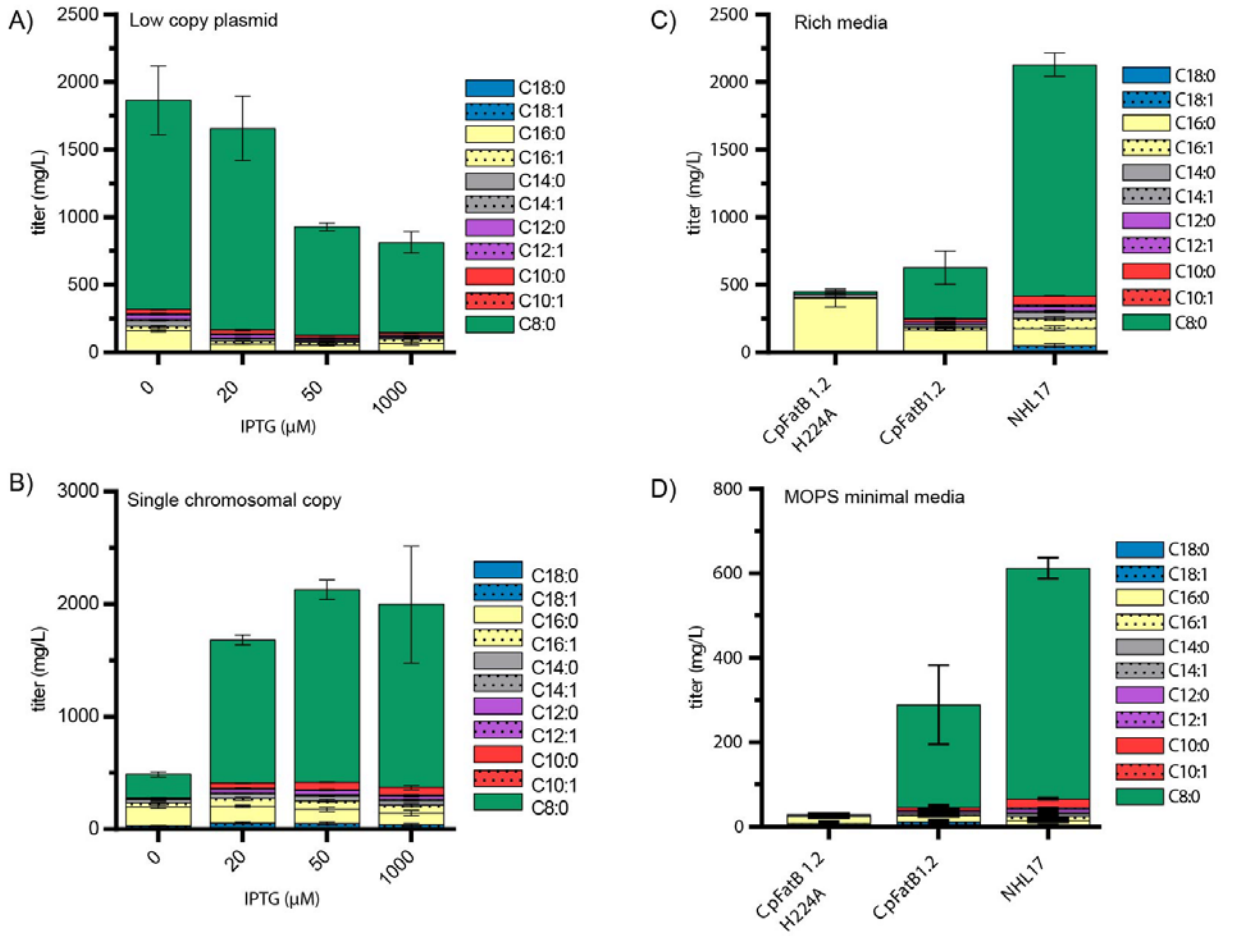


Figure 3.9. Optimization of expression. A) CpFatB1.2-M4-287 was cloned into a pBTRK plasmid with strong RBS test maximum production with plans to move the CpFatB1.2-M4-287 gene into the chromosome. After transforming into RL08ara (ΔfadD) strain and grown in MOPS media enriched[36] with tryptone and yeast 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[36] 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.

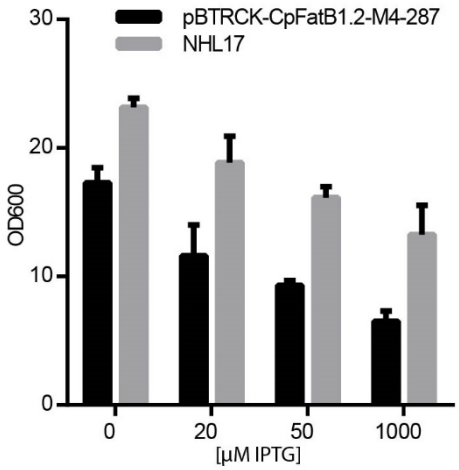


Figure 3.10. OD_{600} data for experiments in Figure 3.9. OD_{600} corresponding to Figure 3.9A (in black) and Figure 3.9B (in gray). Data shows growth burden associated with the level of expression of Cp-FatB1.2-M4-287.

3.3.6. **Structural modeling insights**

It has been experimentally shown that the CpFatB1.2-M4-287 mutant exhibits 15-fold higher specific activity *in vivo* (**Figure 3.6B**) 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 3.6D**). 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[129]) 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 3.11A-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.*[144] 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 3.11C**).

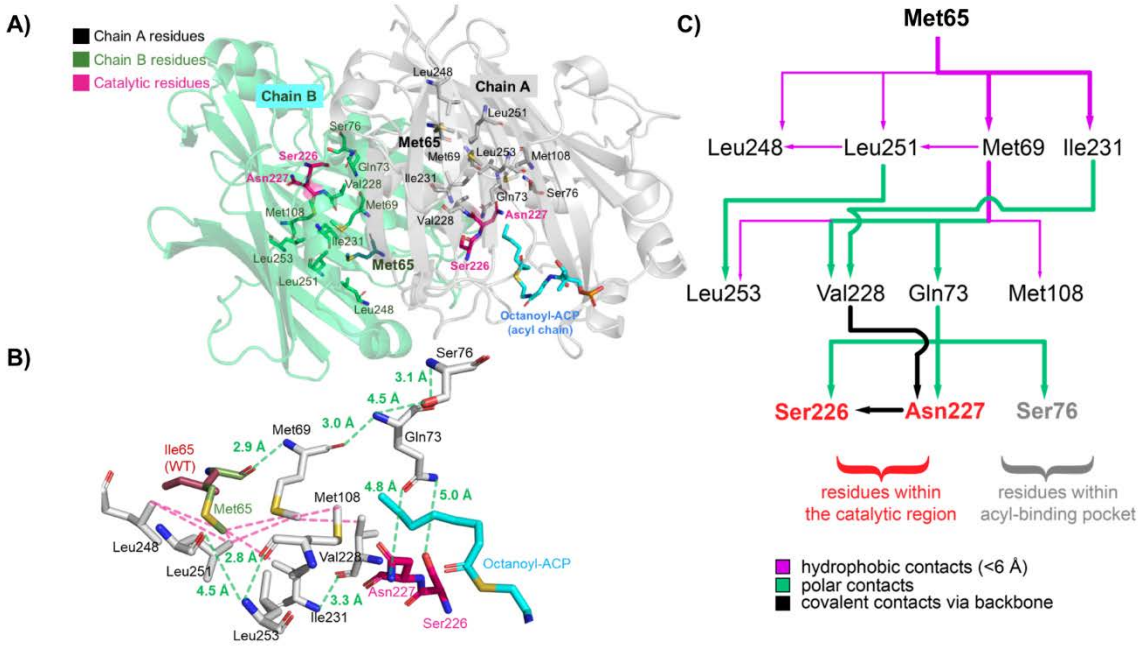


Figure 3.11. Effect of I65M mutation on CpFatB1.2-M4-287 mutant activity. (A) Structural overview of CpFatB1.2-M4-287 mutant. The mutated residues are shown in bold. The catalytic residues are marked in pink. The docked configuration of the substrate (blue) has been shown in one of the two (chain A) 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.

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-M4-287 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 3.12**). Prior to mutation the orientation of the Lys296 side chain did not result in a salt-bridge (~5.3 Å apart). Contact map (**Figure 3.12C**) 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-M4-287 (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 3.12C**). The aliphatic methyl-side chain of Ala59 before mutation prevented any such polar contacts.

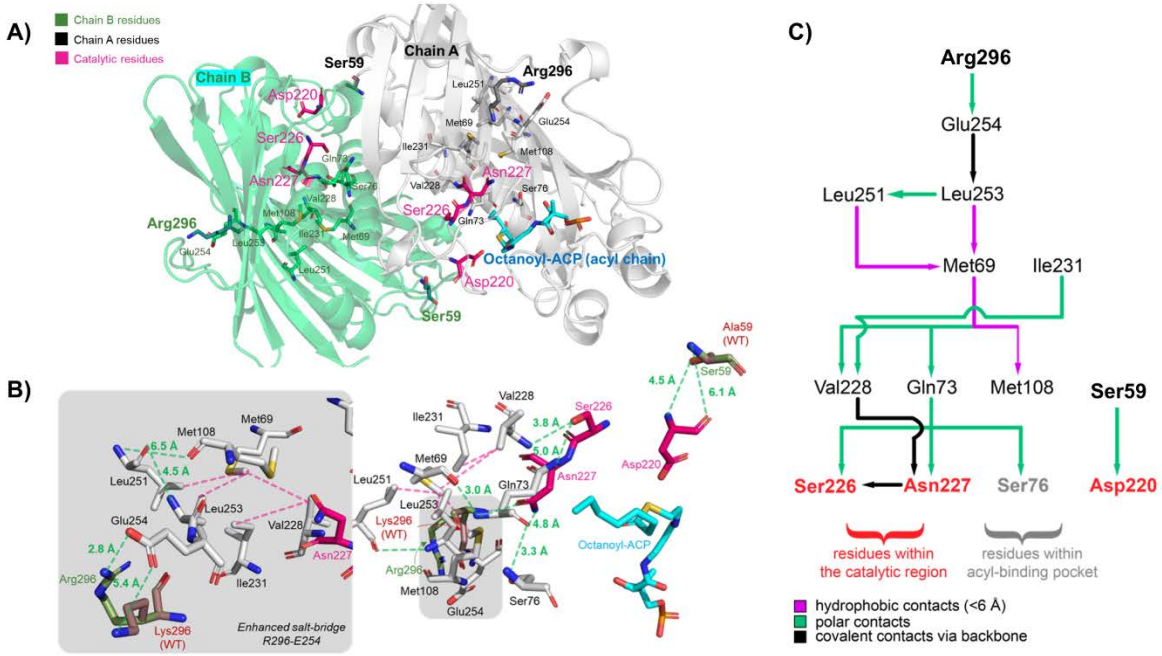


Figure 3.12. Effect of A59S and K296R mutations on CpFatB1.2-M3 mutant activity. (A) Structural overview of CpFatB1.2-M3 mutant. The mutated residues are shown in bold. The catalytic residues are marked in pink. The docked configuration of the substrate (blue) has been

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.

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.

3.4. Conclusion

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[10]. 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.

3.5. 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.

4. Production of 1-octanol in Escherichia coli by a high flux thioesterase route

4.1. Introduction

In Chapters 2 and 3 we focused on the thioesterase reaction to convert octanoyl-ACP into octanoic acid. We showed that the thioesterase reaction was a major bottleneck when using the fatty acid biosynthesis pathway. Using crystal structure guided computational approaches (Chapter 2) and biological selection (Chapter 3) we were able to engineer thioesterases that were more active and specific for octanoyl-ACP and achieved high flux to octanoic acid. The best variant thioesterase found (CpFatB1.2-M4-287 was able to achieve titers of 1.7 g/L octanoic acid in rich media and 0.6 g/L in minimal media from a single chromosomal copy. In this Chapter we will use this thioesterase to spearhead production of 1-octanol from endogenously made octanoic acid. Later, in Chapter 5 we will focus on other two oleochemicals, poly(3-hydroxyoctanoic acid) (PHO) and 2-heptanone.

1-octanol is a valuable aliphatic molecule that can be used as plasticizers, to synthesize surfactants and fatty amines, as a plant sucker growth regulator in the tobacco industry[149], as biodiesel additive[150] as well as dehydrated to 1-octene[151] to use as co-monomer in the production of LLDPE. The current price of 1-octanol is $1.66/lb and it is the highest valued fatty alcohol currently (C12 and C16 chain alcohols sell for ~$0.79/lb)[2]. 1-octanol is currently produced from petroleum derived ethylene, hydrogen, and

aluminum, via the Ziegler process and generates a linear alcohols mixture out of which 1-octanol is only a small fraction (~17%)[152].

Biological production of 1-octanol has been identified as a viable economical route at the current sugar prices if high yields are achieved[2]. However, unlike other products (i.e. short chain alcohols, short chain acids, amino acids, among others) which are advanced in their development, production of 1-octanol production is still in the proof of concept stage. One of the challenges towards making 1-octanol is the need of engineering biosynthetic pathways that are tailored towards medium chain lengths (C_8-C_{12}). Medium chain oleochemicals are a challenging group of molecule to produce at high titers in microbes for multiple reasons, the two most important being lack of active and specific enzymes to extract medium chain intermediates from the fatty acid biosynthesis pathway or other synthetic pathways such as the reverse β-oxidation[42,43] (see Chapter 1), and the fact that products of the medium chain range are inherently more toxic to the cell because they disrupt the membrane of the bacteria. Thus, addressing these issues are two of the main focus points in current efforts to engineer medium chain oleochemical production[10].

1-octanol can be made biologically from the reduction 1-octanal via an aldehyde reductase (**Figure 4.1, 7**). In a cellular context, fatty aldehydes are often a by-product of reactive oxygen species and long chain phospholipids. Since fatty aldehydes have a propensity to react and form adducts with amino groups in proteins and with DNA causing cellular damage, cells have devised very active aldehyde reductases that protects them from these adduct forming reactions[153]. Therefore, in producing 1-octanol from 1-octanal the major concern is how to make the latter with high titer and specificity. As can be seen

in (**Figure 4.**1), 1-octanal can be made biologically from 3 main intermediate precursors: octanoyl-ACP, octanoyl-CoA, and octanoic acid from two main routes that start from acetyl-CoA as the precursor, the fatty acid biosynthesis pathway (FAB) (**1**) and the reversal of the β-oxidation pathway (**8**). Octanoyl-ACP can be either directly reduced using and acyl-ACP reductase (AAR) (**6**)[41], or hydrolyzed via a C8-ACP thioesterase (**2**), followed by direct reduction with a carboxylic acid reductase (CAR)[150] (**5**) or converted to octanoyl-CoA using an acyl-CoA synthethase (ACS) (**3**) followed by reduction with an acyl-CoA reductase (ACR)[13] (**8**). For an overview of the fatty acid biosynthesis and the reversal of the β-oxidation pathway refer to Chapter 1. Although out of the FAB pathway options the AAR route (**6**) is theoretically the most energy efficient, it requires a reductase that is very active catalytically as well as specific toward C_8 acyl chains. On the other hand, the thioesterase/acyl-CoA reductase route can be benefited from the high flux that thioesterases can have, enough to deregulate the acyl-ACP mediated inhibition of the FAB pathway.

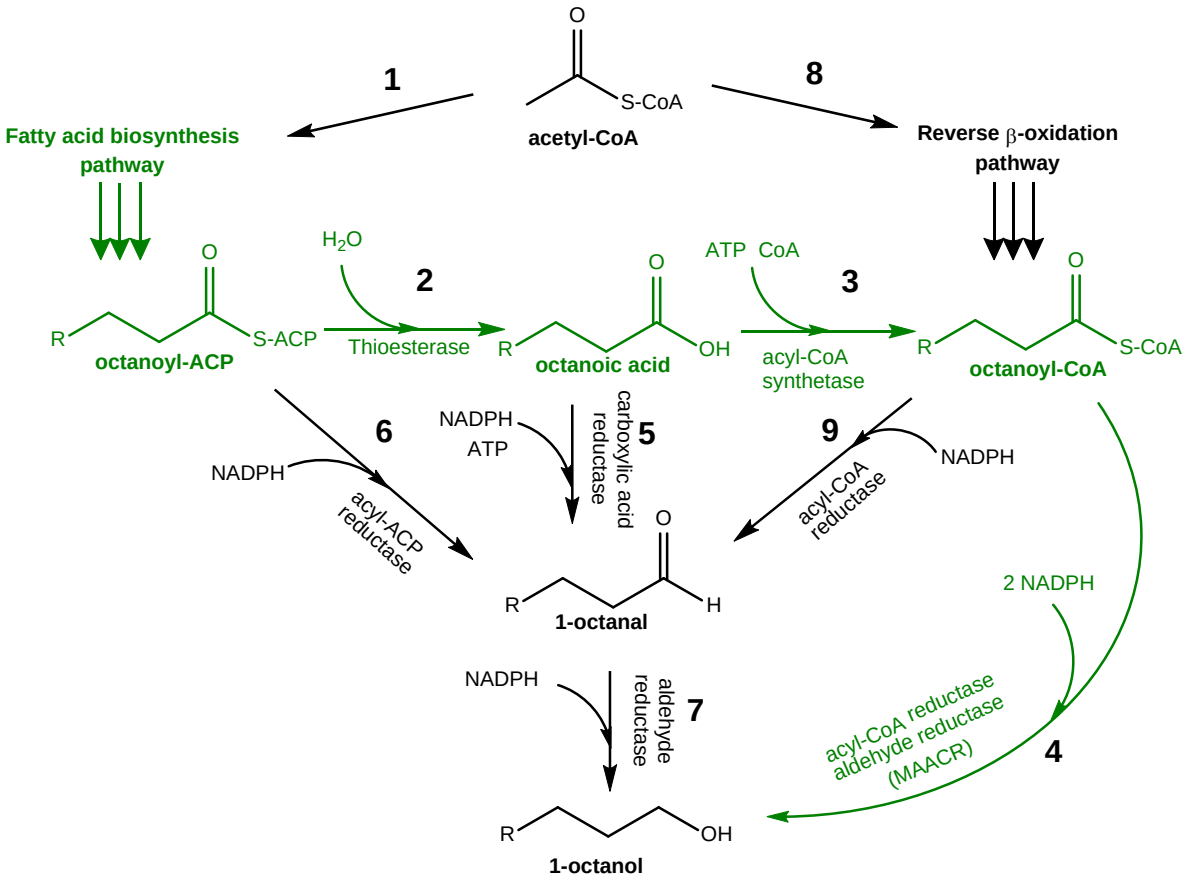


Figure 4.1. Main routes to 1-octanol. (green) the thioesterase/acyl-CoA reductase route to 1-octanol. (black) other routes to 1-octanol.

In Chapter 3, developed an acyl-ACP thioesterase from the shrub *Cuphea palustris* that has high activity as well as ~90% specificity for the C8 acyl chains[154] with a growth based selection platform that relies on the requirement for lipoylated E2 domain of pyruvate dehydrogenase in aerobic conditions[135], which can be supplied through octanoic acid produced by a thioesterase. Leveraging this thioesterase to drive the flux of the pathway we first explored various fermentation variables including host strain, fermentative pathway knockouts, and growth media to test maximum production of our

precursor octanoic acid from sugars. Then we tested a series of acyl-CoA synthethases (ACS) to replace *E. coli* FadD ACS which has been shown to have very low activity on octanoic acid[155]. I addition we characterized the pattern of evaporation of 1-octanol in a shake flask under the conditions that our fermentations would take place and found that dodecane can be used to alleviate evaporation issues. We then screened several acyl-CoA reductases (ACR) in order to identify variants with high catalytic activity towards octanoyl-CoA in aerobic conditions. Finally, we assembled the full pathway to demonstrate 1 g/L 1-octanol production from sugars in shake flasks and partially incorporated the pathway to the chromosome towards building an industrially relevant strain. Future work include increasing flux towards FAB pathway, reduction to a fully chromosomally expressed strain, and bioreactor studies.

4.2. Materials and Methods

4.2.1. Chemicals, Strains and Plasmids

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). See **Appendix I-II** for a list of strains and plasmids used in this study.

4.2.2. Free fatty acid experiments

For experiments designed to produce free fatty acids (**Figure 4.2**) single colonies were grown overnight with the appropriate antibiotics depending on the plasmids present (Chloramphenicol 34 μg/mL for pACYC; Carbenicillin 100 μg/mL for pTRC99a; Kanamycin 50 μg/mL for pBTRK) in LB media. For minimal media experiments (**Figure 4.2F**), 50 μL of LB media overnights were used to inoculate 5 mL minimal media and grown overnight before using the minimal media overnights for the experiments. Overnight cultures from three single colonies were used to inoculate 50 mL of media in a 250 mL shake flask at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2 before inducing with IPTG and moving to 30 °C for 24 h post induction. IPTG concentrations used were as follows: for **Figure 4.2A-C** induction is indicated on X axis; for **Figure 4.2D-F** and **Figure 4.7,** 50 μM IPTG was used which corresponds to the optimal induction level for the pTRC99a-CpFatB1.2-M4-287 plasmid (see **Figure 4.2C**); for **Figures 4.3 and 4.6,** 1 mM IPTG was used. Media used for **Figures 4.2A-D and Figure 4.7** was MOPS rich media glycerol described in Kim, et al.[36] with the following changes: 1.39 mM Na_2HPO_4, no biotin, thiamine or sodium selenite added). For **Figure 4.2E**, we replaced the glycerol from the MOPS rich media described above with glucose.

For **Figure 4.2F** we used MOPS minimal media phosphate limited as described elsewhere[142]. All other experiments were done on LB media.

4.2.3. Fatty acid extraction, methylation and quantification

Fatty acid samples were prepared by collecting the equivalent of 2.5 mL culture into a 10 mL glass centrifuge tube, such that 2.5 mL was collected when no dodecane overlay was added to the culture, 2.75 mL when 1:10 dodecane to media ratio was used and 3.0 mL when 1:5 dodecane to media was used. 50 µL of internal standard containing 12.5 mg/mL nonanoic acid and 1.25 mg/mL pentadecanoic acid was added to the samples. The extraction and methylation procedure was done as described in Chapter 2 (see section 2.2.7)[156]. For GC quantification, samples were analyzed using a GC-FID model Shimadzu GC-2010 equipped with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The column oven temperature protocol was 100°C for 2 min, ramp to 125°C (at 7°C/min), ramp to 235°C (at 10°C/min), ramp to 250°C (at 40°C/min), and hold for 2 min.

4.2.4. Fatty alcohol experiments

For experiments designed to produce fatty alcohols (**Figure 4.5, Figure 4.7**) single colonies were grown overnight with the appropriate antibiotics depending on the plasmids present (Chloramphenicol 34 µg/mL for pACYC; Carbenicillin 100 µg/mL for pTRC99a; Kanamycin 50 µg/mL for pBTRK) in LB media. Overnight cultures from three single colonies were used to inoculate 50 mL of media in a 250 mL shake flask (for **Figure 4.5**

we used LB + 500 mg/L octanoic acid; for **Figure 4.7** MOPS rich media glycerol described in Kim, et al.[36] with the following changes: 1.39 mM Na_2HPO_4, no biotin, thiamine or sodium selenite added) at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2 before inducing with 1mM IPTG (**Figure 4.5)** or 50 μM IPTG (**Figure 4.7**), 10 mL dodecane was added to control evaporation and the cultures were moved to 30 °C. Samples at different time points were collected. For octanoic acid consumption experiments in the presence of ACS and/or ACR (**Figure 4.5**) 100 μL of a 250 mg/mL octanoic acid solution in ethanol was added immediately after induction.

4.2.5. **Fatty alcohol extraction and quantification**

Before collecting fatty alcohol samples, 100 μL of an internal standard containing 125 mg/mL nonanol, and 12.5 mg/mL pentadecanol was added to the shake flasks and allowed to shake for 5 min to ensure mixing. Cultures in the presence of dodecane were constantly mixed right before the sample is taken and the sample was taken immediately after the shake flask stops mixing. Fatty alcohol samples were prepared by collecting the equivalent of 2.5 mL culture (same as free fatty extraction section) into a 10 mL glass centrifuge tube. Samples were then extracted into 1 mL hexane by vortexing the solution for 3 min at max speed. Samples where then centrifuged at 3000 g for 10 min and the hexane layer was injected in a GC-FID. Samples were analyzed using a GC-FID model Shimadzu GC-2010 equipped with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The column oven temperature protocol was 100°C for 2 min, ramp to

125°C (at 7°C/min), ramp to 235°C (at 10°C/min), ramp to 250°C (at 40°C/min), and hold for 2 min.

4.2.6. **Addressing quantification issues**

In **section 4.3.3.** we discussed how 1-octanol quantification in the presence of dodecane had an underestimation bias (see **Figure 4.5**). This was a phenomenon that was observed only when the samples were taken out of a shake flask into a glass centrifuge tube for further processing. It should be noted that a media:dodecane mixture has the characteristic of separating into two layers within several seconds with the top layer being dodecane rich and the bottom the media rich. This makes the sampling of the culture challenging as the shake flask has to be continuously mixed up to the point before the sample is taken and the sample has to be taken immediately after the shake flask stops mixing. We hypothesized that this sampling process might be behind the underestimations if the sample taken was slightly biased towards taking more of the media rich layer than the dodecane rich. Although we routinely add internal standards to account for these type of losses during the sample processing, our protocols were designed to add the internal standard to the sample once it is in the centrifuge tube. This internal standard practice would not be able to correct for any losses prior to adding the sample to the tube (something that is typically not an issue in one-phase systems) and thus we hypothesized that adding the internal standard to the shake flask would correct this issue. To test this, we created an 10X concentrated internal standard such that upon adding the appropriate volume to the shake flask it would result in the same concentration of internal standard in the media as in the internal standard added to the tube. We tested side by samples of media containing known amounts of 1-octanol ranging from 50 to 2000

mg/L with internal standard added directly to the shake flask or added to the tubes post sampling (**Figure 4.2**). As can be seen in **Figure 4.2A**, when the internal standard is added directly to the shake flask the quantified 1-octanol line is closer to the 1:1 line at any given concentration tested. Percent error difference from the expected values was calculated for each sample and shows a large error rate (up to 40%) for the samples when the internal standard was added to the sample in the tubes versus errors ranging from <1% to 12% for the samples with internal standard added to the shake flask (**Figure 4.2B**). Based on these results we decided to modify the sampling protocol to add the internal standard to the shake flask prior to taking samples for all the samples that involved dodecane as an organic layer.

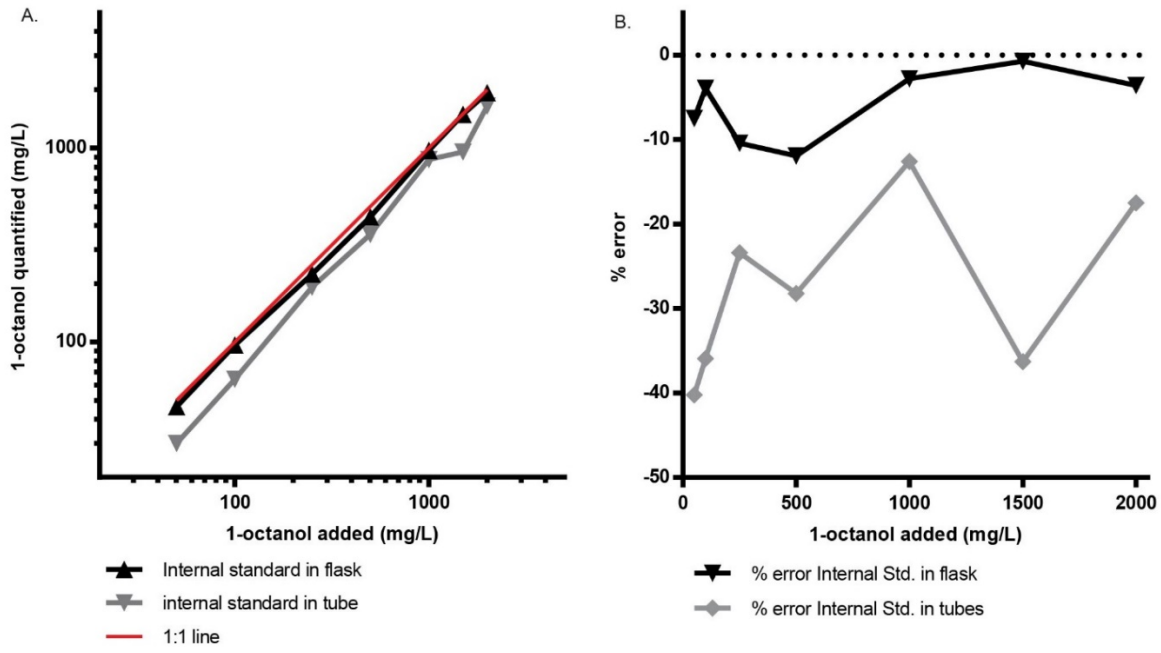


Figure 4.2. Quantification of 1-octanol from shake flasks containing dodecane.

4.3. Results and Discussion

4.3.1. Octanoic acid production

In a Chapter 3[154], we developed a highly active thioesterase (CpFatB1.2-M4-287) for the production of octanoic acid from octanoyl-ACP made through the FAB pathway. Leveraging this thioesterase as a starting point we decided to pursue 1-octanol production. We first explored octanoic acid production from two different strains, the common K12 MG1655 and LS5218[157]. LS5218 is an *E. coli* K12 derived strain that was obtained after selection on first decanoate and then butyrate, and that has shown success for alcohol production in anaerobic conditions in the past[43] as well as poly(3-hydroxyalkanoates)[157]. In Chapter 3[154], we demonstrated that a single chromosomal copy of CpFatB1.2-M4-287 in the *fadD* loci was sufficiently expressed to produce octanoic acid at high titers (up to 1.7 g/L) from MOPS rich media glycerol (see materials and methods). We performed the same chromosomal integration of CpFatB1.2-M4-287 in the *fadD* loci into the LS5218 strain and compared the two strains. Cells were inoculated from overnight cultures into 50 mL of rich media with glycerol (see materials and methods) and induced at OD_{600} of 0.2 with 0 μM, 20 μM, 50 μM and 1000 μM IPTG and grown at 30 °C for 24 h post induction. The LS5218 single chromosomal copy of CpFatB1.2-M4-287 peaked at 0.8 g/L of octanoic acid production at maximum expression compared to 1.7 g/L for MG1655 (**Figure 4.3A**). In order to control for different chromosomal expression levels that may arise from the two different strains, *E. coli* strains MG1655 *ΔfadD* and

LS5218 *ΔfadD* were transformed with either high copy plasmid pTRC99a with a weak ribosome binding site (RBS) or a low copy plasmid pBTRK with a strong RBS. The high copy plasmid was selected on the basis of previous observation that MG1655 *ΔfadD* strain was able to maximally express CpFatB1.2-M4-287 at non saturated induction conditions (see Chapter 3)[154]. The low copy plasmid was selected because the full construct in it (including promoter and RBS) is the same as the chromosomally integrated construct, which serves as an indication of the outcome expected if extra copies were to be added to the chromosome. This low copy plasmid was designed to have strong promoter and RBS to maximize expression from a single copy when moved to the chromosome. For the case where low copy plasmid was used (**Figure 4.3B**), MG1655 peaks in production at 1.5 g/L octanoic acid at no induction while LS5218 peaked at 50 μM IPTG with 0.94 g/L. Moreover, when the high copy plasmid was used octanoic acid production of MG1655 peaked at 50 μM IPTG with 1.5 g/L whereas LS5218 peaked at 1 mM IPTG with 1.1 g/L (**Figure 4.3C**). This data shows that for all, chromosomal, low copy plasmid and high copy plasmid MG1655 was able to outperform LS5218 octanoic acid production in aerobic conditions. Since no improvement was seen from strain LS5218 we decided to continue using MG1655 towards 1-octanol production.

Some of the most common byproducts that *E. coli* cells produce when grown at high carbon source concentrations are acetate and lactate. Research has shown that removing the enzymes related to these pathways can have a positive effect in the yields since more carbon source is available to produce the product of interest[158]. For these reasons, we performed a series of fermentation knockouts of the acetate and lactate pathways, phosphate acetyltransferase (*pta*), pyruvate oxidase (*poxB*), and lactate

dehydrogenase (*ldhA*) in the MG1655 *ΔfadD* strain and made strains NHL09 (MG1655 *ΔfadD Δpta*), NHL10 (MG1655 *ΔfadD Δpta ΔpoxB*) and NHL11 (MG1655 *ΔfadD Δpta ΔpoxB ΔldhA*). These strains were transformed with pTRC99a-CpFatB1.2-M4-287 plasmid using 50 μM IPTG induction level, optimal for this plasmid (**Figure 4.3C**) while using different media conditions (**Figure 4.3D-F**). We tested in total three media conditions, rich media with glycerol (**Figure 4.3D**), rich media with glucose (**Figure 4.3E**), and MOPS minimal media phosphate limited with glucose (**Figure 4.3F**).

For the case rich media with glycerol (**Figure 4.3D**), we saw a decrease in titers from 1.4 g/L octanoic acid in the control strain with no deletions in fermentation pathways to 0.8 g/L in the strain with *Δpta*. For strains with *Δpta ΔpoxB* and *Δpta ΔpoxB ΔldhA* deletions the titers were not statistically significantly different from the control strain. For cultures grown in rich media with glucose (**Figure 4.3E**) the control strain and the *Δpta* strain produced 55% less octanoic acid (0.6 g/L) compared with the same strain in rich media glycerol (1.4 g/L). Interestingly, *Δpta ΔpoxB* and *Δpta ΔpoxB ΔldhA* had a drastic negative effect in production from glucose and accumulated only 0.17 and 0.27 g/L octanoic acid respectively. Finally, for MOPS minimal media phosphate limited with glucose (**Figure 4.3F**) all strains produced 0.40-0.49 g/L and no major impact was seen for having the additional deleted pathways. Overall, we did not find that fermentative pathways were a major bottleneck at this point leading to increased production. This could still be revisited for a strain with fully optimized central metabolism.

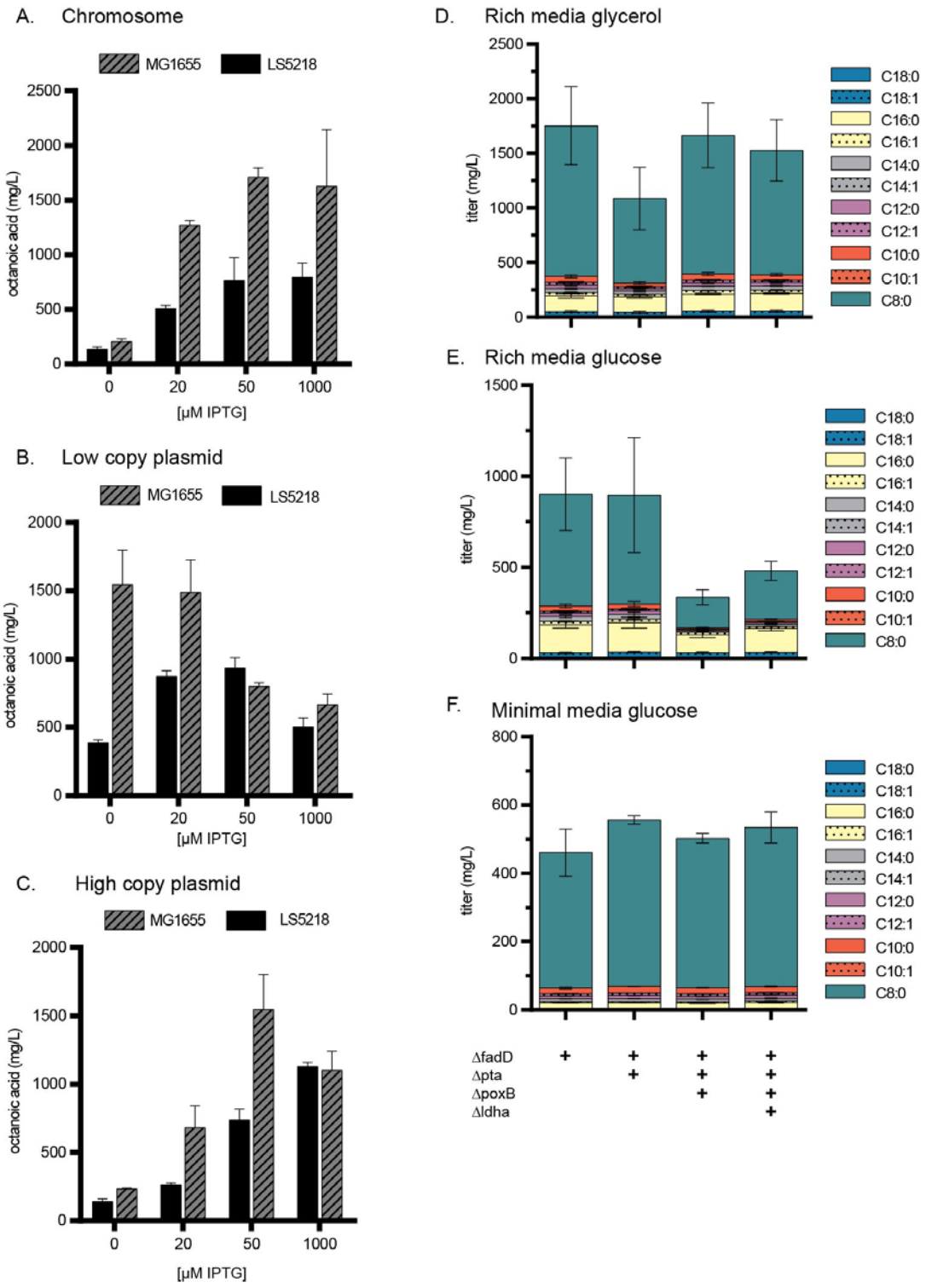


Figure 4.3. Octanoic acid production of various fermentation knockout strains (A) Fatty acid production at different induction levels for chromosomal integration of CpFatB1.2-M4-287 in the *fadD* loci into LS5218 and MG1655 strains. (B-C) Fatty acid production at different induction levels for MG1655 *ΔfadD* and LS5218 *ΔfadD* transformed with either a low copy plasmid pBTRK-CpFatB1.2-M4-287 with a strong RBS (B) or high copy plasmid pTRC99a- CpFatB1.2-

M4-287 with a weak RBS (C) (see discussion for details). (D-F) MG1655 *ΔfadD* with additional fermentation pathway deletions carrying pTRC99a- CpFatB1.2-M4-287 and tested in total three media conditions, rich media with glycerol (D), rich media with glucose (E), and MOPS minimal media phosphate limited with glucose (F).

4.3.2. **Acyl-CoA Synthetase**

One of the challenges that needed to be addressed in order to use the FAB/ACR route is finding an acyl-CoA synthetase (ACS) that will have high activity towards octanoic acid. It has been previously shown that *E. coli* FadD has low enzymatic activity towards acyl chains of less than 10 carbons[159]. Efforts to engineer *E. coli* acyl-CoA synthetase have had only very limited sucess[160]. We tested some alternate ACS from *Pseudomonas putida* and *Mycobacterium tuberculosis* to evaluate their activity towards octanoic acid relative to *E. coli* fadD. For this, we expressed plasmids carrying these enzymes in strain MHS04 (*E. coli* MG1655 Δ*fadR* Δ*fadD*)[13] in the presence of octanoic acid. Strain MHS04 has *fadR* deletion resulting in the deregulation of the β-oxidation pathway and the *fadD* deletion makes the plasmid expression of ACS necessary for the consumption of octanoic acid. Cells were inoculated from LB media overnight cultures into 50 mL of LB media with 250 mg/L octanoic acid (see materials and methods), induced at OD_{600} of 0.2 with 1 mM IPTG and grown at 30 °C for 12 h post induction. Cells carrying *Mycobacterium tuberculosis* (Mt-FadD6) was able to consume most of the octanoic acid by 5 h post induction and *Pseudomonas putida* (Pp-00763) were able to consume most of the octanoic acid by 8 h (**Figure 4.4**). *E. coli* FadD lagged behind unable to consume octanoic acid as fast. This data suggests that both Mt-FadD6 and Pp-00763 are highly active on

octanoic acid. For the rest of this study, we used Mt-FadD6 when expressing an acyl-CoA synthetase.

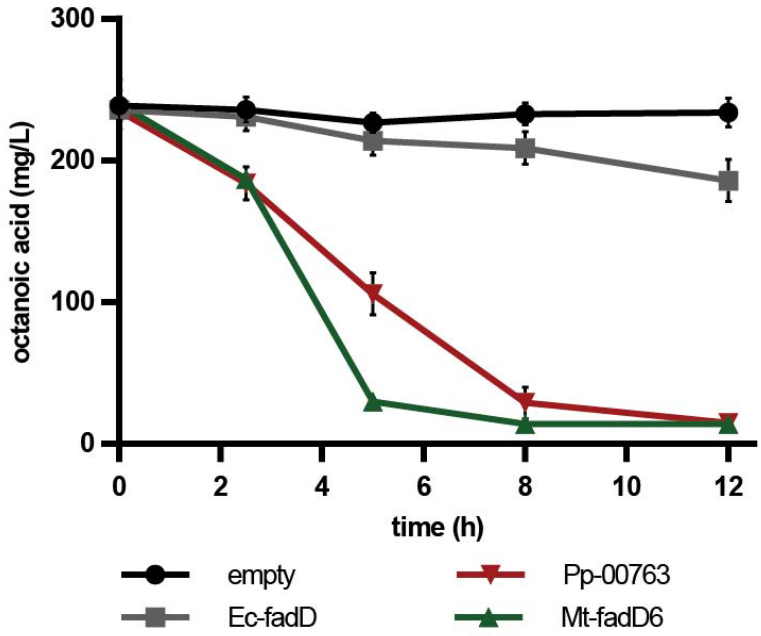


Figure 4.4. acyl-CoA Synthetase with activity on octanoic acid. *E. coli* MG1655 Δ*fadR* Δ*fadD* cells carrying plasmids with various acyl-CoA synthethases were grown from LB media overnight cultures into 50 mL of LB media with 250 mg/L octanoic acid (see materials and methods) and the octanoic acid consumption was tracked over time.

4.3.3. **1-Octanol evaporation**

In some of our preliminary work (not shown here) one of the factors that we noticed was a significant smell of 1-octanol upon producing it in aerobic conditions from fed octanoic acid but a lack of product after 24 h. We decided to explore the possibility that our product was evaporating from the shake flasks. To do this we set up shake flasks containing 50 mL of LB media with a known amount of 1-octanol (250 mg/L) and took

samples at different times (**Figure 4.5**). As can be seen in the black circles, it took only 12 h after adding 1-octanol to evaporate 85% of the 1-octanol added to the LB media.

Previous studies have used dodecane as a secondary organic layer in cell cultures to alleviate toxicity issues that can arise from dodecanol production[13]. We decided to explore additions of dodecane to extract the 1-octanol into this organic layer and lowering the vapor pressure of 1-octanol, alleviating the losses to evaporation. To do this we repeated the experiment with additional dodecane added to the 50 mL LB media with 250 mg/L 1-octanol. We added 5 mL (1:10) and 10 mL (1:5) dodecane and tracked the evaporation of 1-octanol over time. Interestingly, addition of dodecane did assist to reduce evaporation (**Figure 4.5**) although some losses still occurred. After 24 h, in the flasks containing 1:10 dodecane we were able to quantify 56% of the 1-octanol added at the beginning of the experiment and for 1:5 dodecane 70% of the dodecane was quantified. Based on this results we decided to do all of our testing of 1-octanol in shake flasks containing 1:5 v:v ratio of dodecane:media. However, dodecane containing cultures seemed to have a downward bias (approximately 20%) in the calculation of concentration. This can be seen in the difference observed 1-octanol at early time points between the flasks containing only LB media vs the flasks containing LB + dodecane (**Figure 4.5**). This is a problem that was addressed in section 4.2.6.

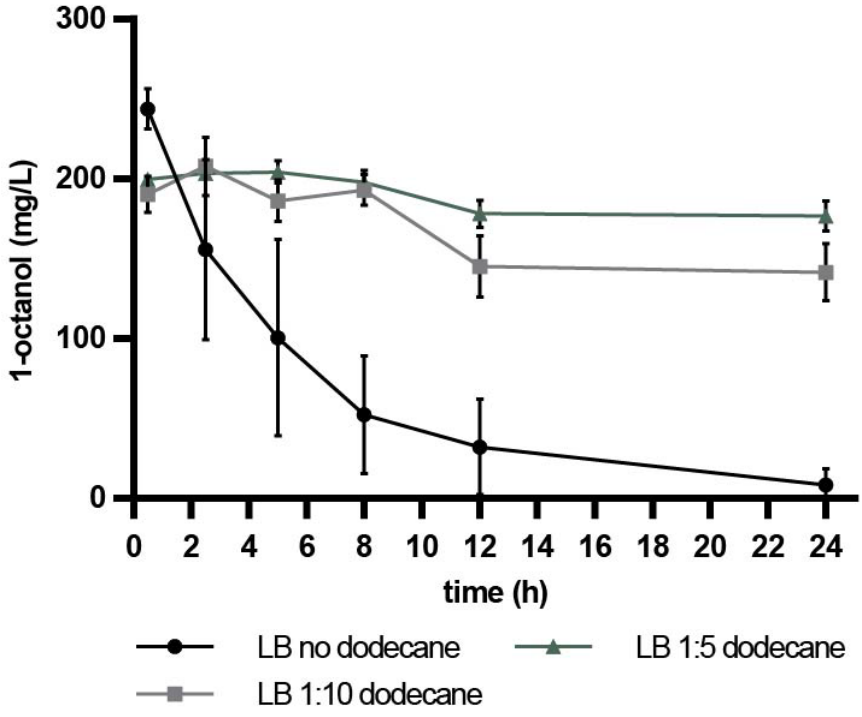


Figure 4.5. Evaporation of 1-octanol. 1-Octanol evaporation in 250 mL shake flasks was tracked over time to understand evaporation behavior in aerobic conditions. When no dodecane is added, 250 mg/L of 1-octanol is lost to evaporation over a period of 24 hr. Adding 1:5 and 1:10 volumetric ratios of dodecane alleviate evaporation. However, values of 1-octanol in the presence of dodecane are underestimated.

4.3.4. **Acyl-CoA reductase**

Given that we found a thioesterase in Chapter 3 (CpFatB1.2-M4-287) and an acyl-CoA synthetase in this Chapter (Mt-FadD6) which have high activity towards octanoyl-ACP and octanoic acid, respectively, we decided to complete the pathway to 1-octanol by adding an acyl-CoA reductase (ACR). Before testing 1-octanol production from sugars we decided to test production from fed octanoic acid in order to ensure that the acyl-CoA reductase was functional towards octanoyl-CoA. To do this we started from a MG1655 *ΔfadD ΔfadE* strain which blocks β-oxidation to metabolize any acyl-CoA into enoyl-CoA

and has only ACS activity if the ACS is provided in a plasmid. This strain was transformed with two plasmids containing the ACS (Mt-FadD6) and the ACR from *Marinobacter aquaeolei* (Ma-ACR)[34]. Three controls were used: (1) plasmids containing the ACS but not the ACR (empty plasmid was used when an enzyme was not expressed), (2) plasmids containing the ACR but not the ACS, and (3) plasmids containing neither the ACS nor ACR. 250 mL shake flasks with 50 mL LB media was inoculated with overnight cultures at OD_{600} = 0.05 and plasmids where induced at OD_{600} = 0.2 with 1 mM IPTG. 500 mg/L octanoic acid was added immediately after induction and samples were taken over time to track the consumption of octanoic acid (**Figure 4.6A**). As can be seen from **Figure 4.6A** the octanoic acid was consumed only in cells containing both the acyl-CoA synthetase and the acyl-CoA reductase. Moreover, Ma-ACR is able to turnover most of the octanoyl-CoA by 20 h. All other cases resulted in unconsumed substrate. It should be noted that the acid catalyzed derivatization method used for the quantification of free fatty acids cannot differentiate between the free acids and the CoA species as it methylates both. Thus, the sample containing ACS but no ACR (which is expected to accumulate CoA species rather than the acid) appears to not have consumed the acid.

In addition, we decided to test eight additional acyl-CoA reductases to look for faster turnover of octanoic acid into alcohols in aerobic conditions. These eight variants, which were obtained using enzyme similarity tools in previous work of our laboratory done by Mehrer, et al.[43], were originally tested anaerobically. In order to test these variants, we repeated the experiment of **Figure 4.6A** but instead we tracked the production of 1-octanol while changing the acyl-CoA reductase enzyme present (**Figure 4.6B**). In addition to Ma-ACR, two other enzymes (Mt-ACR from *Methylibium sp.* and Mb-ACR from

Marinobacter BSs20148) were observed to have high turnover of octanoyl-CoA to 1-octanol. In Mehrer, et al.[43] Mt-ACR was seen to express well and produce alcohols in anaerobic conditions, but Mb-ACR did not express or produce alcohols. Interestingly, we found that both of them produce alcohols at the same levels as Ma-ACR in aerobic conditions (**Figure 4.6B**). It is not clear which of these enzymes produced the most 1-octanol from aerobic conditions since there is some evaporation occurring simultaneously. For this reason, we decided to continue using Ma-ACR throughout the study. Furthermore, after 24 h, the levels of 1-octanol accumulated 30-50% percent lower than the 18 h time point suggesting that there is still considerable evaporation occurring and the actual 1-octanol production might be higher than the accumulation seen in the flask.

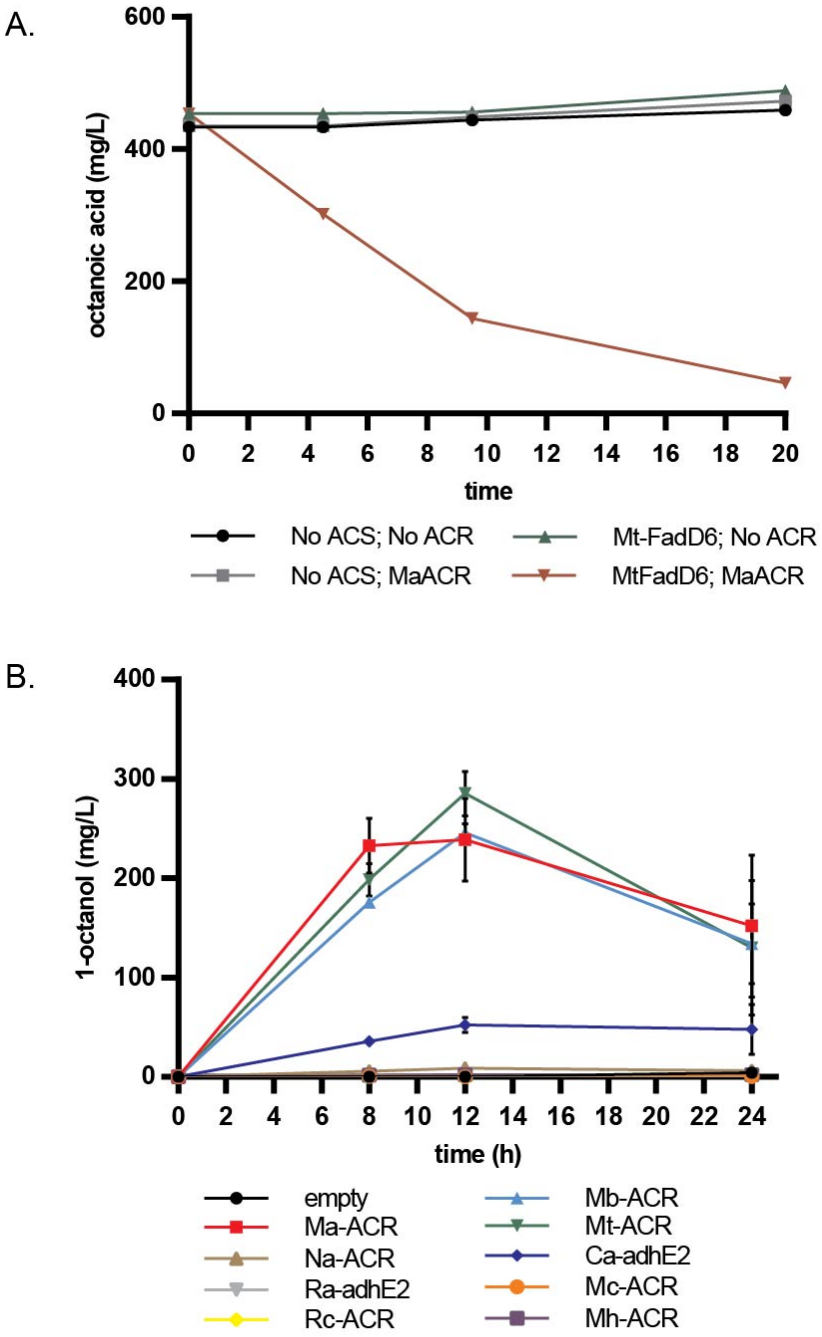


Figure 4.6. Production of 1-octanol from octanoic acid. (A) 500 mg/L of octanoic acid in LB media was fed to MG1655 Δ*fadD* Δ*fadE* strain carrying pACYC-Mt-FadD6 and pBTRK-MaACR to show consumption of octanoic acid upon having the full pathway assembled. Addition of ACS or ACR alone did not result in conversion to 1-octanol. ACS only results in conversion to octanoyl-CoA (Figure 2) but FAME extraction using HCl-MeOH is able to readily methylate it, rendering it indistinguishable from octanoic acid. (B) 500 mg/L of octanoic acid in LB media was fed to MG1655 Δ*fadD* Δ*fadE* strain carrying pACYC-Mt-FadD6 and different acyl-CoA

reductases in a pBTRK plasmid to determine an optimal ACR. Ma-ACR, Mt-ACR and Mb-ACR were the best performer with similar peak production.

4.3.5. **Assembly of the full pathway from glycerol**

In Chapter 3 we engineered a thioesterase from the plant *Cuphea palustris* that has high enzymatic activity and specificity towards octanoyl-ACP to produce octanoic acid. In this Chapter we have shown that an acyl-CoA synthetase from *Mycobacterium tuberculosis* (Mt-FadD6) and an acyl-CoA reductase from *Marinobacter aquaeolei* are able to effectively convert octanoic acid to 1-octanol. In this section we will assemble the complete the pathway in order to test the ability of *E. coli* to produce 1-octanol from sugars. We did this by first expressing all enzymes in separate plasmids and then systematically moving the thioesterase (CpFatB1.2-M4-287) and acyl-CoA synthetase (Mt-FadD6) to the chromosome. The end goal is to have all enzymes chromosomally expressed while maintaining the levels of production since this is an important step towards creating industrially relevant strains.

First, we assembled the full pathway, thioesterase (CpFatB1.2-M4-287), acyl-CoA synthetase (Mt-FadD6), and acyl-CoA reductase (Ma-ACR) into three separate plasmids: pTRC99a, pACYC, and pBTRK, respectively (**Figure 4.7A** and **Figure 4.7B** left two panels). A control where the ACS (Mt-FadD6) was absent was used in order to track the free fatty acid accumulation. Plasmids were transformed into DE strain (*E. coli* MG1655 *ΔaraBAD ΔfadD ΔfadE*) and cultures were inoculated from overnight in 250 mL shake flasks with 50 mL MOPS rich media (see materials and methods) and time points at 12 h, 18 h, and 24 h were collected to track evaporation effects. As can be seen in the left

two panels of **Figure 4.7A**, stains lacking Mt-FadD6 ACS accumulate 1.1 g/L octanoic acid in these conditions after 18 h while not being able to accumulate any alcohols (**Figure 4.7B** left two panels). The observed C_{16} and C_{18} fatty acids in these graphs are representative of the membrane lipids as the acid catalyzed methylation reaction is able to methylate those esters as well as the free fatty acids (see section 2.2.7). When the Mt-FadD6 ACS is included in the strain cells accumulate 1.0 g/L 1-octanol after 18 h and minor amounts of other chain lengths (**Figure 4.7B** left two panels). Interestingly, we have seen a similar strain (*E. coli* MG1655 *ΔaraBAD ΔfadD*) carrying the same pTRC99a-CpFatB1.2-M4-287 and expressed under the same conditions (**Figure 4.1D**) produce 50% more octanoic acid. This suggests that there may be a burden associated with carrying 3 plasmids in the cell.

Next, we sought to introduce the thioesterase (CpFatB1.2-M4-287) developed in Chapter 3 into the chromosome in order to work towards reducing the strain to only chromosomally based. For this we constructed strain NHL19 (K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287 ΔfadE)* and repeated the experiment as described above with pTRC99a empty plasmid taking the place of the pTRC99a-CpFatB1.2-M4-287 plasmid. This time it can be seen in the center two panels of **Figure 4.7A**, stains lacking Mt-FadD6 ACS accumulating 0.9 g/L octanoic acid in these conditions after 18 h while not being able to accumulate any alcohols (**Figure 4.7B** center two panels). When the Mt-FadD6 ACS is included in the strain cells accumulate 0.6 g/L 1-octanol after 18 h and minor amounts of other chain lengths (**Figure 4.7B** center two panels).

Finally, we constructed strain NHL22 (K-12 MG1655 *ΔaraBAD ΔfadD::*PtrcCpFatB1.2-M4-287 *ΔfadE ΔackA-pta::Ptrc-Mt-FadD6*) which in addition to

CpFatB1.2-M4-287 also includes *Mt-FadD6* ACS in the chromosome. We repeated the experiment as described above with and without pACYC-Mt-FadD6 plasmid in order to see the chromosomal expression versus the plasmid on the same strain. Interestingly, strain NHL22 was sufficient to keep the same level of 1-octanol made by the NHL19 (**Figure 4.7B** right two panels) and additional plasmid expression of Mt-FadD6 did not increase the performance of the strain in terms of 1-cotanol production after 18 h, albeit the earlier 12 h time point shows less accumulation of 1-octanol.

Overall, the largest decrease in 1-octanol production was observed from moving the thioesterase to the chromosome. This could be due to faster kinetics for making octanoic acid coupled with evaporation of 1-octanol since the 12 h time points for plasmid expression (**Figure 4.7A**, left two panels) reveal that the cells carrying the plasmid for the thioesterase have significantly more octanoic acid accumulated compared to the 12 h time point for the chromosomally expressed thioesterase (**Figure 4.7A**, center and right two panelss). Another possibility is to express the thioesterase at higher levels the chromosomal thioesterase strain to achieve the levels of octanoic acid produced in **Figure 4.1D**. Finally, this experiment was carried with three plasmids present at all times (when the enzyme was not expressed from plasmid the empty plasmid was used) in order to keep the plasmid load constant. Future work towards reducing the plasmid load may result in higher production of 1-octanol.

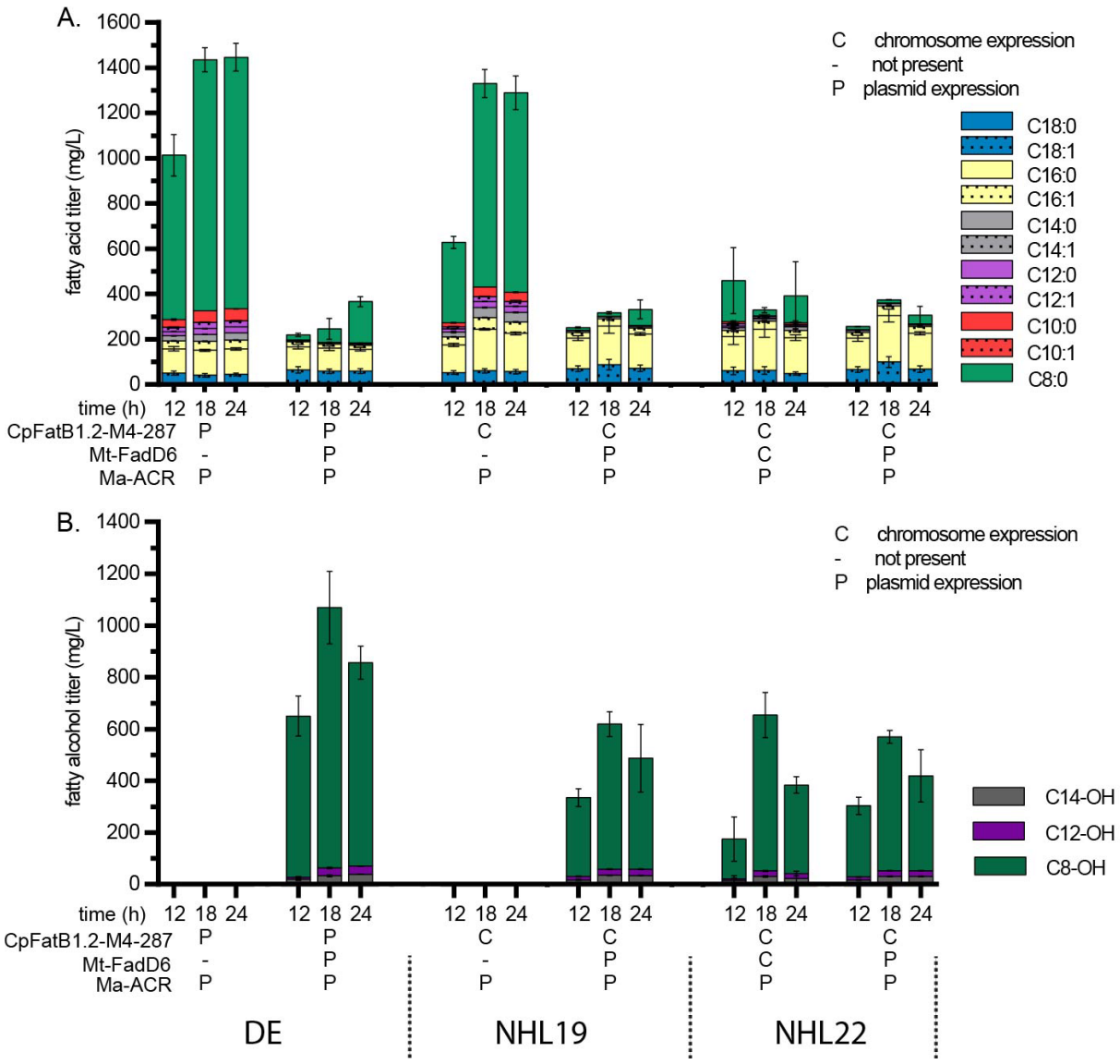


Figure 4.7. Production of 1-octanol from glycerol. (A) Fatty acid accumulation, and (B) fatty alcohol accumulation for strains DE (*E. coli* MG1655 *ΔaraBAD ΔfadD ΔfadE*) (left two panels), NHL19 (K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287 ΔfadE)* (center two panels), and NHL22 (K-12 MG1655 *ΔaraBAD ΔfadD::*PtrcCpFatB1.2-M4-287 *ΔfadE ΔackA-pta::Ptrc-Mt-FadD6*) (right two panels). Cromosomally expressed genes are indicated by a 'C' and plasmid expressed are denoted by a 'P'.

4.4. Conclusions

In this work we showed 1-octanol production from simple sugars through the fatty acid biosynthesis pathway. We employed a highly active thioesterase for the production of octanoic acid developed in Chapter 3 and demonstrated its conversion to octanoyl-CoA via two heterologous acyl-CoA synthethases from *P. putida* and *M. tuberculosis*. We also completed the pathway using a widely used acyl-CoA reductase from *M. aqueaolei* and showed to produce 1 g/L 1-octanol. In addition, we alleviated the challenge that arise when producing volatile compounds by adding an organic overlay and the quantification challenges that such a system poises. Future work will be focused on reducing the strain to fully chromosomally expressed and bioreactor studies to evaluate the performance in a fed-batch context.

4.5. 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. This work was also funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

5. Production of the octanoic acid derived oleochemicals poly(3-hydroxyoctanoic acid) and 2-Heptanone[4,5]

Leveraging on the octanoyl-ACP thioesterase that we developed[154] in Chapter 3, we sought to demonstrate production of other eight carbon oleochemicals. In Chapter 4, I showed how production of 1-octanol can be achieved leveraging on this strategy. In this Chapter I will focus on two other oleochemicals; in section 5.1, I will focus on production of the bio-polymer poly(3-hydroxyoctanoic acid) and in section 5.2, I will demonstrate production of 2-heptanone.

5.1. Production of Poly(3-hydroxyoctanoic acid) from the thioesterase route

5.1.1. Introduction

Poly(3-hydroxyalkanoates) (PHA) are polyesters produced by microorganisms and stored intracellularly as an energy and carbon reserve. PHA are known to be produced at higher quantities when there is an excess of carbon and other nutrients are limited (e.g. N, O_2 or P). PHA can be tuned in their distribution of different chain length to be very diverse in properties and thus potential applications such as the medical industry[161,162], and the plastic industry[163]. Poly(3-hydroxyoctanoic acid)[164] (PHO) is of particular interest

[4] Undergraduate student Ke Xu contributed in the collection of the Poly(3-hydroxyalkanoates) data in this chapter, supervised by Nestor J. Hernandez-Lozada.

[5] Undergraduate student Trevor R. Simmons contributed in the collection of the 2-heptanone data in this chapter, supervised by Nestor J. Hernandez-Lozada.

to the biomedical industry since it is a biocompatible, biodegradable and elastic polymer which can serve as a scaffold for tissue grafts, wound healing and drug delivery applications[161,161,165].

Generally, PHA can be classified as short chain length (C_3-C_5) (scl-PHA) or medium chain (C_6-C_{14}) (mcl-PHA), based on the 3-hydroxyacyl-CoA monomers from which the polyester is synthesized. scl-PHA, such as poly(3-hydroxybutyrate) (PHB) is a brittle and rigid plastic, with T_m of 180 °C whereas mcl-PHA have more elasticity and lower T_m (40-60 °C)[44]. In addition, scl-PHA and mcl-PHA have also different biosynthetic routes in the cell. The most studied scl-PHA, poly(3-hydroxybutyrate) (PHB), is synthesized from 3-hydroxybutyrate. The monomer is made from acetyl-CoA in three enzymetic steps. First, the enzyme β-keto thiolase (PhaA) converts two molecules of acetyl-CoA to acetoacetyl-CoA. Secondly, acetoacetyl-CoA is reduced with a stereo specific acetoacetyl-CoA reductase (PhaB) to (R)-3-hydroxybutyryl-CoA. Finally, the polymer is synthesized by PHA synthase (PhaC)[45]. On the other hand, mcl-PHA are synthesized from one of three routes, (1) transesterification: medium chain 3-hydroxyacyl-ACP from the fatty acid biosynthesis pathway can be trans esterified to (R)-3-hydroxyacyl-CoA by PhaG followed by polymerization by PHA synthase (PhaC); (2) the thioesterase route: medium chain acyl-ACP from the fatty acid biosynthesis pathway can be hydrolyzed to produce medium chain free fatty acid via a thioesterase, followed by reactivation into acyl-CoA by an acyl-CoA synthethase (ACS), then the acyl chain is oxidized by acyl-CoA dehydrogenase (FadE) and hydrated by an stereo specific enoyl-CoA hydratase (PhaJ) into (R)-3-hydroxyacyl-CoA which is finally polymerized with a PHA synthase (PhaC) (**Figure 5.1**). This route also requires other β-oxidation genes, *fadBA* and *fadIJ* to be

knocked out in order to prevent the cell from metabolizing the enoyl-CoA back into acetyl-CoA. The third route follows as (2) from free fatty acids present in the media.

As mentioned at the beginning of this Chapter, in Chapter 3 we engineered a plant acyl-ACP thioesterase (CpFatB1.2-M4-287) with high specificity and increased activity towards C_8 chain lengths for the production of octanoic acid from octanoyl-ACP originated in the fatty acid biosynthesis (FAB) pathway. As mentioned in Chapters 3 and 4 this work opened up the possibility of pursuing compounds downstream of octanoic acid with high interest for the oleochemicals industry. In another work from our laboratory Agnew, et al.[14] used the thioesterase route of mcl-PHA production (**Figure 5.1**) to engineer a pathway to produce C_{12} and C_{14} PHA from sugars leveraging on a thioesterase from *Umbellularia californica* (BTE) with high specificity for dodecanoyl-ACP and some specificity tetradecanoyl-ACP. This, combined with the expression of an ACS from *P. putida* and a specific combination of phaJ and phaC genes from *P. aeruginosa PAO1* yielded 15% DCW of mcl-PHA of C_{12} and C_{14} carbon chain lengths.

In this study we foced on the eight carbon Poly(3-hydroxyoctanoic acid) (PHO) made from the thioesterase route. First, we elucidated the best combination of enoyl-CoA hydratase (phaJ) and PHA synthase (phaC) genes that yielded the highest PHO synthesis from fed octanoic acid. Secondly, we balanced the expression of the intermediate enzymes in the pathway (acyl-CoA synthethase and acyl-CoA dehydrogenase) by overexpressing the Pp_0763 and FadE enzymes under fed octanoic acid. Finally, we introduced a chromosomal copy of our C_8 specific thioesterase[154] carrying into a strain carrying the β-oxidation pathway knockouts necessary to avoid metabolizing our C_8 precursors (*ΔfadBA*, *ΔfadIJ*) as well as a knockout of the fatty acid

degradation regulator to render the chromosomal copy of acyl-CoA dehydrogenase (*fadE*) constitutively expressed. With this strain we pursued production of PHO from glycerol yielding a 15% DCW PHO. This work opens the possibility of generating strains with highly defined eight carbon chain length PHA with potential applications in the biomedical and oleochemicals industries.

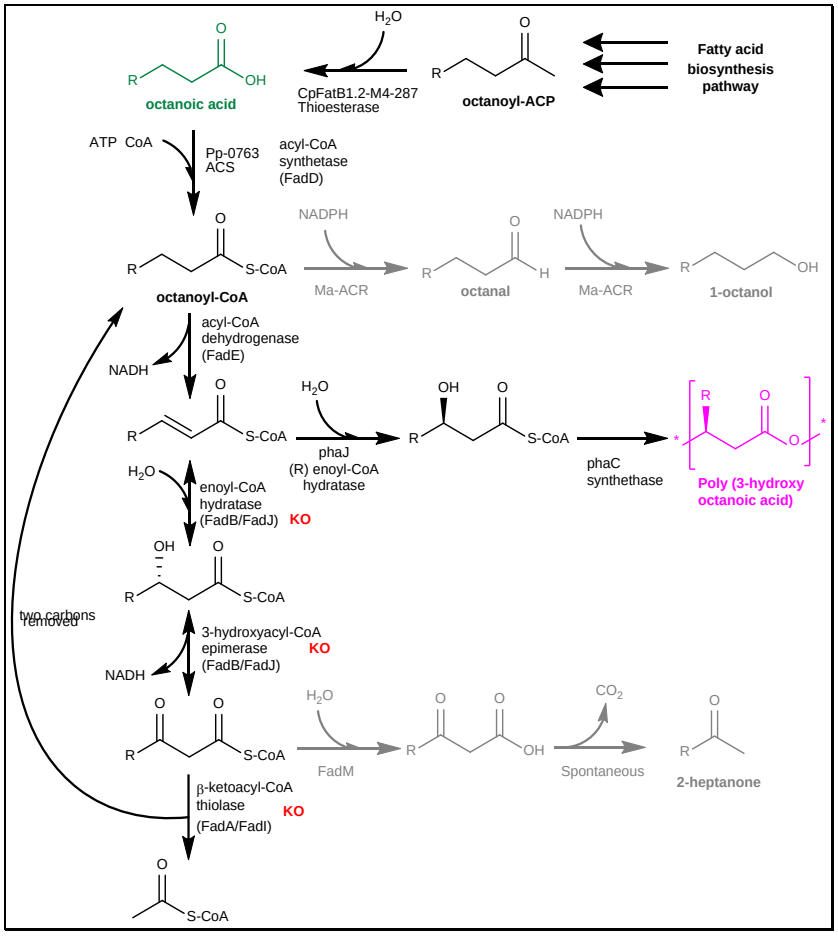


Figure 5.1. Thioesterase route to PHO synthesis. Octanoyl-ACP from the fatty acid biosynthesis pathway can be hydrolyzed to produce octanoic acid via a C_8 specific thioesterase, followed by reactivation into octanoyl-CoA by an acyl-CoA synthethase (ACS), then the acyl chain is oxidized by acyl-CoA dehydrogenase (fadE) and hydrated by an stereo specific enoyl-CoA hydratase (phaJ) into (R)-3-hydroxyacyl-CoA which is finally polymerized with a PHA synthase (phaC). Other β-oxidation genes are knocked out (in red) in order to prevent the cell from metabolizing the enoyl-CoA.

5.1.2. Materials and Methods

5.1.2.1. Strains and Plasmids

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. See **Appendix I-II** for a list of strains and plasmids used in this study.

5.1.2.2. PHA production experiments

For experiments designed to accumulate PHO from exogenously supplied octanoate, single colonies were grown overnight with the appropriate antibiotics depending on the plasmids present (Chloramphenicol 34 µg/mL for pACYC; Carbenicillin 100 µg/mL for pTRC99a; Kanamycin 50 µg/mL for pBTRK) in LB media. Overnight cultures from three single colonies were used to inoculate 50 mL of LB media supplemented with 2 g/L octanoate (2.32 g/L sodium octanoate) and carbenicillin in a 250 mL shake flask at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2-0.3 before inducing with 1 mM IPTG and moving to 30 °C for 48 h post induction.

Cells were then harvested for PHA and FAME analysis. For experiments designed to accumulate PHO from endogenously produced FFA, the procedure follows as above with the media changed to MOPS enriched media described elsewhere[36] with the following changes: 1.39 mM Na_2HPO_4, no biotin, thiamine or sodium selenite added. Cultures were then induced with 1 mM IPTG at an OD_{600} of 0.2-0.3 and incubated at 30°C for 48 hr with shaking. Cells were then harvested for PHA and FAME analysis.

5.1.2.3. Fatty acid extraction, methylation and quantification

Fatty acid samples were prepared by collecting of 2.5 mL culture into a 10 mL glass centrifuge tube. 50 μL of internal standard containing 12.5 mg/mL nonanoic acid and 1.25 mg/mL pentadecanoic acid was added to the samples. The extraction and methylation procedure was done as described in Chapter 2[156]. For GC quantification, samples were analyzed using a GC-FID model Shimadzu GC-2010 equipped with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The column oven temperature protocol was 100°C for 2 min, ramp to 125°C (at 7°C/min), ramp to 235°C (at 10°C/min), ramp to 250°C (at 40°C/min), and hold for 2 min.

5.1.2.4. PHO extraction and quantification

For PHA analysis, cells were harvested by centrifuging 35 mL of culture at 8000 x g for 10 min, washed with 25 mL deionized water, and lyophilized overnight. Biomass concentration was then determined as the CDW of lyophilized cell pellet was weighted.

PHA content as a percentage of CDW was determined by quantifying corresponding 3-hydroxy fatty acid methyl esters. PHA extraction and methylation were performed based on modifications of the method as described previously.[166] 10 mg of lyophilized cells were transfer to a 10 mL disposable glass centrifuge tube followed by the addition of 2 mL of chloroform and 2 mL of 3% H_2SO_4 in methanol (v/v) solution. 50 µL of 12.5 mg/mL nonanoic acid in ethanol was added as an internal standard. The mixture was heated at 100°C in a silicon oil bath for 4 hr before quenched by the addition of 5 mL of 100 mg/mL aqueous $NaHCO_3$. The mixture was vortexed for 3 minutes at max speed and centrifuged at 1000 x g for 10 min. The aqueous layer and cell debris were removed by aspiration and the chloroform layer was analyzed by a Shimadzu GC-2010 GC-FID system equipped with an AOC-20i auto-injector and a Restek 30 m, 0.25 mm ID RTX-5 column. The GC temperature program used was as follows: 60 °C for 1 min, ramp to 230°C (at 10°C/min).

5.1.3. **Results and Discussion**

5.1.3.1. PHA genes screen

In order to elucidate the best route to PHO we first tested combinations of genes from *P. aeruginosa* PAO1 that encode for (1) stereo specific enoyl-CoA hydratase (phaJ1-4) to produce (R)-3-hydroxyoctanoyl-CoA and (2) PHA synthase (phaC1-2). Plasmids were generated containing operons with 3 genes: *phaJ*, *phaC* and ACS from *P. putida Pp_0763* (see **Figure 5.2**). Genes were assembled into one plasmid encoded in an

operon under the inducible expression of the P_{trc} promoter so as to test combinations for *phaJ1-4* and *phaJ1-2* genes while keeping *Pp_0763* unchanged. To test this we transformed these plasmids into the strain of *E. coli* ΔfadRABIJ (K-12 MG1655 Δ*araBAD* Δ*fadR* Δ*fadAB* Δ*fadIJ*)[14]. Cells were inoculated from LB media overnight cultures in triplicates into 50 mL of LB media with 2000 mg/L sodium octanoate (see materials and methods), induced at OD_{600} of 0.2 with 1 mM IPTG and grown at 30 °C for 48 h post induction. Cells transformed with *P. aeruginosa phaJ2-phaC2* were able to consume the most of the octanoate (**Figure 5.3A**). Moreover, *phaJ2-phaC2* yielded the most PHO production with 20% DCW (**Figure 5.3B**). Interestingly, all other combinations, including *phaJ1- phaC2* and *phaJ2-phaC1* were unable to surpass 2.5% PHO suggesting that both genes are driving the specificity towards C_8 acyl chains. In addition, dry cell weight (DCW) in most combinations remained roughly constant while *phaJ2-phaC2* as well as *phaJ2-phaC1* DCW increased by 30% over the empty plasmid carrying strain suggesting that no growth defects were associated with PHO production under these conditions. Taken together, this data strongly suggests moving forward in the study with the *phaJ2-phaC2* combination.

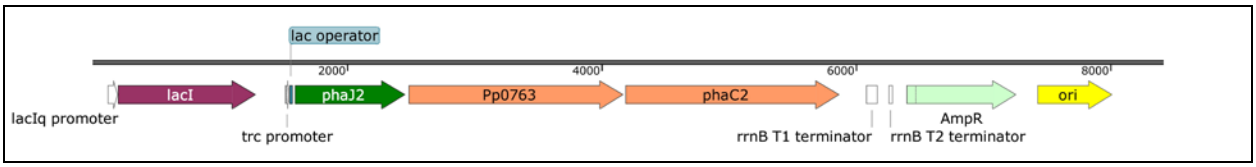


Figure 5.2. Linear map of pKX-J2AC2. Genes from *P. aeruginosa* PAO1 that encode for stereo specific enoyl-CoA hydratase (phaJ2) to produce (R)-3-hydroxyoctanoyl-CoA, PHA synthase (phaC2) as well as the acyl-CoA synthetase from *P. putida* Pp_0763 are encoded in an operon under the inducible expression of the trc promoter. Similar combinations for phaJ1-4 and phaJ1-2 were made while keeping Pp_0763 unchanged.

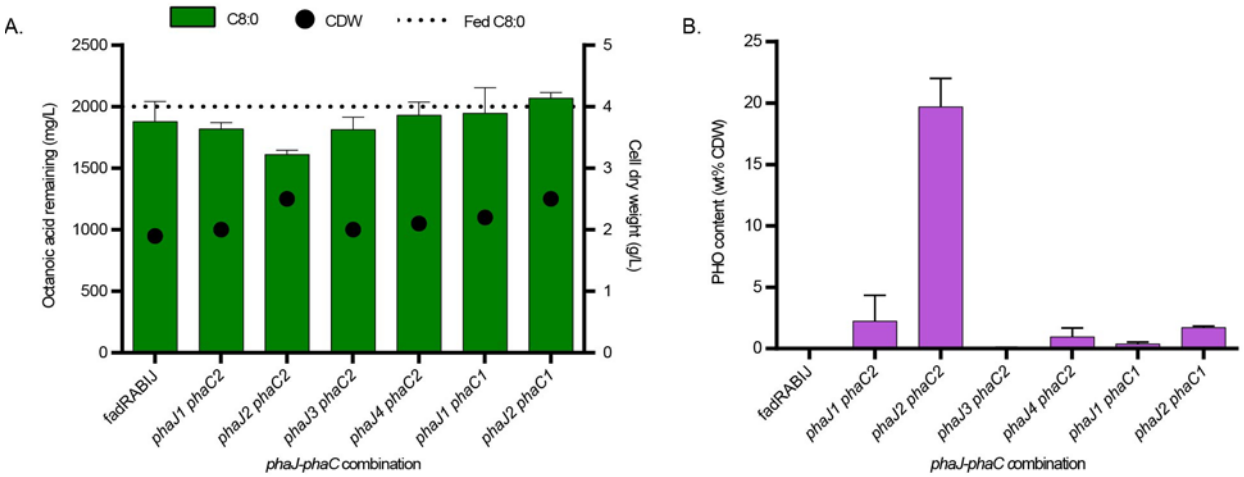


Figure 5.3.Testing combinations of *P. aeruginosa phaJ and phaC* genes. (A) Cells carrying *P. aeruginosa phaJ2-phaC2* were able to consume the most of the octanoate. (B) *phaJ2-phaC2* also yielded the most PHO production with 20% DCW. All other combinations, including *phaJ1-phaC2* and *phaJ2-phaC1* were unable to surpass 2.5% PHO suggesting that both genes are driving the specificity towards C_8 acyl chains.

5.1.3.2. Balancing expression of FadE and Pp_0763

We were interested in exploring the existence expression imbalances in the pathway since acyl-CoA synthethases are often specific for long chain fatty acids[155] and the acyl-CoA dehydrogenase we used was not altered from *E. coli* genome (albeit deregulated through *ΔfadR* deletion). In order to do this, we introduced two more plasmids to our system carrying an extra copy of Pp_0763 ACS under a pACYC plasmid (see Appendix II) and FadE under a pBTRK plasmid. This way, at any point there where 3 plasmids present: pKX-J2AC2 (see **Figure 5.2**), pACYC (either empty or with Pp_0763) and pBTRK (either empty or with fadE). Cells were grown on LB media with 2000 mg/L sodium octanoate following the same procedure as **Figure 5.3** (see section 5.1.3.1 and

materials and methods)**.** As can be seen in **Figure 5.4A**, octanoate consumption upon addition of additional Pp--0763, both by itself and with additional fadE overexpression, decreased and cell dry weight dropped by at least 50%, showing a deleterious effect of overexpression of Pp_0763. Moreover, PHA content decreased by 33% when only additional Pp_0763 overexpression was done and was abolished when overexpressing both Pp_0763 and fadE at the same time. On the other hand, overexpression of fadE alone increased the PHO content by 83% over the strain carrying the pKX-J2AC2 alone with pBTRK and pACYC empty plasmids yielding a total PHO content of 32% DCW (**Figure 5.4B**). For this case of FadE overexpression only, the DCW dropped 35% showing some burden on the cell growth (**Figure 5.4A**). This data points at chromosomal FadE expression as being imbalanced and the need to express it at higher levels.

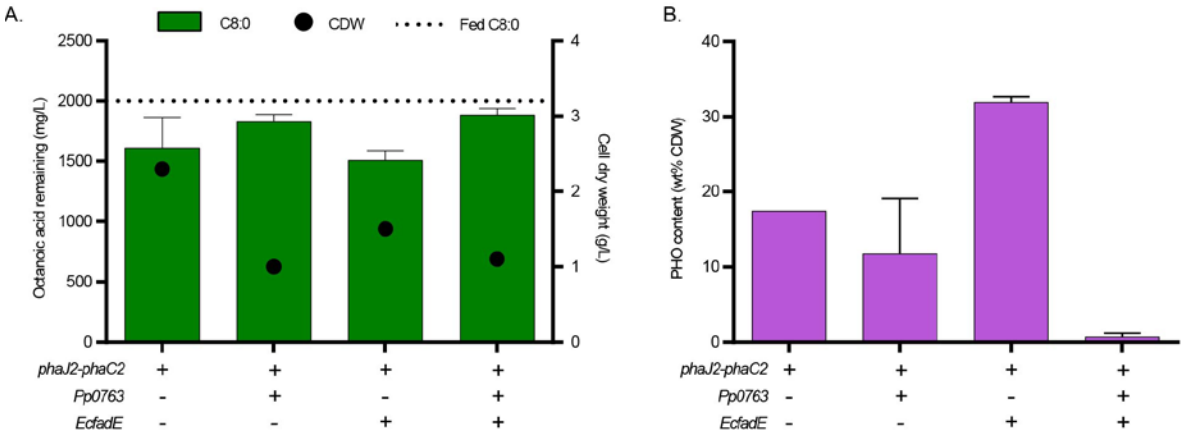


Figure 5.4. Balancing expression of Pp_0763 and fadE. (A) Octanoate consumption upon addition of additional Pp_0763, both by itself and with additional fadE overexpression, decreased and cell dry weight dropped by at least 50%, showing a deleterious effect of overexpression of Pp_0763. (B) PHA content decreased by 33% when only additional Pp_0763 overexpression was done and was abolished when overexpressing both Pp_0763 and fadE at the same time. Overexpression of fadE alone increased the PHO content by 83% over the strain carrying the pKX-J2AC2 alone with pBTRK and pACYC empty plasmids yielding a total PHO content of 32% DCW.

5.1.3.3. Production of PHO from sugars

We decided to explore the potential of producing octanoic acid endogenously using our thioesterase from Chapter 3 and its subsequent conversion to PHO (See **Figure 5.1**). To do this, we constructed a strain carrying a chromosomal copy of CpFatB1.2-M4-287 thioesterase on the *fadD* loci and the inducible P_{trc} promoter: NHL18 (*E. coli* K-12 MG1655 Δ*araBAD* Δ*fadR* Δ*fadAB* Δ*fadIJ* Δ*fadD::Ptrc-CpFatB1.2-M4-287).* We transformed pKX-J2AC2 (or the empy plasmid) and with pBTRK-fadE (or empty) into NHL18 strain carrying the C_8 thioesterase chromosomal copy. Cells were inoculated from LB media overnight cultures in triplicates into 50 mL of MOPS media enriched[36] with tryptone and yeast extract (see methods section), induced at OD_{600} of 0.2 with 50 μM IPTG and grown at 30 °C for 48 h post induction. Cells carrying both empty plasmids were able to accumulate 1.3 g/L octanoic acid from the chromosomal copy of CpFatB1.2-M4-287 thioesterase but were not able to transform it to PHO as expected (**Figure 5.5A**) while cells carrying pKX-J2AC2 with pBTRK-fadE or the empty pBTRK plasmid were able to completely transform all the octanoic acid made. This suggests that under these conditions there was enough chromosomally expressed fadE to not be a limiting reaction (**Figure 5.5A**). We also did not see a significant growth defect upon expressing *fadE* in the plasmid (**Figure 5.5A**), an effect that we saw under the fed octanoic acid experiments (**Figure 5.4A**). In terms of PHO, we saw accumulations of 15% DCW PHO in cells carrying pKX-J2AC2 with the empty pBTRK and no significant difference when pBTRK-fadE was present (**Figure 5.4B**). This suggest that at these conditions octanoic acid supply is the limiting factor.

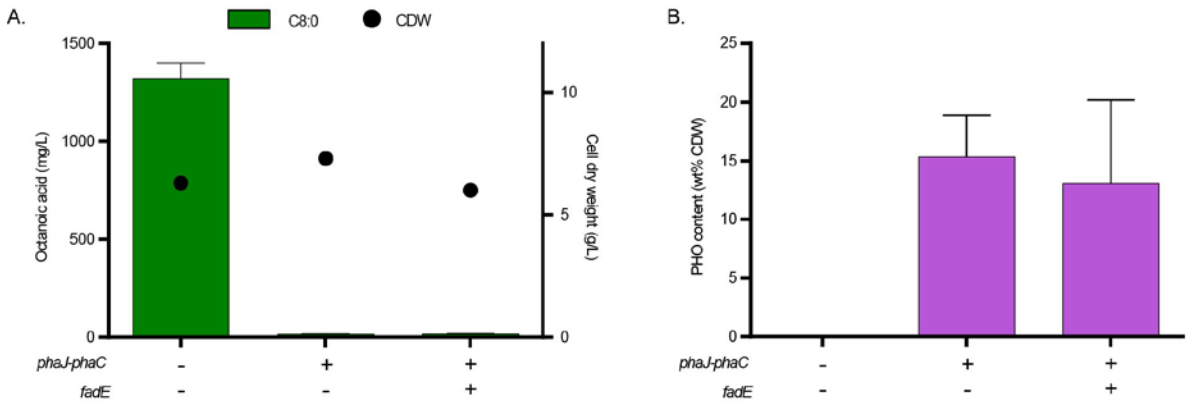


Figure 5.5. PHO production from endogenously made octanoic acid.

5.1.4. **Conclusions**

In this work we demonstrated PHO production of 15% dry cell weight from an unrelated carbon source. We did this by first elucidating the combination of enoyl-CoA hydratase (phaJ2) and PHA synthase (phaC2) genes as one that results in the highest PHO content from fed octanoate. Then explored expression imbalances and found that the acyl-CoA dehydrogenase expression from the chromosome was not sufficient for maximum flux through the pathway. Finally, we introduced a chromosomal copy of a highly active acyl-ACP thioesterase developed in Chapter 3 and showed that PHO production can be achieved through this route from unrelated carbon sources. Future work will focus on optimizing this pathway for higher octanoic acid production, exploring additional variables such as the introduction of chaperones that stabilize the PHO aggregates, moving the full pathway to the chromosome and fermentation in fed-batch cultures.

5.2. Production of 2-Heptanone from the thioesterase route

5.2.1. Introduction

2-heptanone is a methyl ketone with a cheese-like distinctive aroma and flavor. In particular, 2-heptanone is commonly present in blue-veined cheese at concentrations ranging 17-70 mg/kg cheese and it is known as a key contributor to its flavor[167]. Microbial production of 2-heptanone could serve as a process to make blue cheese flavor concentrates for the food industry[50]. Another application that it is proposed for methyl ketones is their use in the biofuel industry[47], as they possess properties similar to gasoline (C_5-C_{10}) and diesel (C_{11}-C_{15}) depending upon the chain length.

Leveraging on the octanoyl-ACP thioesterase that we developed[154] we pursued 2-heptanone production via octanoyl-ACP hydrolysis into octanoic acid followed by its reactivation to octanoyl-CoA via Pp_0763 acyl-CoA synthetase (see Chapter 4) and using the *Escherichia coli* native acyl-CoA dehydrogenase (FadE), enoyl-CoA hydratase/epimerase (FadB) to produce β-ketooctanoyl-CoA. Finally, 2-heptanone can be synthesized from the spontaneous decarboxylation of β-ketooctanoic acid which can be derived from the hydrolysis of β-ketooctanoyl-CoA species via a thioesterase reaction (**Figure 5.6**).

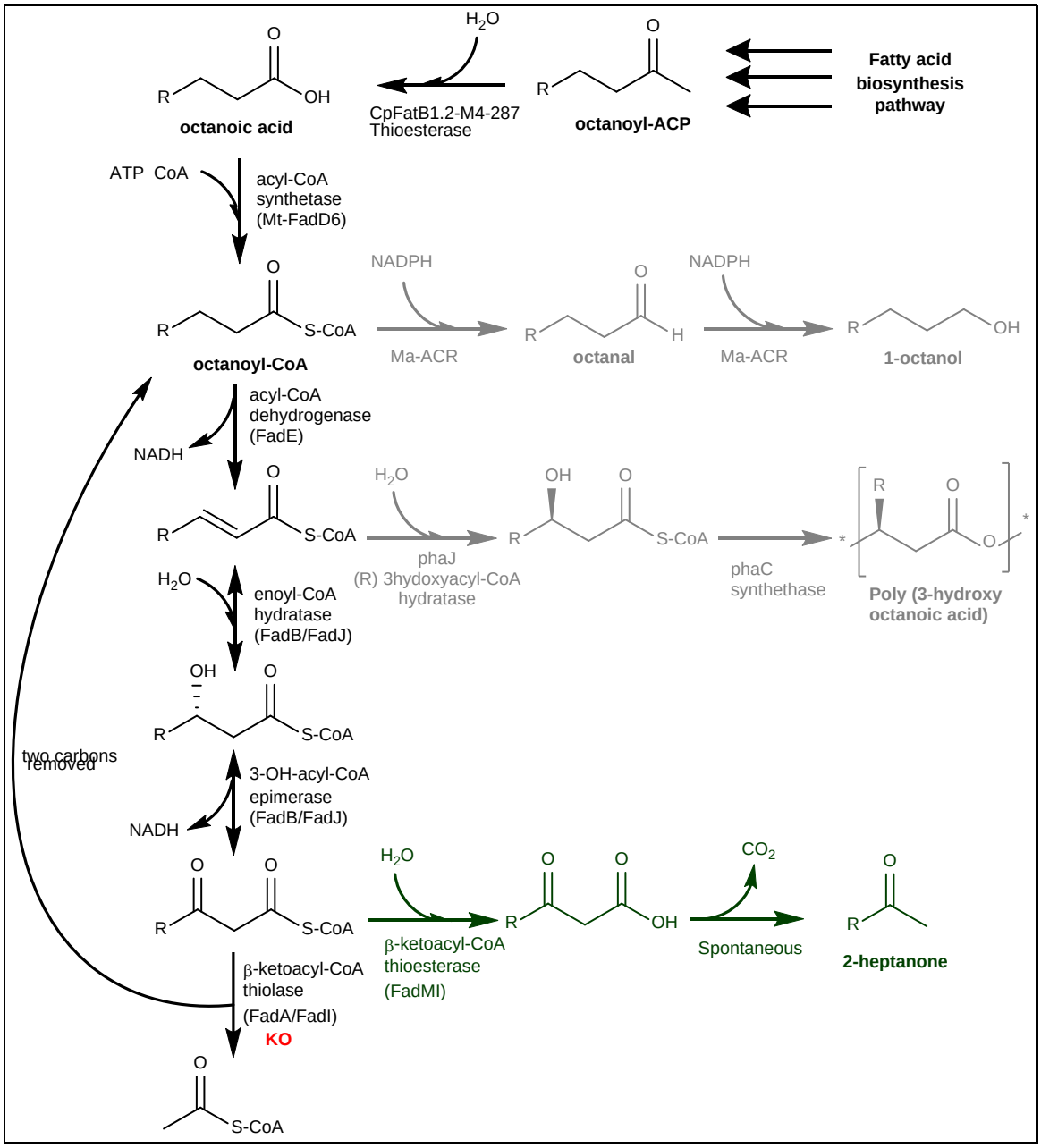


Figure 5.6. Biosynthesis of 2-Heptanone from the thioesterase route. Octanoyl-ACP from the fatty acid biosynthesis pathway can be hydrolyzed to produce octanoic acid via a C_8 specific thioesterase, followed by reactivation into octanoyl-CoA by an acyl-CoA synthethase (ACS), then the acyl chain is oxidized by acyl-CoA dehydrogenase (fadE) followed by enoyl-CoA hydratase/epimerase (FadB) to produce β-ketooctanoyl-CoA. Finally, 2-heptanone can be synthesized from the spontaneous decarboxylation of β-ketooctanoic acid which can be derived from the hydrolysis of β-ketooctanoyl-CoA species via a thioesterase reaction.

5.2.2. Materials and methods

5.2.2.1. Chemicals, Strains and plasmids

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. See **Appendix I-II** for a list of strains and plasmids used in this study.

5.2.2.2. Fatty acid and methyl ketone experiments

For experiments designed to produce and/or consume free fatty acids and methyl ketones (**Figure 5.8** and **Figure 5.10**) single colonies were grown overnight with the appropriate antibiotics depending on the plasmids present (Carbenicillin 100 μg/mL for pTRC99a; Kanamycin 50 μg/mL for pBTRK) in LB media. Overnight cultures from three single colonies were used to inoculate 50 mL of MOPS enriched media described elsewhere[36] (with the following changes: 1.39 mM Na_2HPO_4, no biotin, thiamine or sodium selenite added) and 10mL dodecane as an organic layer in a 250 mL shake flask

at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2 and at that point induced with 1mM IPTG and moved to 30 °C for 24 h post induction.

For experiments used to test homologs of *E. coli fadM* (**Figure 5.10**) overnight cultures from three single colonies were used to inoculate 50 mL of LB media supplied with 1g/L octanoic acid in a 250 mL shake flask at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2 and at that point induced with 1mM IPTG and moved to 30 °C for 24 h post induction.

5.2.2.3. Fatty acid extraction and quantification

Fatty acid extractions followed with the same procedure described in section 4.2.3. of this thesis.

5.2.2.4. Methyl ketone quantification

Methyl ketone samples were prepared by collecting the equivalent of 2.5 mL culture (same as free fatty extraction section) into a 10 mL glass centrifuge tube. When cultures were in the presence of dodecane (**Figure 5.8**) shake flask were constantly mixed right before the sample is taken and the sample was taken immediately after the shake flask stops mixing. After collecting the samples, 50 µL of an internal standard containing 12.5 mg/mL 2-octanone, and 12.5 mg/mL 2-nonanone was added to the samples. Samples were then extracted into 1 mL hexane by vortexing the solution for 3 min at max speed and centrifuged at 3000 g for 10 min. The hexane layer was injected in a GC-FID. Samples were analyzed using a GC-FID model Shimadzu GC-2010 equipped

with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The column oven temperature protocol was 100°C for 2 min, ramp to 125°C (at 7°C/min), ramp to 235°C (at 10°C/min), ramp to 250°C (at 40°C/min), and hold for 2 min.

5.2.2.5. Homologs of *E. coli* FadM thioesterase

Sequence of the homologs of *E. coli* FadM thioesterase were obtained from the Thyme thioesterase database from Ohio state university[26,168]. Sequences were compared using NCBI protein Blast tool[169] and codon optimized and purchased using the IDT technologies codon optimizing tool. Homologs of Ec-FadM where cloned into the same pTRC99a plasmid originally carrying Ec-FadM (**Figure 5.7**) under the guide of the same promoter and ribosome binding site.

5.2.3. Results and discussion

5.2.3.1. 2-Heptanone synthesis demonstration from sugars

In Chapter 3 (section 3.3.5) we showed that a high production of octanoic acid can be achieved from a single chromosomal copy of the engineered CpFatB1.2-M4-287 thioesterase from *Cuphea palustris*[154]. In addition, we showed in Chapter 4 (section 4.3.2.) that the acyl-CoA synthethase from *Mycobacterium tuberculosis* (Mt-FadD6) has high activity for octanoic acid to produce octanoyl-CoA. Given this, we set to demonstrate production of 2-heptanone from simple sugars. We first created strains TRS10 (K-12 MG1655 *ΔaraBAD ΔfadD::trcCpFatB1.2-M4-287 ΔfadA ΔfadR*) and TRS11 (K-12

MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287 ΔfadA ΔfadR ΔfadE*). The strains have deletions of *fadA* which blocks thiolase activity to degrade β-ketooctanoyl-CoA to hexanoyl-CoA and acetyl-CoA (see **Figure 5.6)** as well as *fadR* deletion which deregulates expression of the pathway genes *fadE* and *fadB* when chromosomally present. In TRS11, and additional *fadE* deletion was done to compare chromosomal vs plasmid expression between TRS10 and TRS11 since in the case of PHO production (section 5.1.3.2) we had found *fadE* to be underexpressed from the chromosome. Finally, both strains have chromosomal insertions of the *CpFatB1.2-M4-287* thioesterase developed in Chapter 3 to enable octanoic acid production endogenously.

In order to test 2-heptanone production strains TRS10 and TRS11 were transformed with a pTRC99a plasmid (see Appendix II) carrying the acyl-CoA synthethase from *Mycobacterium tuberculosis* (Mt-FadD6) as well as the β-ketoacyl-CoA thioesterase from *E. coli* (Ec-FadM) both under separate promoters (**Figure 5.7**). TRS11 was also transformed with a pBTRK-Ec-FadE plasmid to allow acyl-CoA dehydrogenase activity from plasmid rather than the chromosome as per TRS10. Cells were inoculated from LB media overnight cultures in triplicates into 50 mL of MOPS enriched media + 10mL dodecane as an organic layer (see materials and methods), induced at OD_{600} of 0.2 with 1 mM IPTG and grown at 30 °C for 24 h post induction. Fatty acid data after 24 h post induction (**Figure 5.8A**) shows 0.40 g/L octanoic acid accumulated in the strain TRS10 which has only chromosomal expression of FadE compared to 0.12 g/L for TRS11 suggesting, as in the PHO case (see section 5.1.3.2) that FadE is also likely underexpressed in this pathway. In addition, NHL11 strain also accumulated 10-fold more 2-heptanone (22 mg/L) compared to 2.4 mg/L for NHL10.

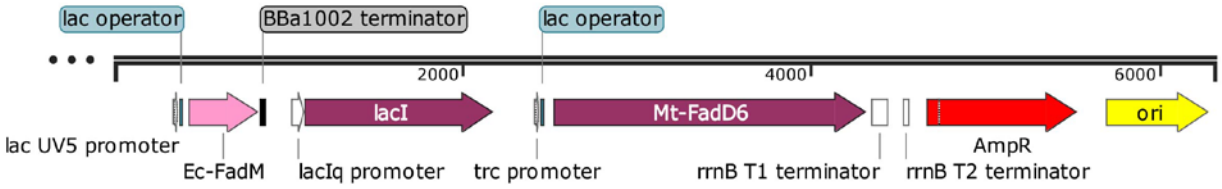


Figure 5.7. Plasmid map of β-ketoacyl-CoA thioesterase and Acyl-CoA synthethase.

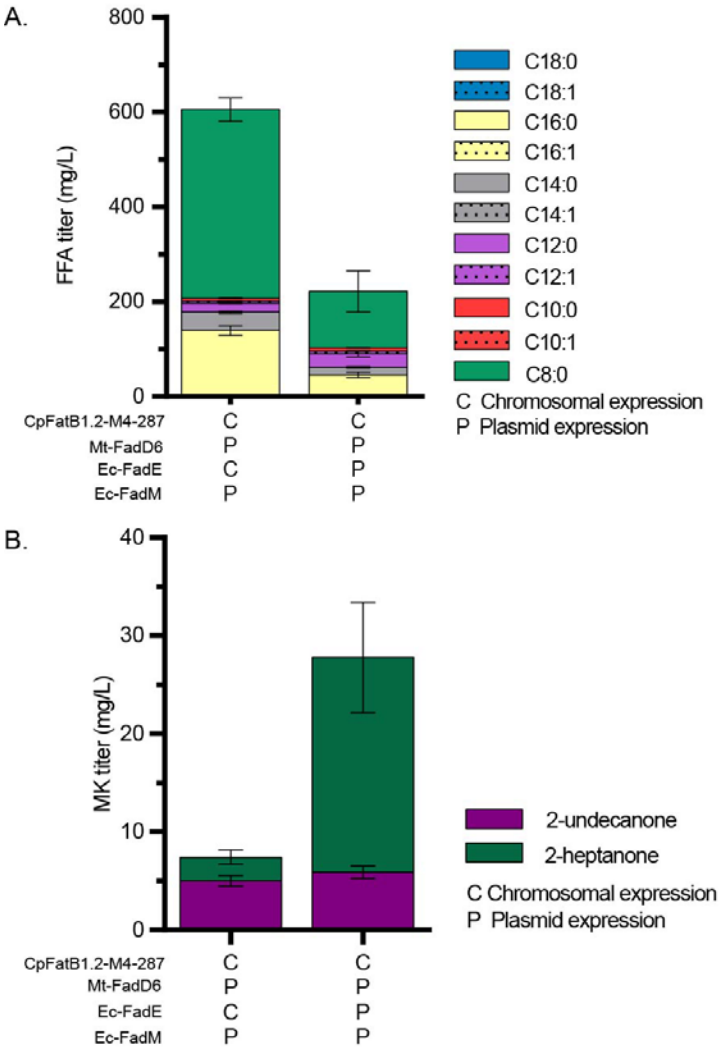


Figure 5.8. Production of 2-Heptanone from the thioesterase route. (A) Free fatty acid accumulation at 24 h on strain TRS10 (left) carrying chromosomal copies of CpFatB1.2-M4-287 and Ec-fadE as well as the plasmid with Mt-Fad6 and Ec-FadM (**Figure 5.7**), and TRS11 (right) carrying chromosomal copy CpFatB1.2-M4-287 and plasmid expression of Mt-Fad6 and Ec-

FadM and Ec-fadE. (B) methyl ketone accumulation at 24 h corresponding to the same cultures as part (A).

There are various considerations to be taken in order to evaluate the reasons for the observed low titers of 2-heptanone compared to the octanoic acid accumulated in the strains. One potential reason is that FadI, another thiolase typically active in anaerobic conditions but known to be somewhat active in aerobic conditions as well is still present on both TRS10 and TRS11 strains. This could be degrading a fraction of the β-ketooctanoyl-CoA pool into hexanoyl-CoA and hampering the accumulation of 2-heptanone. Another potential reason that 2-heptanone is known to be very volatile compound in aqueous solution[51]. Although we added dodecane to our cultures to avoid large evaporation effects, the efficacy of this method before this experiment had only been tested on the case of 1-octanol. Future work should be focused on both characterizing 2-heptanone evaporation and deleting *fadI* chromosomal copy. A third potential reason for the low titers obtained for 2-heptanone, the inherent activity of fadM toward β-ketooctanoyl-CoA, is the focus of next section.

5.2.3.2. Exploring β-ketoacyl-CoA thioesterase diversity

As mentioned above, a possibility for the low titers obtained for 2-heptanone is that it has been shown in vitro that fadM has most activity on β-ketotetradecanoyl-CoA (C_{14}) chains lengths[170]. We compiled a focused library of FadM homologs based on the Thyme database sequences available[26,168]. The main goal of this was to find homologs with different degrees of protein sequence similarity to the *E. coli* fadM to test some biological diversity. As can be seen in **Figure 5.9**, a pairwise comparison of our library of

thioesterases reveal sequence similarities ranging from 30-95% of the protein sequence of the *E. coli* fadM. We also ensured that none of our proteins were similar to each other. Interestingly, we noticed that *Mycobacterium tuberculosis*, the same organism that contained highly active acyl-CoA synthetase for octanoic acid has 5 diferent fadM-like genes (**Figure 5.9B**). We hypothesized that there could be different roles for these fadM in the cell and decided to test them as a sub-library towards finding a strong β-ketooctanoyl-CoA thioesterase.

A.

		EcFadM	SeFadM	NgFadM	PsFadM	MuFadM	SoFadM	CsFadM	BbFadM	BaFadM	CtFadM	PaFadM
Escherichia coli	EcFadM	100	95	41	48	39	76	91	82	82	87	30
Salmonella enterica	SeFadM	95	100	45	50	45	77	93	80	79	89	38
Neisseria gonorrhoeae	NgFadM	41	45	100	36	40	41	43	43	41	41	26
Providencia sneebia	PsFadM	48	50	36	100	36	49	48	44	46	51	30
Marinomonas ushuaiensis	MuFadM	39	45	40	36	100	39	42	38	38	41	30
Serratia odorifera	SoFadM	76	77	41	49	39	100	77	80	77	76	31
Cucumis sativus	CsFadM	91	93	43	48	42	77	100	82	82	88	32
Beauveria bassiana	BbFadM	82	80	43	44	38	80	82	100	94	78	32
Buttiauxella agrestis	BaFadM	82	79	41	46	38	77	82	94	100	80	34
Cronobacter turicensis	CtFadM	87	89	41	51	41	76	88	78	80	100	31
Pseudomonas aeruginosa	PaFadM	30	38	26	30	30	31	32	32	34	31	100

B.

		E. coli	*MtFadM1*	*MtFadM2*	*MtFadM3*	*MtFadM4*	*MtFadM5*
Escherichia coli	*E. coli*	100	61	37	31	30	31
Mycobacterium tuberculosis	*MtFadM1*	58	100	25	26	44	44
	MtFadM2	31	25	100	25	24	36
	MtFadM3	26	26	25	100	74	70
	MtFadM4	44	44	24	74	100	84
	MtFadM5	44	44	36	70	84	100

Figure 5.9. Protein sequence comparison among fadM homologs.

In order to evaluate these Ec-FadM homologs, we decided to track the consumption of exogenously added octanoic acid and track all the intermediates between

octanoic acid and 2-heptanone. Since all these intermediates can be quantified, and *E. coli* β-oxidation pathway is well understood we hypothesized that this method can reliably be used to quantify the turnover of octanoic acid to 2-heptanone. This approach also allows to see potential activities from these homologs toward other intermediates such as 3-hydroxyoctanoyl-CoA or trans-3-octenoyl-CoA.

In order to test this, we used strain RAI (K-12 MG1655 *ΔaraBAD ΔfadR ΔfadA ΔfadI*) from earlier work in our lab by Agnew, et al.[14]. The *fadM* homologs were cloned in the same pTRC99a plasmid as previously discussed (**see Appendix II**, **Figure 5.7**) carrying the acyl-CoA synthethase from *Mycobacterium tuberculosis* (*Mt-FadD6*) except now replacing Ec-FadM with the homolog β-ketoacyl-CoA thioesterases (**Figure 5.9**). Strains RAI was transformed with a this pTRC99a plasmid as well as pBTRK-Ec-FadE to avoid chromosomal FadE underexpression issues found in section 5.1. Cells were inoculated from LB media overnight cultures in triplicates into 50 mL of LB media + 1 g/L sodium octanoate (see materials and methods), induced at OD_{600} of 0.2 with 1 mM IPTG and grown at 30 °C for 24 h post induction while taking samples over time to track consumption of metabolites.

Fatty acid data over time (**Figure 5.10A**) shows gram-negative bacteria *Providencia sneebia* Ps-FadM consumed 0.86 g/L octanoic acid by 8 h and entomopathogenic fungi *Beauveria bassiana* Bb-FadM consumed 0.71 g/L octanoic while *E. coli* Ec-FadM had consumed 0.63 g/L octanoic acid by the same time. Meanwhile, intermediate 3-hydroxyoctanoyl-CoA remained at low levels (< 0.035 g/L) across all samples (**Figure 5.10B**) and times and trans-3-octenoyl-CoA was not detected in any sample (not shown). β-ketooctanoyl-CoA accumulated at 0.07-0.09 g/L after 18 h in Ec-

FadM carrying strain and 0.05-0.07 g/L for Bb-FadM whereas for Ps-FadM it accumulated at 0.007-0.013 g/L. Overall, this data suggests that Ps-FadM has the most activity on β-ketooctanoyl-CoA. However, the differences among the variants tested are not drastic. It should be noted that this test was done at high protein expression and that these differences might become sharper when testing low expression of FadM. Future work will be done to test this hypothesis.

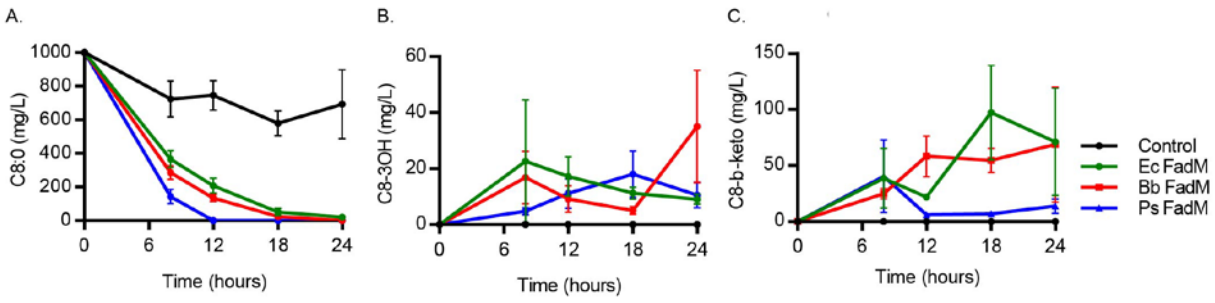


Figure 5.10. Comparison of among fadM homologs from the thioesterase route towards 2-Heptanone. Strain RAI (K-12 MG1655 *ΔaraBAD ΔfadR ΔfadA ΔfadI*) was transformed with plasmids containing *Mt-FadD6,* Ec-FadE and various FadM. Cells were cultured in 50 mL of LB media + 1 g/L sodium octanoate. Consumtion of octanoic acid (A) and accumulation of intermediates 3-hydroxyoctanoyl-CoA (B) and β-ketooctanoyl-CoA (C) was tracked over time.

5.2.4. Conclusions and future directions

In this work we demonstrated production of 2-heptanone from simple sugars. We did this by first creating strains with a chromosomal copy of the thioesterase CpFatB1.2-M4-287 developed in Chapter 3, deregulation of the β-oxidation pathway through a *fadR* deletion, and the deletion of FabA thiolase to avoid β-ketoacyl-CoA side reactions. These strains,

were then transformed with plasmids containing acyl-CoA synthethase and β-ketoacyl-CoA thioesterase to demonstrate production of 2-heptanone at 2.4 mg/L. We also found that overexpression of fadE results in 10-fold increase to 22 mg/L 2-heptanone. Some considerations such as the presence of another thiolase in our strains (*fadI*), and 2-heptanone volatility may be behind the low titers observed. We also evaluated various Ec-FadM homologs and tracked the consumption of exogenously added octanoic acid and all the intermediates between octanoic acid and 2-heptanone in order to find higher activity mutants. Overall, the data suggests that Ps-FadM has the most activity on β-ketooctanoyl-CoA among the homologs tested. These differences might become important when testing systems in which low expression of FadM is required such as chromosomally integrated *fadM*. Future work will be focused on characterizing more FadM homologs in low expression, addressing evaporation issues and reducing the strain to fully chromosomally expressed.

5.3. Acknowledgements

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy. N.H.L. is the recipient of a NIH Chemistry-Biology Interface Training Program fellowship (No. T32 GM008505) and the Graduate Engineering Research Scholars fellowship from the UW−Madison College of Engineering.

6. Other studies[6]

6.1. Dynamic degradation of *E. coli* FabA to control the membrane phospholipid composition

6.1.1. Introduction

In addition to producing C_8 oleochemicals, which has been the subject of most of this thesis, one of the overarching goals in our laboratory has been to make C_{12} oleochemicals through the thioesterase route. When a C_{12} specific thioesterase is expressed in *E. coli* to overproduce mostly saturated C_{12} free fatty acids (FFA), Lennen et al.[148] found that it leads to growth defects and that some of these defects can be related to impaired membrane homeostasis[148,171]. For example, a SYTOX green assay was done by Lennen et al.[148] that shows that the membrane becomes more permeable when a thioesterase is actively expressed. Moreover, gene transcription data suggests that the consequences of this problem on the membrane could be multiple such as inappropriate unsaturated free fatty acid regulation, futile cycles in the electron transport in the membrane, and toxicity of the FFA due to inability to export the products fast enough[148,172]. From this, a hypothesis derives that, if the mechanisms by which the membrane of *E. coli* is being affected are elucidated and fixed, higher thioesterase induction could be pursued resulting in a higher FFA yield and healthier cells. However, efforts to correct these problems, although somewhat positive for cell health, have not necessarily translated to higher FFA yields[140,173]. Therefore, the overall goal of this

[6] This work was done in collaboration with Dr. Mark Politz

section of the project is to find a way to restore the cell membrane composition to a degree that lead to increased free fatty acid yields.

When *E. coli* is forced to overproduce C_{12} FFA Lennen et al.[148] found that it leads to a change in the membrane composition of the cell, specifically, the saturated:unsaturated phospholipids ratio in the cell is decreased from 2:1 in healthy cells to 1:1 in overproducing cells. Transcription studies have identified wild-type unsaturated fatty acid biosynthesis related genes to be down-regulated during this context of production. However, when quantified, the unsaturated fatty acid levels are still too high relative to the saturated; suggesting that further suppression of these genes is required to ensure restauration of saturated to unsaturated fatty acid balance in the membrane. The two main enzymes that control the unsaturated fatty acid biosynthesis, FabA and FabB, are responsible for initiation and elongation of unsaturated fatty acids on the cell, respectively (see **Figure 6.1 & Figure 1.3**). Specifically, FabA enzyme isomerizes trans-2-decenoyl-ACP to the cis form (cis-3-decenoyl-ACP) and FabB does the subsequent elongation of the unsaturated acyl-ACP produced. When a C_{12} thioesterase (BTE) is expressed, the saturated fraction of acyl-ACP pools is depleted in order to produce C_{12} FFA rendering it unavailable to make as much saturated phospholipids and thus depleting the amount of saturated phospholipids. The cell responds to this pressure by downregulating *fabA* and *fabB* genes through FabR[23] which is a transcriptional regulator of both *fabA* and *fabB* genes, resulting in less production of unsaturated acyl-ACP. Despite of this natural response of the cell, the ratio of saturated:unsaturated phospholipids in the cell is remains unbalanced. Efforts to solve the unsaturated FFA issue had partially positive results. Lennen et. al.[140], were able to partially restore the

unsaturated to saturated ratio by overexpressing FabR on a plasmid but it lead to severe growth defects. They also co-expressed BTE with a second thioesterase that has more specificity towards unsaturated acyl-ACP, which resulted in increased viability, partially restored unsaturated FFA content, but did not increase the yields. Here, we adopted a dynamic protein degradation approach to degrade some or all of the FabA protein and aid FabR natural system of *E. coli* to decrease the total protein content of FabA. This way, we hypothesized that FabR will be able to restore the membrane composition of the cell resulting healthier producing cells.

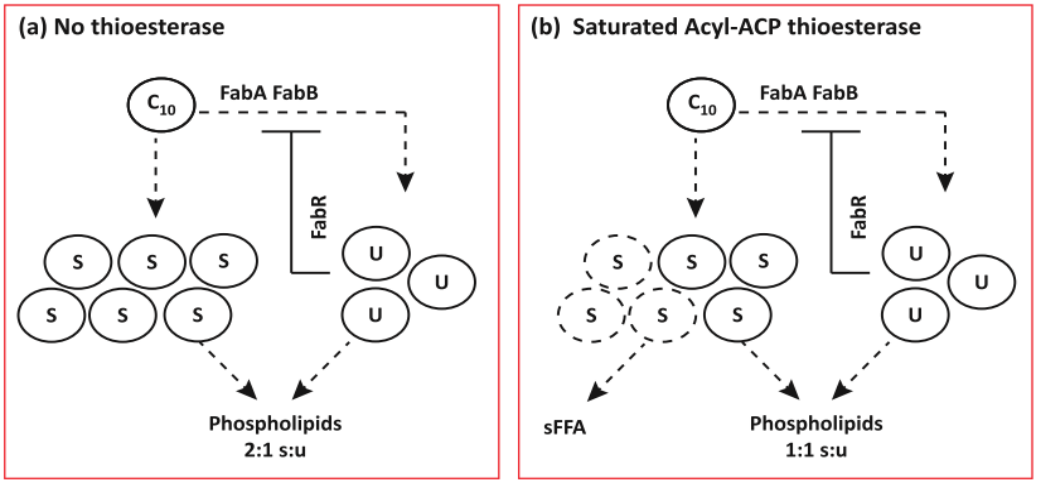


Figure 6.1. Effect of expressing a C12:0 thioesterase on the membrane lipid composition. (a) Native *E. coli* saturated:unsaturated ratio is 2:1. (b) When a thioesterase specific for saturated FFA is expressed the saturated:unsaturated ratio is affected due to the depletion of the saturated pool and the inability of the cell to downregulate *fabA* and *fabB* expression strongly enough. Figure adapted from Lennen, R. M., & Pfleger, B. F. (2012). *Trends in Biotechnology*, *30*(12), 659–67.

6.1.2. Materials and methods

6.1.2.1. Chemicals

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). See **Appendix I-II** for a list of strains and plasmids used in this study.

6.1.2.2. DNA synthesis

Strains NHL02 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3*), NHL03 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3B*) and NHL04 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3C*) were created from RLO8ara (*E. coli* K12 MG1655 Δ*araBAD ΔfadD*) via CRISPR-Cas9 assisted homologous recombination protocol, modified from Li et al.,[137] in by using a linear piece of dsDNA containing a flag tag, a GSG linker and the degradation tag (pdt#3, pdt#3B and pdt#3C)[174] between 40 base pairs of homology flanking the c-terminal of *fabA gene* and downstream of *fadA* sequence (see **Figure 6.2**).

All plasmids made were constructed using Gibson Assembly of PCR products[139]. Mf-Lon protease was cloned out from Addgene plasmid #21867 (pBAD33-mf-lon) into a pBTRK (Kan^R, pBBR1 origin) under the inducible P_{tet} promoter to create plasmid pMP009 (see **Appendix II**).

6.1.2.3. Fatty acid experiments

Single colonies were grown overnight with the appropriate antibiotics depending on the plasmids present (Kanamycin 50 μg/mL for pMP009-mf-lon; Carbenicillin 100 μg/mL for pTRC99a-BTE) in LB media. 50 μL of LB media overnights were used to inoculate 5 mL MOPS minimal media phosphate limited as described elsewhere[142] and grown overnight. Overnight cultures from three single colonies in minimal media were used to inoculate 50 mL of media in a 250 mL shake flask at an initial OD_{600} = 0.05. Cells were allowed to grow at 37 °C to OD_{600} = 0.2 before inducing with 50 μM IPTG and 20 ng/mL ATC (where applies) allowing them to grow for 48 h post induction.

6.1.2.4. Lipid extraction and methylation

Fatty acid samples were prepared by collecting of 2.5 mL culture into a 10 mL glass centrifuge tube. 50 μL of internal standard containing 12.5 mg/mL undecanoic acid and tridecanoic acid as well was 1.25 mg/mL pentadecanoic acid was added to the samples. The extraction and methylation procedure was done as described in Chapter 2[156]. Samples were analyzed using a GC-FID model Shimadzu GC-2010 equipped with an AOC-20i auto-injector and a 30 m, 0.25 mm ID RTX-5 column. The GC temperature protocol was 100°C for 2 min, ramp to 150°C (at a rate of 80°C/min), hold for 4 min, ramp to 218°C (at a rate of 4°C/min), ramp to 250°C (at a rate of 8°C/min), and hold for 2.5 min.

6.1.2.5. Flow cytometry

In order to determine the percentage of live vs dead cells we used the Thermofisher LIVE/DEAD™ BacLight™ Bacterial Viability and Counting Kit for flow cytometry following the manufacture's instructions. To prepare the samples for flow cytometry, cells were centrifuged at 10000 g for 1 minute and the pellets washed in PBS buffer three times. In order to determine the compensation and set the voltages for the FACScalibur flow cytometer, samples with only live cells and only dead cells (killed using 70% isopropanol) were prepared with single colors each (Propidium iodide and SYTO9). In addition, a 1:1 mix of the live and isopropanol killed cells with both colors was used as a control.

6.1.3. Results and discussion

In a study by Torella et al.[31] they showed that FabB can be modified to include a degradation tag sequence that combined with a protease enhancer induced during production phase lead its degradation. In that study the strategy was to force the acyl-ACP pool to be condensed with malonyl-ACP in fabF in order to increase the medium chain acyl-ACP pool. We hypothesized that a similar strategy can be used targeted towards FabA to induce its degradation in order to reduce unsaturated fatty acid buildup. The hypothesis is that dynamically degrading FabA would result in a decrease in the fraction of unsaturated phospholipids relative to the saturated. Since FabA is an essential gene and its product are essential to the membrane, this degradation must be tuned to

an optimum level rather than completely degraded. Such an approach can be done by resorting to testing multiple degradation rates.

In a study by McGuinnes and Baker[175] they engineered the native *E. coli* ssrA-SspB-ClpXP system for protein degradation of unrelated proteins to control their abundance in the cell. The way this works is by adding a peptide sequence tag (ssrA) to the protein of interest and controlling the induction of SspB enhancer which binds to ssrA and delivers the protein to the ClpXP degradation machinery of the cell. Although this works in principle, it raises question about the disrupting effects of such a system in the native ssrA-tagged proteins in *E. coli.* A similar system was developed by Cameron and Collins[176], they identified an orthogonal system which relies on the *Mesoplasma florum mf*-ssrA degradation tag and *mf*-Lon protease. These degradation tags were modified so that upon the induction of the protease mf-Lon the protein containing the tag will have degradation rates at different a steady state level. This way they created variants (pdt#3, pdt#3A, pdt#3B, pdt#3C, pdt#3D, pdt#3E) that start at 100% protein expressed and 4 h after inducing mf-Lon the protein containing the tag would reach the following steady state protein levels: 3% for pdt#3, 30% for pdt#3A, 45% for pdt#3B, 50% for pdt#3C, 70% for pdt#3D, and 80% for pdt#3E.

In order to apply this approach in the context of cells producing C_{12} FFA, we created strains with engineered degradation tags in their chromosomal *fabA* c-terminus region (**Figure 6.2**). We chose the following degradation tags: pdt#3, pdt#3A, pdt#3B for pdt#3C and integrated them as part of the protein coding sequence of *fadA* using CRISPR-Cas9 guided homologous recombination (see meterials and methods). This way, we engineered strain RLO8ara (*E. coli* K12 MG1655 ∆*araBAD ΔfadD*) and created

we created strains NHL02 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3*), NHL03 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3B*) and NHL04 (*E. coli* K12 MG1655 Δ*araBAD ΔfadD fabA-pdt#3C*).

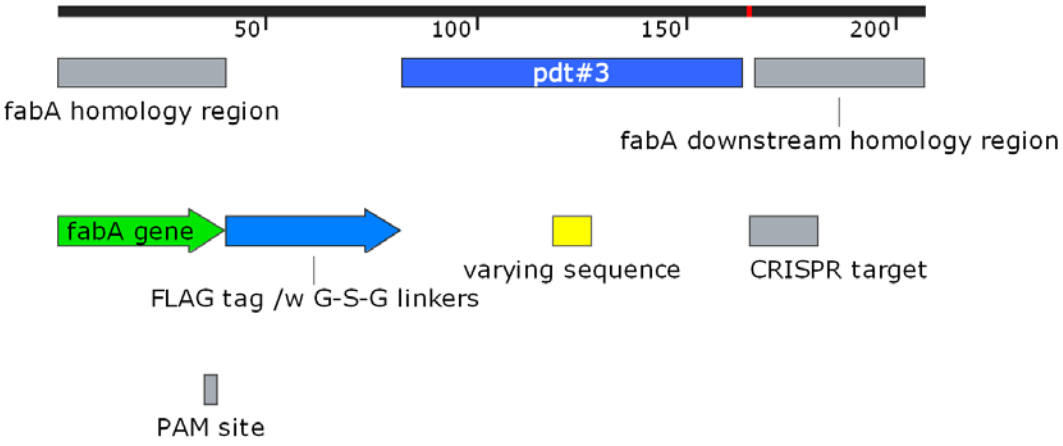


Figure 6.2. Degradation tag design at the *fabA* locus.

In order to test our hypothesis, strains RLO8ara, NHL02, NHL03, and NHL04 were cultivated under C_{12} FFA producing conditions with and without inducing FabA degradation and measured (1) FFA production, (2) cell count and (3) cell viability. To do this, each strain was transformed with two plasmids: pTRC99a-BTE containing the C_{12} specific acyl-ACP thioesterase from *Umbellularia californica*[147] under P_{trc} inducible promoter and plasmid pMP009 containing the protease mf-lon[174] under the guide of a P_{tet} inducible promoter. In addition, RLO8ara strain was transformed with pTRC99a-BTE-H204A, a catalytically inactive variant of BTE as a control. Cells were inoculated from MOPS minimal media phosphate limited overnights into 50 mL of the same media (see

materials and methods) in 250 mL shake flasks and induced at OD_{600} of 0.2 with 50 μM IPTG and 20 ng/mL ATC (where is applies) and grown at 30 °C for 48 h post induction. After 48 h, cultures where sampled for FFA content and viability assays through flow cytometry as described in the materials and methods section.

As can be seen in **Figure 6.3**, fatty acid composition for the controls strain RL08ara carrying inactive BTE-H204A and active BTE behaved as expected. When BTE is inactive, the cells do not accumulate medium chain FFA, while for active BTE RL08ara cells accumulate 154 mg/L MCFA (C_8-C_{14}) (**Figure 6.3A**). In addition, RL08ara membrane composition in BTE active cells became more unsaturated changing to 40 % unsaturated from 28% unsaturated for inactive BTE. In terms of viability, RL08ara cells carrying active BTE became less viable after 48h, going from 41 % dead in the inactive BTE to 76% dead cells when BTE is active. This data is consistent with previous studies[177]. Results for strains carrying FabA degradation tags were mixed. Our hypothesis was that strains NHL02, NHL03, and NHL04 would accumulate very similar amounts of MCFA when the protease was not induced since for that case FabA protein levels would likely remain constant. Instead we saw a decrease in the MCFA content of 59-63% in all three cases compared to RL08ara with active BTE. Interestingly, when de degradation tags are induced, the MCFA content increases is statistically equal to RL08ara (**Figure 6.3A**). In terms of viability, the trends for all the strains follow closely the expected values for cells that are producing more MCFA rather than the expected increase in saturated membrane contents and viability. That is, strains NHL02, NHL03, and NHL04 have viabilities of 34 % dead, 45% dead and 20 % dead, respectively, when the degradation tag is not induced and cells are producing low amounts of MCFA (**Figure 6.3A**). When

cells have induced degradation tags and are producing higher amounts of MCFA, the viabilities go down to 81 % dead for NHL02, 83 % dead for NHL03, and 87 % dead for NHL04 (**Figure 6.3B**). Finally, the percent of unsaturated fatty acid content did not improve NHL02, NHL03, and NHL04 cells with induced degradation of FabA. RL08ara cells expressing active BTE had 40% unsaturated FA contents whereas NHL02, NHL03, and NHL04 had 44%, 48%, and 47% unsaturated FA content, respectively. Taken together, this data suggests that the MCFA production in NHL02, NHL03, and NHL04 strains did not increased. In addition, the cells did not become more viable or less unsaturated. Overall the hypothesis was not supported by the data.

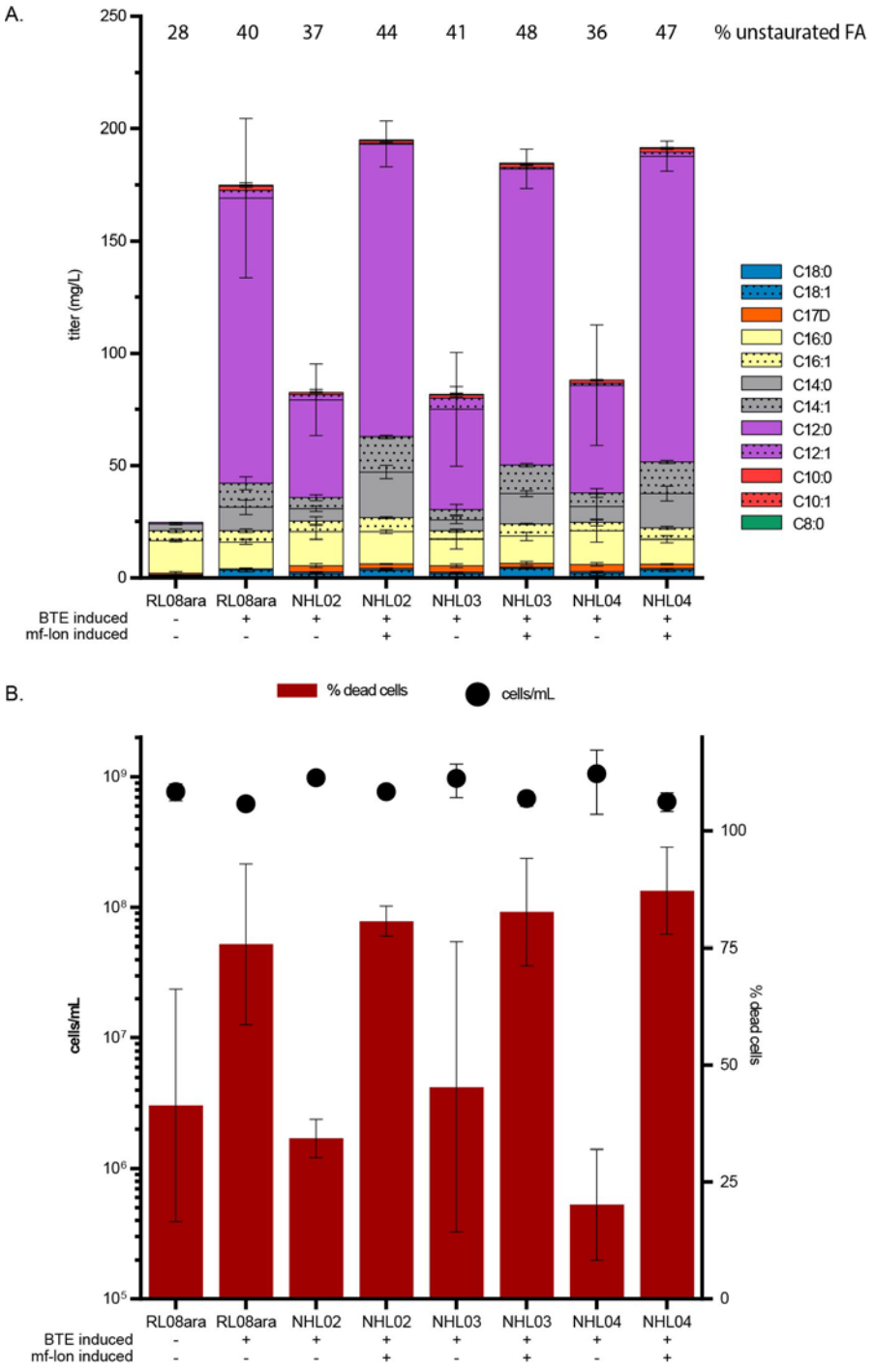


Figure 6.3. Fatty acid production and viability upon dynamic degradation of FabA.

Conclusions

In this work we implemented an inducible protein degradation system to test the hypothesis that the FabA mediated isomerization of trans-2-decenoyl-ACP to the cis form (cis-3-decenoyl-ACP) can be tuned through FabA protein degradation to change the unsaturated:saturated phospholipid contents of the cell. This in turn would restore the viability of the cell and lead to higher production of MCFA. We tested this hypothesis on C_{12} FFA producing cells and showed no correlation between the degradation of FabA and the saturated content in the phospholipids. We also found that FadA depleted cells did not become more viable or produced higher amounts of MCFA. Overall our hypothesis was not supported by the data collected.

References

1. Huber, G. W., Iborra, S. & Corma, A. Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem. Rev.* **106,** 4044–98 (2006).

2. Pfleger, B. F., Gossing, M. & Nielsen, J. Metabolic engineering strategies for microbial synthesis of oleochemicals. *Metab. Eng.* **29,** 1–11 (2015).

3. Biermann, U., Bornscheuer, U., Meier, M. A. R., Metzger, J. O. & Schäfer, H. J. Oils and Fats as Renewable Raw Materials in Chemistry. *Angew. Chemie Int. Ed.* **50,** 3854–3871 (2011).

4. Salimon, J., Salih, N. & Yousif, E. Industrial development and applications of plant oils and their biobased oleochemicals. *Arab. J. Chem.* **5,** 135–145 (2012).

5. Rupilius, W. & Ahmad, S. Palm oil and palm kernel oil as raw materials for basic oleochemicals and biodiesel. *Eur. J. Lipid Sci. Technol.* **109,** 433–439 (2007).

6. Fargione, J., Hill, J., Tilman, D., Polasky, S. & Hawthorne, P. Land Clearing and the Biofuel Carbon Debt. *Science (80-.).* **319,** 1235–1238 (2008).

7. Luterbacher, J. S. *et al.* Nonenzymatic sugar production from biomass using biomass-derived γ-valerolactone. *Science* **343,** 277–80 (2014).

8. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8,** 17 (2009).

9. Lennen, R. M. & Pfleger, B. F. Engineering Escherichia coli to synthesize free fatty acids. *Trends Biotechnol.* **30,** 659–67 (2012).

10. Sarria, S., Kruyer, N. S. & Peralta-Yahya, P. Microbial synthesis of medium-chain chemicals from renewables. *Nat. Biotechnol.* **35,** 1158–1166 (2017).

11. Lennen, R. M., Braden, D. J., West, R. a, Dumesic, J. a & Pfleger, B. F. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. *Biotechnol. Bioeng.* **106,** 193–202 (2010).

12. Youngquist, J. T. *et al.* Kinetic modeling of free fatty acid production in Escherichia coli based on continuous cultivation of a plasmid free strain. *Biotechnol. Bioeng.* **109,** 1518–27 (2012).

13. Youngquist, J. T. *et al.* Production of medium chain length fatty alcohols from glucose in Escherichia coli. *Metab. Eng.* **20,** 177–86 (2013).

14. Agnew, D. E., Stevermer, A. K., Youngquist, J. T. & Pfleger, B. F. Engineering Escherichia coli for production of C12-C14 polyhydroxyalkanoate from glucose. *Metab. Eng.* **14,** 705–13 (2012).

15. Cronan, Jr., J. E. & Rock, C. O. Biosynthesis of Membrane Lipids. *EcoSal Plus* **3,** (2008).

16. Zhang, Y.-M. & Rock, C. O. Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* **6,** 222–33 (2008).

17. Davis, M. S., Solbiati, J. & Cronan, J. E. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli. *J. Biol. Chem.* **275,** 28593–8 (2000).

18. Weber, A. L. Origin of fatty acid synthesis: thermodynamics and kinetics of reaction pathways. *J. Mol. Evol.* **32,** 93–100 (1991).

19. Yu, X., Liu, T., Zhu, F. & Khosla, C. In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli. *Proc. Natl. Acad. Sci.* **108,** 18643–18648 (2011).

20. Fujita, Y., Matsuoka, H. & Hirooka, K. Regulation of fatty acid metabolism in bacteria. *Mol. Microbiol.* **66,** 829–839 (2007).

21. Zhang, Y.-M. & Rock, C. O. Transcriptional regulation in bacterial membrane lipid synthesis. *J. Lipid Res.* **50 Suppl,** S115–9 (2009).

22. Zhang, F. *et al.* Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. *Metab. Eng.* **14,** 653–60 (2012).

23. Campbell, J. W. & Cronan, J. E. Escherichia coli FadR positively regulates transcription of the fabB fatty acid biosynthetic gene. *J. Bacteriol.* **183,** 5982–90 (2001).

24. Kunst, L. & Samuels, L. Plant cuticles shine: advances in wax biosynthesis and export. *Curr. Opin. Plant Biol.* **12,** 721–727 (2009).

25. Benning, C. Mechanisms of lipid transport involved in organelle biogenesis in plant cells. *Annu. Rev. Cell Dev. Biol.* **25,** 71–91 (2009).

26. Cantu, D. C., Chen, Y., Lemons, M. L. & Reilly, P. J. ThYme: a database for thioester-active enzymes. *Nucleic Acids Res.* **39,** D342–6 (2011).

27. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30,** 354–9 (2012).

28. Cho, H. & Cronan, J. E. Defective export of a periplasmic enzyme disrupts regulation of fatty acid synthesis. *J. Biol. Chem.* **270,** 4216–4219 (1995).

29. Voelker, T. a & Davies, H. M. 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).

30. 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).

31. Torella, J. P. *et al.* Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11290–5 (2013).

32. Fontaine, L. *et al.* Molecular characterization and transcriptional analysis of adhE2, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of Clostridium acetobutylicum ATCC 824. *J. Bacteriol.* **184,** 821–30 (2002).

33. Wahlen, B. D., Oswald, W. S., Seefeldt, L. C. & Barney, B. M. Purification, characterization, and potential bacterial Wax production role of an nadph-dependent fatty aldehyde reductase from Marinobacter aquaeolei VT8. *Appl. Environ. Microbiol.* **75,** 2758–2764 (2009).

34. Willis, R. M., Wahlen, B. D., Seefeldt, L. C. & Barney, B. M. Characterization of a Fatty Acyl-CoA Reductase from Marinobacter aquaeolei VT8: A Bacterial Enzyme Catalyzing the Reduction of Fatty Acyl-CoA to Fatty Alcohol. *ACS Biochem.* **50,** 10550–10558 (2011).

35. Machado, H. B., Dekishima, Y., Luo, H., Lan, E. I. & Liao, J. C. A selection platform for carbon chain elongation using the CoA-dependent pathway to produce linear higher alcohols. *Metab. Eng.* **14,** 504–511 (2012).

36. Kim, S., Clomburg, J. M. & 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).

37. Liu, A., Tan, X., Yao, L. & Lu, X. Fatty alcohol production in engineered E. coli expressing Marinobacter fatty acyl-CoA reductases. *Appl. Microbiol. Biotechnol.* **97,** 7061–7071 (2013).

38. Reiser, S. & Somerville, C. Isolation of mutants of Acinetobacter calcoaceticus deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *J. Bacteriol.* **179,** 2969–75 (1997).

39. Zheng, Y.-N. *et al.* Optimization of fatty alcohol biosynthesis pathway for selectively enhanced production of C12/14 and C16/18 fatty alcohols in engineered Escherichia coli. *Microb. Cell Fact.* **11,** 65 (2012).

40. Akhtar, M. K., Turner, N. J. & Jones, P. R. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 87–92 (2013).

41. Liu, R. *et al.* Metabolic engineering of fatty acyl-ACP reductase-dependent pathway to improve fatty alcohol production in Escherichia coli. *Metab. Eng.* **22,** 10–21 (2014).

42. Dellomonaco, C., Clomburg, J. M., Miller, E. N. & Gonzalez, R. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476,** 355–9 (2011).

43. Mehrer, C. R., Incha, M. R., Politz, M. C. & Pfleger, B. F. Anaerobic production of medium-chain fatty alcohols via a β-reduction pathway. *Metab. Eng.* **48,** 63–71

(2018).

44. Rai, R., Keshavarz, T., Roether, J. A., Boccaccini, A. R. & Roy, I. Medium chain length polyhydroxyalkanoates, promising new biomedical materials for the future. *Mater. Sci. Eng. R Reports* **72,** 29–47 (2011).

45. Verlinden, R. A. J., Hill, D. J., Kenward, M. A., Williams, C. D. & Radecka, I. Bacterial synthesis of biodegradable polyhydroxyalkanoates. *J. Appl. Microbiol.* **102,** 1437–1449 (2007).

46. Zhuang, Q., Wang, Q., Liang, Q. & Qi, Q. Synthesis of polyhydroxyalkanoates from glucose that contain medium-chain-length monomers via the reversed fatty acid β-oxidation cycle in Escherichia coli. *Metab. Eng.* **24,** 78–86 (2014).

47. Goh, E. B., Baidoo, E. E. K., Keasling, J. D. & Beller, H. R. Engineering of bacterial methyl ketone synthesis for biofuels. *Appl. Environ. Microbiol.* **78,** 70–80 (2012).

48. Goh, E.-B. *et al.* Substantial improvements in methyl ketone production in E. coli and insights on the pathway from in vitro studies. *Metab. Eng.* **26,** 67–76 (2014).

49. Lan, E. I., Dekishima, Y., Chuang, D. S. & Liao, J. C. Metabolic engineering of 2-pentanone synthesis in Escherichia coli. *AIChE J.* **59,** 3167–3175 (2013).

50. Welsh, F. W., Murray, W. D., Williams, R. E. & Katz, I. Microbiological and enzymatic production of flavor and fragrance chemicals. *Crit. Rev. Biotechnol.* **9,** 105–169 (1989).

51. Creuly, C., Larroche, C., Gros, J., Biologique, L. D. G. C. & Pascal, U. B. Bioconversion of fatty acids into methyl ketones by spores of Penicillium roquefortii in a water-organic solvent , two-phase system. *Enzyme Microb. Technol.* **14,** 669–678 (1992).

52. Creuly, C., Larroche, C. & Gros, J. B. A fed-batch technique for 2-heptanone production by spores of Penicillium roquefortii. *Appl. Microbiol. Biotechnol.* **34,** 20–25 (1990).

53. M., L. R. & Pfleger, B. F. Microbial production of fatty acid-derived fuels and chemicals. LID - S0958-1669(13)00042-6 [pii] LID - 10.1016/j.copbio.2013.02.028 [doi]. *Curr Opin Biotechnol* (2013).

54. Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463,** 559–62 (2010).

55. Knothe, G. ‘Designer’ biodiesel: Optimizing fatty ester composition to improve fuel properties. *Energy and Fuels* **22,** 1358–1364 (2008).

56. Howard, T. P. *et al.* Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli. *Proc. Natl. Acad. Sci.* **110,** 7636–7641 (2013).

57. Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. Fuelling the future: Microbial engineering

for the production of sustainable biofuels. *Nature Reviews Microbiology* **14,** 288–304 (2016).

58. Magnuson, K., Jackowski, S., Rock, C. O. & Cronan, J. E. J. Regulation of fatty acid biosynthesis in Escherichia coli. *Microbiol. Mol. Biol. Rev.* **57,** 522–542 (1993).

59. Choi, Y. J. & Lee, S. Y. Microbial production of short-chain alkanes. *Nature* **502,** 571–4 (2013).

60. Zhang, X., Li, M., Agrawal, A. & San, K.-Y. Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. *Metab. Eng.* **13,** 713–722 (2011).

61. Lu, X., Vora, H. & Khosla, C. Overproduction of free fatty acids in E. coli: Implications for biodiesel production. *Metab. Eng.* **10,** 333–339 (2008).

62. Dörmann, P., Voelker, T. A. & Ohlrogge, J. B. Cloning and expression in escherichia coli of a novel thioesterase from arabidopsis thaliana specific for long-chain acyl-acyl carrier proteins. *Arch. Biochem. Biophys.* **316,** 612–618 (1995).

63. Hom, L., Trinh, N. & Alibhai, M. Methods and compositions related to thioesterase enzymes. US Patent US 9,587,231 B2. (2010).

64. Jing, F. Characterization of acyl-ACP thioesterases for the purpose of diversifying fatty acid synthesis pathway. *ProQuest Diss. Publ.* (2013).

65. Mayer, K. M. & Shanklin, J. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC Plant Biol.* **7,** (2007).

66. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30,** 354–359 (2012).

67. Liu, T., Vora, H. & Khosla, C. Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. *Metab. Eng.* **12,** 378–86 (2010).

68. Lo, Y., Lin, S., Shaw, J. & Liaw, Y. Substrate Specificities of Escherichia coli Thioesterase I / Protease I / Lysophospholipase L 1 Are Governed by Its Switch Loop Movement †. *Biochemistry* 1971–1979 (2005).

69. Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Current Opinion in Structural Biology* **15,** 447–452 (2005).

70. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).

71. Gajewski, J., Pavlovic, R., Fischer, M., Boles, E. & Grininger, M. Engineering fungal de novo fatty acid synthesis for short chain fatty acid production. *Nat. Commun.* **8,** (2017).

72. Pantazes, R. J., Grisewood, M. J., Li, T., Gifford, N. P. & Maranas, C. D. The

Iterative Protein Redesign and Optimization (IPRO) suite of programs. *J. Comput. Chem.* **36,** 251–263 (2015).

73. Pavelka, A., Chovancova, E. & Damborsky, J. HotSpot Wizard: A web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* **37,** (2009).

74. Grosdidier, S. & Fernández-Recio, J. Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics* **9,** (2008).

75. Zheng, H. & Reetz, M. T. Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis. *J. Am. Chem. Soc.* **132,** 15744–15751 (2010).

76. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci.* **99,** 14116–14121 (2002).

77. Kortemme, T., Kim, D. E. & Baker, D. Computational Alanine Scanning of Protein-Protein Interfaces. *Sci. Signal.* **2004,** pl2–pl2 (2004).

78. Darnell, S. J., Page, D. & Mitchell, J. C. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins* **68,** 813–823 (2007).

79. Kawarasaki, Y. Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. *Nucleic Acids Res.* **31,** 126e–126 (2003).

80. Meyer, M. M., Hochrein, L. & Arnold, F. H. Structure-guided SCHEMA recombination of distantly related β-lactamases. *Protein Eng. Des. Sel.* **19,** 563–570 (2006).

81. Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L. & Arnold, F. H. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9,** 553–558 (2002).

82. Pantazes, R. J., Saraf, M. C. & Maranas, C. D. Optimal protein library design using recombination or point mutations based on sequence-based scoring functions. *Protein Eng. Des. Sel.* **20,** 361–373 (2007).

83. Moore, G. L. & Maranas, C. D. Computational Challenges in Combinatorial Library Design for Protein Engineering. *AIChE Journal* **50,** 262–272 (2004).

84. Saraf, M. C. & Maranas, C. D. Using a residue clash map to functionally characterize protein recombination hybrids. *Protein Eng.* **16,** 1025–34 (2003).

85. Moore, G. L. & Maranas, C. D. Identifying residue-residue clashes in protein hybrids by using a second-order mean-field approach. *Proc. Natl. Acad. Sci.* **100,** 5091–5096 (2003).

86. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).

87. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science (80-.).* **319,** 1387–1391 (2008).

88. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537,** 320–327 (2016).

89. Garrabou, X., Wicky, B. I. M. & Hilvert, D. Fast Knoevenagel Condensations Catalyzed by an Artificial Schiff-Base-Forming Enzyme. *J. Am. Chem. Soc.* **138,** 6972–6974 (2016).

90. Li, T., Pantazes, R. J. & Maranas, C. D. OptMAVEn - A new framework for the de novo design of antibody variable region models targeting specific antigen epitopes. *PLoS One* **9,** (2014).

91. Fazelinia, H., Cirino, P. C. & Maranas, C. D. Extending Iterative Protein Redesign and Optimization (IPRO) in protein library design for ligand specificity. *Biophys. J.* **92,** 2120–2130 (2007).

92. Saraf, M. C. *et al.* IPRO: An Iterative Computational Protein Library Redesign and Optimization Procedure. *Biophys. J.* **90,** 4167–4180 (2006).

93. Pantazes, R. J., Grisewood, M. J. & Maranas, C. D. Recent advances in computational protein design. *Current Opinion in Structural Biology* **21,** 467–472 (2011).

94. Samish, I., MacDermaid, C. M., Perez-Aguilar, J. M. & Saven, J. G. Theoretical and Computational Protein Design. *Annu. Rev. Phys. Chem.* **62,** 129–149 (2011).

95. Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci.* **109,** 3790–3795 (2012).

96. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Science* **19,** 1817–1819 (2010).

97. Roujeinikova, A. *et al.* 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 (2007).

98. Brooks, B. R. *et al.* CHARMM : The Biomolecular Simulation Program. (2009). doi:10.1002/jcc

99. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **52,** 3144–3154 (2012).

100. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **52,** 3155–3168 (2012).

101. Lebigot, E. O. Uncertainties: a Python package for calculations with uncertainties. Available at: http://pythonhosted.org/uncertainties/.

102. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222–D226 (2015).

103. Hall, M. A. *et al.* The WEKA data mining software: an update. *SIGKDD Explor.* **11,**

10–18 (2009).

104. Richardson, J. S., Keedy, D. A. & Richardson, D. C. in *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map* 46–61 (2013). doi:10.1142/9789814449144_0004

105. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci.* **97,** 10383–10388 (2000).

106. Guzman, L.-M., Belin, D., Carson, M. J. & Beckwith, J. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter. *J. Bacteriol.* **177,** 4121–4130 (1995).

107. Thoden, J. B. & Holden, H. M. The molecular architecture of human N-acetylgalactosamine kinase. *J. Biol. Chem.* **280,** 32784–32791 (2005).

108. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

109. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* **53,** 240–255 (1997).

110. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

111. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

112. Shin, K. S., Kim, S. & Lee, S. K. Improvement of free fatty acid production using a mutant acyl-CoA thioesterase I with high specific activity in Escherichia coli. *Biotechnol. Biofuels* **9,** 208 (2016).

113. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

114. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26,** 1781–1802 (2005).

115. Grisewood, M. J. *et al.* OptZyme: Computational Enzyme Redesign Using Transition State Analogues. *PLoS One* **8,** (2013).

116. Khoury, G. A. *et al.* Computational design of Candida boidinii xylose reductase for altered cofactor specificity. *Protein Sci.* **18,** 2125–2138 (2009).

117. Fazelinia, H., Cirino, P. C. & Maranas, C. D. OptGraft: A computational procedure for transferring a binding site onto an existing protein scaffold. *Protein Sci.* **18,** 180–195 (2009).

118. Feldmeier, K. & Höcker, B. Computational protein design of ligand binding and catalysis. *Curr. Opin. Chem. Biol.* **17,** 929–933 (2013).

119. Khare, S. D. *et al.* Computational redesign of a mononuclear zinc metalloenzyme

for organophosphate hydrolysis. *Nat. Chem. Biol.* **8,** 294–300 (2012).

120. Gordon, S. R. *et al.* Computational Design of an alpha-Gliadin Peptidase. *J. Am. Chem. Soc.* **134,** 20513–20520 (2012).

121. Chen, C.-Y., Georgiev, I., Anderson, A. C. & Donald, B. R. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci.* **106,** 3764–3769 (2009).

122. Lo, Y. C., Lin, S. C., Shaw, J. F. & Liaw, Y. C. Crystal structure of Escherichia coli thioesterase I/protease I/lysophospholipase L1: Consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. *J. Mol. Biol.* **330,** 539–551 (2003).

123. Morozov, A. V., Havranek, J. J., Baker, D. & Siggia, E. D. Protein-DNA binding specificity predictions with structural models. *Nucleic Acids Res.* **33,** 5781–5798 (2005).

124. Meiler, J. & Baker, D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins Struct. Funct. Genet.* **65,** 538–548 (2006).

125. Lennen, R. M. & Pfleger, B. F. Microbial production of fatty acid-derived fuels and chemicals. *Curr. Opin. Biotechnol.* **24,** 1044–53 (2013).

126. Graham, S. A., Hirsinger, F. & Robbelen, G. Fatty Acids of Cuphea (Lythraceae) Seed Lipids and Their Systematic Significance. *Am. J. Bot.* **68,** 908–917 (1981).

127. Grisewood, M. J. *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

128. Yuan, L., Voelker, T. A. & Hawkins, D. J. Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. **92,** 10639–10643 (1995).

129. Feng, Y. *et al.* Structural Insight into Acyl-ACP Thioesterase toward Substrate Specificity Design. *ACS Chem. Biol.* **12,** 2830–2836 (2017).

130. Jing, F., Zhao, L., Yandeau-Nelson, M. D. & Nikolau, B. J. Two distinct domains contribute to the substrate acyl chain length selectivity of plant acyl-ACP thioesterase. *Nat. Commun.* **9,** 860 (2018).

131. Ziesack, M. *et al.* Chimeric Fatty Acyl-ACP Thioesterases Provide Mechanistic Insight into Enzyme Specificity and Expression. *Appl. Environ. Microbiol.* AEM.02868–17 (2018). doi:10.1128/AEM.02868-17

132. Mukherjee, K., Bhattacharyya, S. & Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth. Biol.* **4,** 1261–1269 (2015).

133. Sarria, S., Bartholow, T. G., Verga, A., Burkart, M. D. & Peralta-Yahya, P. Matching Protein Interfaces for Improved Medium-Chain Fatty Acid Production. *ACS Synth. Biol.* **7,** 1179–1187 (2018).

134. Cronan, J. E., Zhao, X. & Jiang, Y. *Function , Attachment and Synthesis of Lipoic Acid in Escherichia coli*. *Advances in Microbial Physiology* **50,** (Elsevier Masson SAS, 2005).

135. Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A. & Cronan, J. E. Assembly of the Covalent Linkage between Lipoic Acid and Its Cognate Enzymes. *Chem. Biol.* **10,** 1293–1302 (2003).

136. Dehesh, K., Edwards, P., Hayes, T., Cranmer, A. M. & 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).

137. Li, Y. *et al.* Metabolic engineering of Escherichia coli using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31,** 13–21 (2015).

138. Datsenko, K. a & Wanner, B. L. 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).

139. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–5 (2009).

140. Lennen, R. M. & Pfleger, B. F. Modulating membrane composition alters free fatty acid tolerance in Escherichia coli. *PLoS One* **8,** e54031 (2013).

141. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture Medium for Enterobacteria. *J. Bacteriol.* **119,** 736–747 (1974).

142. Youngquist, J. T., Rose, J. P. & Pfleger, B. F. Free fatty acid production in Escherichia coli under phosphate-limited conditions. *Appl. Microbiol. Biotechnol.* **97,** 5149–59 (2013).

143. Beld, J., Finzel, K. & Burkart, M. D. Versatility of acyl-acyl carrier protein synthetases. *Chem. Biol.* **21,** 1293–1299 (2014).

144. Mendonça, L. M. F. & Marana, S. R. Single mutations outside the active site affect the substrate specificity in a β-glycosidase. *Biochim. Biophys. Acta - Proteins Proteomics* **1814,** 1616–1623 (2011).

145. Lu, X., Vora, H. & Khosla, C. Overproduction of free fatty acids in E. coli: implications for biodiesel production. *Metab. Eng.* **10,** 333–9 (2008).

146. Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A. & Cronan, J. E. Assembly of the Covalent Linkage between Lipoic Acid and Its Cognate Enzymes. *Chem. Biol.* **10,** 1293–1302 (2003).

147. Hoover, S. W. *et al.* Bacterial production of free fatty acids from freshwater macroalgal cellulose. *Appl. Microbiol. Biotechnol.* **91,** 435–46 (2011).

148. Lennen, R. M. *et al.* Membrane stresses induced by overproduction of free fatty acids in Escherichia coli. *Appl. Environ. Microbiol.* **77,** 8114–28 (2011).

149. USDA, A. A. D. for the U. N. O. P. *Fatty Alcohols (Octanol and Decanol)*. (2016).

150. Akhtar, M. K., Dandapani, H., Thiel, K. & Jones, P. R. Microbial production of 1-octanol: A naturally excreted biofuel with diesel-like properties. *Metab. Eng. Commun.* **2,** 1–5 (2015).

151. Van Leeuwen, P. W. N. M., Clément, N. D. & Tschan, M. J. L. New processes for the selective production of 1-octene. *Coord. Chem. Rev.* **255,** 1499–1517 (2011).

152. Falbe, J., Bahrmann, H., Lipps, W., Mayer, D. & Guido D. Frey. in *Ullmann's Encyclopedia of Industrial Chemistry* 235–261 (2011).

153. Pérez, J. M., Arenas, F. A., Pradenas, G. A., Sandoval, J. M. & Vásquez, C. C. Escherichia coli YqhD exhibits aldehyde reductase activity and protects from the harmful effect of lipid peroxidation-derived aldehydes. *J. Biol. Chem.* **283,** 7346–7353 (2008).

154. Hernandez-Lozada, N. J. *et al.* Highly active C8-acyl-ACP thioesterase variant isolated by a synthetic selection strategy. *ACS Synth. Biol.* 2205–2215 (2018). doi:10.1021/acssynbio.8b00215

155. Iram, S. H. & Cronan, J. E. The Beta-Oxidation Systems of Escherichia coli and Salmonella enterica Are Not Functionally Equivalent. *J. Bacteriol.* **188,** 599–608 (2006).

156. Grisewood, M. J. *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

157. Rand, J. M., Gordon, G. C., Mehrer, C. R. & Pfleger, B. F. Genome sequence and analysis of Escherichia coli production strain LS5218. *Metab. Eng. Commun.* **5,** 78–83 (2017).

158. Atsumi, S. *et al.* Metabolic engineering of Escherichia coli for 1-butanol production. *Metab. Eng.* **10,** 305–311 (2008).

159. Iram, S. H. & Cronan, J. E. The B-Oxidation Systems of Escherichia coli and Salmonella enterica Are Not Functionally Equivalent. *J. Bacteriol.* **188,** 599–608 (2006).

160. Ford, T. J. & Way, J. C. Enhancement of E. coli acyl-CoA synthetase FadD activity on medium chain fatty acids. *PeerJ* **3,** (2015).

161. Chen, G. Q. & Wu, Q. The application of polyhydroxyalkanoates as tissue engineering materials. *Biomaterials* **26,** 6565–6578 (2005).

162. Brigham & Sinskey. Applications of Polyhydroxyalkanoates in the Medical Industry. *Int. J. Biotechnol. Wellness Ind.* 53–60 (2012). doi:10.6000/1927-3037.2012.01.01.03

163. Urtuvia, V., Villegas, P., González, M. & Seeger, M. Bacterial production of the biodegradable plastics polyhydroxyalkanoates. *Int. J. Biol. Macromol.* **70,** 208–213 (2014).

164. Rai, R. *et al.* Poly-3-hydroxyoctanoate P(3HO), a medium chain length polyhydroxyalkanoate homopolymer from pseudomonas mendocina. *Biomacromolecules* **12,** 2126–2136 (2011).

165. Zinn, M., Witholt, B. & Egli, T. Occurrence, synthesis and medical application of bacterial polyhydroxyalkanoate. *Adv. Drug Deliv. Rev.* **53,** 5–21 (2001).

166. Braunegg, G., Sonnleitner, B. & Lafferty, R. M. A rapid gas chromatographic method for the determination of poly-β-hydroxybutyric acid in microbial biomass. *Eur. J. Appl. Microbiol. Biotechnol.* **6,** 29–37 (1978).

167. Gripon, J. C. in *Cheese: Chemistry, Physics and Microbiology* 124–127 (1993).

168. Cantu, D. C., Chen, Y. & Reilly, P. J. Thioesterases: A new perspective based on their primary and tertiary structures. *Protein Sci.* **19,** 1281–1295 (2010).

169. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).

170. Nie, L., Ren, Y. & Schulz, H. Identification and Characterization of Escherichia coli Thioesterase III that Functions in Fatty Acid ß-Oxidation. *Biochemistry* **47,** 7744–7751 (2008).

171. Zhu, K., Zhang, Y.-M. & Rock, C. O. Transcriptional regulation of membrane lipid homeostasis in Escherichia coli. *J. Biol. Chem.* **284,** 34880–8 (2009).

172. He, L. *et al.* Central metabolic responses to the overproduction of fatty acids in Escherichia coli based on 13C-metabolic flux analysis. *Biotechnol. Bioeng.* **111,** 575–85 (2014).

173. Lennen, R. M., Politz, M. G., Kruziki, M. a & Pfleger, B. F. Identification of transport proteins involved in free fatty acid efflux in Escherichia coli. *J. Bacteriol.* **195,** 135–44 (2013).

174. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32,** (2014).

175. McGinness, K. E., Baker, T. a & Sauer, R. T. Engineering controllable protein degradation. *Mol. Cell* **22,** 701–7 (2006).

176. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.3053

177. Lennen, R. M. *et al.* Membrane stresses induced by overproduction of free fatty acids in Escherichia coli. *Appl. Environ. Microbiol.* **77,** 8114–8128 (2011).

178. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).

179. Agnew, D. E., Stevermer, A. K., Youngquist, J. T. & Pfleger, B. F. Engineering Escherichia coli for production of C12-C14 polyhydroxyalkanoate from glucose. *Metab. Eng.* **14,** 705–13 (2012).

180. Amann, E., Ochs, B. & Abel, K. J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene* **69,** 301–315 (1988).

181. Dormann, P., Voelker, T. A. & Ohlrogge, J. B. CLONING AND EXPRESSION IN ESCHERICHIA-COLI OF A NOVEL THIOESTERASE FROM ARABIDOPSIS-THALIANA SPECIFIC FOR LONG-CHAIN ACYL-ACYL CARRIER PROTEINS. *Arch. Biochem. Biophys.* **316,** 612–618 (1995).

182. Ghosh, S. K. *et al.* Characterization and cloning of a stearoyl/oleoyl specific fatty acyl-acyl carrier protein thioesterase from the seeds of Madhuca longifolia (latifolia). *Plant Physiol. Biochem.* **45,** 887–897 (2007).

183. Jha, J. K. *et al.* Cloning and functional expression of an acyl-ACP thioesterase FatB type from Diploknema (Madhuca) butyracea seeds in Escherichia coli. *Plant Physiol. Biochem.* **44,** 645–655 (2006).

184. Dong, S. *et al.* Cloning, characterization, and expression analysis of acyl-acyl carrier protein (ACP)-thioesterase B from seeds of Chinese Spicehush (Lindera communis). *Gene* **542,** 16–22 (2014).

185. Chen, G. *et al.* Cloning of acyl-ACP thioesterase FatA from Arachis hypogaea L. and its expression in Escherichia coli. *J. Biomed. Biotechnol.* **2012,** 652579 (2012).

186. Zheng, Y. *et al.* Boosting the free fatty acid synthesis of Escherichia coli by expression of a cytosolic Acinetobacter baylyi thioesterase. *Biotechnol. Biofuels* **5,** (2012).

187. Jha, J. K. *et al.* Functional expression of an acyl carrier protein (ACP) from Azospirillum brasilense alters fatty acid profiles in Escherichia coli and Brassica juncea. *Plant Physiol. Biochem.* **45,** 490–500 (2007).

188. Gong, Y., Guo, X., Wan, X., Liang, Z. & Jiang, M. Characterization of a novel thioesterase (PtTE) from Phaeodactylum tricornutum. *J. Basic Microbiol.* **51,** 666–672 (2011).

189. Serrano-Vega, M. J., Garcés, R. & Martínez-Force, E. Cloning, characterization and structural model of a FatA-type thioesterase from sunflower seeds (Helianthus annuus L.). *Planta* **221,** 868–880 (2005).

190. Lee, S., Park, S., Park, C., Pack, S. P. & Lee, J. Enhanced free fatty acid production by codon-optimized Lactococcus lactis acyl-ACP thioesterase gene expression in Escherichia coli using crude glycerol. *Enzyme Microb. Technol.* **67,** 8–16 (2014).

191. Jing, F. Characterization of acyl-ACP thioesterases for the purpose of diversifying fatty acid synthesis pathway. Ph.D. Thesis, Iowa State University, Ames, IA, 2013. (Iowa State University, 2013).

Appendices

Appendix I: Strains used

Strain/plasmid	Genotype	Chapter	Source
E. coli K12 MG1655	F^- λ^- $ilvG^-$ rfb-50 rph-1	2-6	CGSG
E. coli BL21 DE3	F^- *hddS gal lacI::PlacUV5::T7 gene1*	2-3	178
ΔlipB	K12 MG1655 *ΔlipB*	3	This work
RL08ara	K-12 MG1655 *ΔaraBAD ΔfadD*	2-6	11
araBAD	K-12 MG1655 *ΔaraBAD*	3	179
NHL17	K-12 MG1655 *ΔaraBAD ΔfadD::trcCpFatB1.2-M4-287*	3	This work
NHL09	K-12 MG1655 *ΔaraBAD ΔfadD ΔackA-pta*	4	This work
NHL10	K-12 MG1655 *ΔaraBAD ΔfadD ΔackA-pta ΔpoxB*	4	This work
NHL11	K-12 MG1655 *ΔaraBAD ΔfadD ΔackA-pta ΔpoxB ΔldhA*	4	This work
MHS04	K-12 MG1655 *ΔfadR ΔfadE ΔfadD*	4	13
DE	K-12 MG1655 *ΔaraBAD ΔfadD ΔfadE*	4	179
NHL19	K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287 ΔfadE*	4	This work
NHL22	K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287 ΔfadE Δpta::Ptrc-Mt-FadD6*	4	This work
LS5218	*E. coli* LS5218	4	157
LS5218 ΔfadD	LS5218 *ΔfadD*	4	This work
LS5218 ΔfadD::CupTE2M4-287	LS5218 with chromosomally integrated *ΔfadD::CpFatB1.2-M4-287*	4	This work
RABIJ	K-12 MG1655 Δ*araBAD* Δ*fadR* Δ*fadAB* Δ*fadIJ*	5	14
NHL18	K-12 MG1655 Δ*araBAD* Δ*fadR* Δ*fadAB* Δ*fadIJ fadD::Ptrc-CpFatB1.2-M4-287*	5	This work
TRS10	K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287* ΔfadA ΔfadR	5	This work
TRS11	K-12 MG1655 *ΔaraBAD ΔfadD::PtrcCpFatB1.2-M4-287* ΔfadA ΔfadR *ΔfadE*	5	This work
RAI	K-12 MG1655 Δ*araBAD* Δ*fadR* Δ*fadA* Δ*fadI*	5	14
NHL02	*K-12 MG1655 ΔaraBAD ΔfadD fabA-pdt#3;* pdt sequence from Cameron et al.[174]	6	This work
NHL03	*K-12 MG1655 ΔaraBAD ΔfadD fabA-pdt#3B;* pdt sequence from Cameron et al.[174]	6	This work
NHL04	*K-12 MG1655 ΔaraBAD ΔfadD fabA-pdt#3C;* pdt sequence from Cameron et al.[174]	6	This work

Appendix II: Plasmids used

Strain/plasmid	Genotype	Chapter	Source
pBad18	P_{BAD} promoter, pBR322 origin, Amp^R	2	106
pBad18-tesA'	pBad18 with leaderless *E.coli tesA'* thioesterase gene. For a list of mutants created in this study see Appendix 4.	2	This work
pBTRKtrc	P_{trc} promoter, pBBR1 origin, Kan^R	3-6	13
pTRC99a	P_{trc} promoter, pBR322 origin, Amp^R	3-6	180
pBad33	P_{BAD} promoter, pACYC origin, Cm^R	3	106
pACYC	P_{trc} promoter, pACYC origin, Cm^R	3-6	13
pBad33-CpFatB1.2	pBad33 with *Cuphea palustris* CpFatB1 gene truncated at MLLTAIT	3	This work
pTRC99a-CpFatB1	pTRC99a with *Cuphea palustris* CpFatB1 full gene sequence	3	This work
pTRC99a-CpFatB1.2	pTRC99a with *Cuphea palustris* CpFatB1 gene truncated at MLLTAIT	3	This work
pTRC99a-CpFatB1.3	pTRC99a with *Cuphea palustris* CpFatB1 gene truncated at MKSKRPN	3	This work
pTRC99a-CpFatB1.4	pTRC99a with *Cuphea palustris* CpFatB1 gene truncated at MGLVFRQ	3	This work
pBTRK-CpFatB1.2	pBTRKtrc with *Cuphea palustris* CpFatB1.2 under Ptrc promoter	3	This work
pBTRK-CpFatB1.2-M4	pBTRKtrc with CpFatB1.2-M4 sequence (see Appendix X)	3	This work
pBTRK-CpFatB1.2-M4-287	pBTRKtrc with CpFatB1.2-M4 sequence truncated (see **Figure** 3.6)	3	This work
pBTRK-CpFatB1.2-M4-288	pBTRKtrc with CpFatB1.2-M4 sequence truncated (see **Figure** 3.6)	3	This work
pBTRK-CpFatB1.2-M4-289	pBTRKtrc with CpFatB1.2-M4 sequence truncated (see **Figure** 3.6)	3	This work
pBTRK-CpFatB1.2-M4-290	pBTRKtrc with CpFatB1.2-M4 sequence truncated (see **Figure** 3.6)	3	This work
pBTRK-CpFatB1.2-M4-291	pBTRKtrc with CpFatB1.2-M4 sequence truncated ((see **Figure** 3.6)	3	This work
pTRC99a-CpFatB1.2-M3	CpFatB1.2-M3 variant isolated from the ΔlipB based selection subcloned into pTRC99a plasmid	3	This work
pTRC99a-CpFatB1.2-M4	CpFatB1.2-M4 variant isolated from the ΔlipB based selection subcloned into pTRC99a plasmid	3	This work
pTRC99a-CpFatB1.2-M4-CHis	CpFatB1.2-M4 variant isolated from the ΔlipB based selection subcloned into pTRC99a plasmid with an additional Cterminus Histidine tag with sequence: AAHHHHHH	3	This work
pTRC99a-CpFatB1.2-M9	CpFatB1.2-M9 variant isolated from the ΔlipB based selection subcloned into pTRC99a plasmid	3	This work
pTRC99a-CpFatB1.2-M4-287	pTRC99a with CpFatB1.2-M4 sequence truncated (see **Figure** 3.6)	3	This work
pET28-CpFatB1.2	CpFatB1.2 thioesterase gene cloned into pET28t containing a TEV protease site. Translated sequence contains the following peptide: MGSSHHHHHHSSENLYFQGGGG	3	This work
pET28-CpFatB1.2-M4-287	CpFatB1.2-M4-287 thioesterase gene cloned into pET28t containing a TEV protease site. Translated sequence contains the following peptide: MGSSHHHHHHSSENLYFQGGGG	3	This work
pET28-VhAasS	*Vibrio harveii* acyl-ACP synthethase gene cloned into pET28t containing a TEV protease site. Translated sequence contains the following peptide: MGSSHHHHHHSSENLYFQGGGG	3	This work
pET28-BsSFP	*Bacillus subtilis* 4'-phosphopantetheinyl transferases gene cloned into pET28t containing a TEV protease site. Translated sequence contains the following peptide: MGSSHHHHHHSSENLYFQGGGG	3	This work
pET28-EcACP	*Escherichia coli* K12 MG1655 acyl carrier protein gene cloned into pET28t containing a TEV protease site. Translated sequence contains the following peptide: MGSSHHHHHHSSENLYFQGGGG	3	This work

pBTRK-Ma-ACR	pBTRKtrc with *Marinobacter aquaeolei VT8* acyl-CoA reductase	4	43
pBTRK-Mb-ACR	pBTRKtrc with *Marinobacter BSs20148* acyl-CoA reductase	4	43
pBTRK-Mt-ACR	pBTRKtrc with *Methylibium Sp. T29* acyl-CoA reductase	4	43
pBTRK-Na-ACR	pBTRKtrc with *Nocardia asteroids* acyl-CoA reductase	4	43
pBTRK-Ca-adhE2	pBTRKtrc with *Clostridium acetobutylicum* acyl-CoA reductase	4	43
pBTRK-Ra-ACR	pBTRKtrc with *Rhodococcus aetherivorans* acyl-CoA reductase	4	43
pBTRK-Mc-ACR	pBTRKtrc with *Mycobacterium chlorophenolicum* acyl-CoA reductase	4	43
pBTRK-Mh-ACR	pBTRKtrc with *Mycobacterium hassiacum* acyl-CoA reductase	4	43
pBTRK-Rc-ACR	pBTRKtrc with *Rhodococcus sp. PBTS 2* acyl-CoA reductase	4	43
pACYC-FadD6	pACYC with *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase	4	13
pACYC-Pp_0763	pACYC with *Pseudomonas putida Pp_0763* acyl-CoA synthetase	4	13
pTRC99a-fadD	pTRC99a with *E. Coli FadD* acyl-CoA synthetase	4	13
pMSB6	pTRC99a with altered MCS	5	4
pMSB6-J1	pMSB6 containing *P. aeruginosa phaJ1*	5	4
pMSB6-J2	pMSB6 containing *P. aeruginosa phaJ2*	5	4
pMSB6-J3	pMSB6 containing *P. aeruginosa phaJ3*	5	4
pMSB6-J4	pMSB6 containing *P. aeruginosa phaJ4*	5	4
pBAD33-C280*	pBAD33 *araE* C280* ∆281-292	5	4
pBAD33*–C1	pBAD33-C280* containing *P. aeruginosa phaC1*	5	4
pBAD33*–C2	pBAD33-C280* containing *P. aeruginosa phaC2*	5	4
pDA-JAC	pMSB6 containing *P. aeruginosa phaJ3* and *phaC2* genes and *P. putida PP_0763* cloned between *phaJ3* and *phaC2*	5	14
pKX-J1AC2	pDA-JAC with *P. aeruginosa phaJ1* substitution	5	This work
pKX-J2AC2	pDA-JAC with *P. aeruginosa phaJ2* substitution	5	This work
pKX-J3AC2	pDA-JAC with corrected point mutation in *phaJ3*	5	This work
pKX-J4AC2	pDA-JAC with *P. aeruginosa phaJ4* substitution	5	This work
pKX-J1AC1	pKX-J1AC2 with *P. aeruginosa phaC1* substitution	5	This work
pKX-J2AC1	pKX-J2AC2 with *P. aeruginosa phaC1* substitution	5	This work
pKX-J3AC1	pKX-J3AC2 with *P. aeruginosa phaC1* substitution	5	This work
pKX-J4AC1	pKX-J4AC2 with *P. aeruginosa phaC1* substitution	5	This work
pBTRK-fadE	pBTRKtrc containing *E. coli fadE*	5	This work
pTRC99a-Ec-FadM-Mt-FadD6	pTRC99a with *E. Coli fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters	5	This work
pTRC99a-Ec-FadM-Mt-FadD6	pTRC99a with *E. Coli fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters	5	This work
Ec-FadM	pTRC99a with *E. Coli fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Se-FadM	pTRC99a with *Salmonella enterica fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Ng-FadM	pTRC99a with *Neisseria gonorrhoeae fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Ps-FadM	pTRC99a with *Providencia sneebia fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work

Mu-FadM	pTRC99a with *Marinomonas ushuaiensis fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
So-FadM	pTRC99a with *Serratia odorifera fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Cs-FldM	pTRC99a with *Cucumis sativus fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Bb-FudM	pTRC99a with *Beauveria bassiana fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Ba-FadM	pTRC99a with *Buttiauxella agrestis fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Ct-FadM	pTRC99a with *Cronobacter turicensis fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Pa-FadM	pTRC99a with *Pseudomonas aeruginosa fadM* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Mt-FadM1	pTRC99a with *Mycobacterium tuberculosis fadM1* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Mt-FadM2	pTRC99a with *Mycobacterium tuberculosis fadM2* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Mt-FadM3	pTRC99a with *Mycobacterium tuberculosis fadM3* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Mt-FadM4	pTRC99a with *Mycobacterium tuberculosis fadM4* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
Mt-FadM5	pTRC99a with *Mycobacterium tuberculosis fadM5* thioesterase and *Mycobacterium tuberculosis FadD6* acyl-CoA synthetase under separate P_{trc} promoters (sequence in Appendix XII)	5	This work
pMP009	P_{tet} promoter, pBBR1 origin, Kan^R with protease mf-lon from Cameron, et al.[174]	6	This work
pTRC99a-BTE	pTRC99a with *Umbellularia californica* acyl-ACP thioesterase	6	147
pTRC99a-BTE-H204A	pTRC99a with *Umbellularia californica* acyl-ACP thioesterase catalytically inactive through H204A mutation	6	147

Appendix III. FFA production profiles for various thioesterases expressed in *E. coli*.

This table provides the mole fractions of various FFAs and total FFA in the cell lysate for enzymes considered in this study as well as those found in literature. Computationally-predicted mutants are named R*x*.M*y*, rationally-designed mutants are named R*x*.RD*y*, and randomly-generated mutants are labeled RM.M*y*, where *x* is the Round number and *y* is the Mutant number. Any mutant labeled in this manner derives from this study. Heterologously-expressed and overexpressed thioesterases in *E. coli* are named using their GenBank accession label and engineered mutants from literature (beginning with "CvB2MT", "rTE", "TEGm" or "UcFatB") are named as they are in the original source. "nd" indicates that the value was not determined or provided.

Name	Organism	8:0	10:0	12:0	12:1	14:0	14:1	16:0	16:1	18:0	18:1	Total FFA (μM)	Notes	Source
Control	*E. coli*	2.4±0.3	0.6±0.2	3.6±0.8	0.6±0.9	17±2	0.51±0.06	57±6	12±2	1.3±0.1	4.5±0.6	300±20	-	-
WT	*E. coli*	5.9±0.3	1.09±0.06	20±1	7.5±0.5	29±2	7.1±0.4	14.2±0.8	11.6±0.8	0.22±0.01	2.1±0.1	1900±50	-	-
R1.M1	*E. coli*	19±2	6.0±0.5	12.1±0.7	31±2	4.4±0.3	14.9±0.7	4.7±0.7	6.4±0.8	1.52±0.10	0.11±0.01	4300±100	-	-
R1.M2	*E. coli*	13±2	2.2±0.5	5.8±0.5	2.0±0.1	28±2	3.7±0.3	27±2	8.5±0.6	0.9±0.2	4.1±0.5	650±30	-	-
R1.M3	*E. coli*	5±4	1±2	3±2	3±2	11±2	0.0±0.0	60±10	2.5±0.5	1.9±0.4	4±1	190±40	-	-
R1.M4	*E. coli*	0.0±0.0	0.0±0.0	10±10	3±6	12±2	3±1	64±8	3.2±0.7	2.5±0.5	3.5±0.7	230±30	-	-
R1.M5	*E. coli*	0.0±0.0	0.0±0.0	3±5	2±3	13±3	2±3	70±6	4±1	2.3±0.3	4.3±0.6	160±20	-	-
R1.M6	*E. coli*	0.0±0.0	0.0±0.0	10±10	4±6	14±5	2±4	60±10	4±2	2.3±0.5	4±1	180±30	-	-
R1.M7	*E. coli*	0.0±0.0	0.0±0.0	2±3	0.0±0.0	13±2	3±3	72±5	3.5±0.8	2.4±0.5	4.2±0.8	170±20	-	-
R1.M8	*E. coli*	0.0±0.0	0.0±0.0	1±2	0.0±0.0	13±4	1±2	74±5	3.3±0.8	2.5±0.7	4±1	170±30	-	-
R1.M9	*E. coli*	0.0±0.0	0.0±0.0	5±4	2±3	14±2	2±3	69±4	3.0±0.4	2.6±0.5	4.1±0.3	180±10	-	-
R1.M10	*E. coli*	0.0±0.0	0.0±0.0	4±3	0.0±0.0	13±1	1±2	72±3	3.3±0.2	2.4±0.2	4.8±0.9	164±8	-	-
R1.M11	*E. coli*	0.0±0.0	0.0±0.0	3±3	0.0±0.0	14±3	0.0±0.0	73±4	4±1	2.6±0.5	3.4±0.7	180±20	-	-
R1.M12	*E. coli*	0.0±0.0	0.0±0.0	4±3	2±3	13±2	3±3	69±4	2.9±0.3	2.5±0.3	4.0±0.6	160±10	-	-
R2.M1	*E. coli*	7±1	0.0±0.0	1.2±0.1	0.0±0.0	12±1	0.7±0.1	47±2	13.6±0.7	0.7±0.0	4.2±0.3	510±20	-	-
R2.M2	*E. coli*	7.3±0.4	0.7±0.0	1.4±0.1	0.0±0.0	14.7±0.7	0.9±0.0	53.8±0.9	15.7±0.8	0.8±0.0	4.7±0.2	447±8	-	-
R2.M3	*E. coli*	8±1	0.0±0.0	1.4±0.1	0.0±0.0	11.8±0.4	0.7±0.0	47±1	14.6±0.7	0.8±0.0	4.1±0.3	480±10	-	-
R2.M4	*E. coli*	6.3±0.7	0.6±0.0	1.3±0.1	0.0±0.0	12.4±0.8	0.7±0.1	47±2	14±1	0.4±0.0	3.7±0.4	520±20	-	-
R2.M5	*E. coli*	5.9±0.4	2.4±0.2	12.9±0.8	6.1±0.4	38±2	6.4±0.5	8.4±0.6	17±2	0.1±0.0	1.8±0.2	3500±100	-	-
R2.M6	*E. coli*	1.2±0.2	0.0±0.0	1.3±0.1	0.0±0.0	11.6±0.6	0.7±0.0	49±2	15.3±0.9	0.7±0.1	4.0±0.3	500±10	-	-
R2.M7	*E. coli*	1.3±0.1	0.0±0.0	1.4±0.0	0.0±0.0	12.4±0.5	0.7±0.0	48±1	16.2±0.4	0.8±0.0	4.1±0.2	510±10	-	-
R2.M8	*E. coli*	1.5±0.1	0.0±0.0	1.6±0.1	0.0±0.0	15±1	0.9±0.1	57±2	19±1	0.8±0.0	4.9±0.5	440±20	-	-

R2.M9	*E. coli*	1.0±0.1	0.0±0.0	1.2±0.1	0.0±0.0	11.0±0.7	0.6±0.0	48±1	15.4±0.5	0.8±0.0	4.1±0.3	420±10	-	-
R2.M10	*E. coli*	1.2±0.2	0.8±0.1	1.5±0.2	0.0±0.0	14±2	0.8±0.1	56±3	20±3	0.9±0.1	4.7±0.4	380±30	-	-
R2.M11	*E. coli*	1.4±0.1	0.0±0.0	1.6±0.1	0.0±0.0	14.3±0.9	0.8±0.0	57±1	18.5±0.8	1.0±0.0	5.1±0.2	355±10	-	-
R2.M12	*E. coli*	1.3±0.2	0.0±0.0	1.6±0.1	0.0±0.0	13.1±0.8	0.8±0.0	57.1±0.9	18.9±0.5	2.0±0.1	5.2±0.2	343±7	-	-
R2.M13	*E. coli*	1.1±0.3	0.0±0.0	1.4±0.3	0.0±0.0	9±3	0.6±0.2	62±8	11±5	0.7±0.7	4±2	480±90	-	-
R2.M14	*E. coli*	1.8±0.2	0.9±0.1	2.6±0.3	0.0±0.0	15±1	1.1±0.1	44±2	15±1	0.9±0.1	4.1±0.4	460±20	-	-
R2.RD1	*E. coli*	2.9±0.2	2.8±0.2	20.5±0.9	1.7±0.1	31.2±0.7	7.8±0.3	17.1±0.7	13.2±0.5	0.1±0.0	2.0±0.1	2100±30	-	-
R2.RD2	*E. coli*	1.2±0.0	0.8±0.0	1.5±0.1	0.0±0.0	12.2±0.3	0.7±0.0	46.4±0.9	15.4±0.6	0.8±0.0	4.5±0.1	380±8	-	-
R2.RD3	*E. coli*	1.4±0.1	0.9±0.0	1.9±0.1	0.0±0.0	16±1	0.8±0.0	55.7±0.9	17.8±0.6	1.0±0.0	4.7±0.4	336±6	-	-
R2.RD4	*E. coli*	1.2±0.1	0.7±0.0	1.4±0.1	0.0±0.0	12.7±0.8	0.6±0.0	46±2	15±1	0.8±0.1	4.1±0.3	420±10	-	-
R2.RD5	*E. coli*	2.4±0.1	0.9±0.1	2.2±0.1	0.0±0.0	14±1	1.3±0.1	44±2	15±1	0.8±0.1	3.9±0.2	430±10	-	-
R3.M1	*E. coli*	15±4	6±2	35±7	14±4	14±4	5±2	6±2	3.8±0.8	0.1±0.0	1.4±0.4	3500±400	-	-
R3.M2	*E. coli*	26±5	3.2±0.7	13±3	9±2	21±4	4.4±0.8	11±3	11±2	0.1±0.0	1.6±0.5	4100±400	-	-
R3.M3	*E. coli*	25±6	3±1	19±5	5±1	23±7	4±1	12±4	7±2	0.2±0.1	3±1	2300±300	-	-
R3.M4	*E. coli*	50±3	3.9±0.5	9±1	5.6±0.7	13±1	1.8±0.2	8.9±0.7	5.8±0.5	0.1±0.0	1.7±0.1	4300±300	-	-
R3.M5	*E. coli*	3±1	0.7±0.6	2.5±0.5	0.0±0.0	17±3	0.9±0.2	53±5	13±2	7±7	3.7±0.5	380±40	-	-
R3.M6	*E. coli*	9.4±0.8	1.4±0.1	11.9±0.8	4.3±0.3	36±2	6.3±0.6	12.2±0.9	16±1	0.2±0.0	2.5±0.2	3400±100	-	-
R3.M7	*E. coli*	20±9	6±3	29±10	11±5	16±7	4±2	7±3	5±2	0.1±0.1	1.7±0.8	3100±600	-	-
R3.M8	*E. coli*	16±2	6.1±0.9	33±4	14±2	13±2	5.4±0.9	7±1	4.1±0.9	0.1±0.0	1.8±0.4	3900±200	-	-
R3.M9	*E. coli*	21±3	4.7±0.4	22±2	7.7±0.8	21±2	4.1±0.4	11±2	6.3±0.9	0.2±0.0	2.9±0.4	2500±100	-	-
R3.M10	*E. coli*	3±1	1.0±0.3	1.8±0.6	0.0±0.0	15±5	0.8±0.3	58±9	15±5	1.0±0.3	4±2	390±70	-	-
R3.RD1	*E. coli*	20±20	4±3	20±10	7±4	20±30	4±5	10±10	10±10	0.2±0.2	2±4	1300±600	-	-
R3.RD2	*E. coli*	19.2±0.4	4.8±0.1	20.5±0.6	7.3±0.2	24.7±0.6	4.3±0.1	11±1	6.5±0.3	0.2±0.0	2.0±0.1	1980±30	-	-
R3.RD3	*E. coli*	10±3	3±1	35±9	13±5	18±6	6±3	7±3	6±2	0.1±0.0	1.5±0.6	3300±500	-	-
R3.RD4	*E. coli*	13±1	2.1±0.1	22±1	7.3±0.3	29±2	6.9±0.4	9±1	7.9±0.6	0.1±0.0	1.8±0.2	4600±100	-	-
R3.RD5	*E. coli*	21±4	5±1	21±4	8±2	25±5	5±1	8±2	6±2	0.1±0.0	1.8±0.5	2700±300	-	-
R4.M1	*E. coli*	10.6±0.4	2.6±0.2	13.0±0.9	4.1±0.4	34±1	5.2±0.5	14.3±0.4	11.5±0.7	0.0±0.0	2.2±0.1	2850±60	-	-
R4.M2	*E. coli*	23±1	4.1±0.2	12.3±0.3	9.0±0.3	24±1	4.2±0.1	10.6±0.7	11.5±0.6	0.0±0.0	1.3±0.1	3430±70	-	-
R4.M3	*E. coli*	3.1±0.2	0.9±0.0	9.1±0.3	3.8±0.1	43.8±0.9	7.1±0.2	8.9±0.3	19.7±0.7	0.0±0.0	2.1±0.1	3370±60	-	-
R4.M4	*E. coli*	6.8±0.6	1.2±0.1	8.8±0.9	3.9±0.4	42±3	6.2±0.6	8.4±0.8	19±2	0.0±0.0	1.8±0.3	4700±200	-	-
R4.M5	*E. coli*	11±1	3.0±0.2	15±1	8.9±0.6	30±3	4.3±0.5	11±1	14±2	0.0±0.0	1.8±0.3	4700±200	-	-
R4.M6	*E. coli*	6.0±0.6	2.3±0.2	12.5±0.8	5.9±0.4	38±2	6.5±0.5	8.3±0.6	17±1	0.0±0.0	2.0±0.2	3800±100	-	-
R4.M7	*E. coli*	2.7±0.4	1.2±0.1	17±2	2.7±0.4	37±2	4.9±0.6	17.6±0.9	11.1±0.4	0.0±0.0	2.9±0.1	2200±70	-	-
R4.M8	*E. coli*	12.0±0.8	2.4±0.1	18.6±0.7	6.7±0.3	29.7±0.9	5.2±0.2	13.2±0.6	10.2±0.4	0.0±0.0	1.9±0.1	2890±50	-	-
R4.M9	*E. coli*	3.5±0.4	1.0±0.0	20±1	4.7±0.5	35.9±0.9	4.8±0.3	15.7±0.9	9.5±0.3	0.0±0.0	2.6±0.2	2550±60	-	-
R4.M10	*E. coli*	10.7±0.4	2.6±0.1	8.3±0.3	5.5±0.4	28.3±0.8	6.6±0.3	17.9±0.5	15.7±0.4	0.0±0.0	2.0±0.1	2280±30	-	-
R4.M11	*E. coli*	4.3±0.7	0.0±0.0	3.4±0.8	4.2±0.9	15±2	28±3	4.6±0.3	30±2	0.0±0.0	2.9±0.3	1110±60	-	-
R4.M12	*E. coli*	14.5±0.5	0.0±0.0	3.6±0.2	0.0±0.0	17.5±0.8	0.0±0.0	35.9±0.7	12.8±0.3	0.0±0.0	3.3±0.1	660±10	-	-

R4.M13	*E. coli*	37.9±0.8	3.1±0.1	4.7±0.2	4.4±0.2	17.0±0.6	4.7±0.2	13.4±0.6	12.2±0.3	0.0±0.0	1.5±0.1	2830±40	-	-
R4.M14	*E. coli*	20.0±0.5	2.6±0.1	15.3±0.4	8.5±0.4	26±1	6.1±0.3	6.5±0.3	12.7±0.8	0.0±0.0	1.1±0.1	2910±50	-	-
R4.M15	*E. coli*	33.2±0.7	3.6±0.2	6.0±0.4	5.8±0.3	20±1	5.8±0.4	10.9±0.4	12.0±0.9	0.0±0.0	1.2±0.1	3490±70	-	-
R4.M16	*E. coli*	9.6±0.4	1.5±0.1	8.9±0.4	3.7±0.2	39.3±0.9	4.9±0.2	11.5±0.3	16.5±0.6	0.0±0.0	2.0±0.1	4000±70	-	-
R4.M17	*E. coli*	18±1	2.0±0.2	10.9±0.9	7.5±0.5	26±1	6.8±0.5	13.3±0.4	12.8±0.7	0.0±0.0	1.6±0.1	3050±80	-	-
R4.M18	*E. coli*	15±1	1.7±0.1	9.2±0.6	6.2±0.3	28.0±0.9	6.8±0.3	14.4±0.5	14.9±0.5	0.0±0.0	1.7±0.1	3150±60	-	-
R4.RD1	*E. coli*	18±2	2.6±0.1	8.7±0.4	6.7±0.3	29±1	4.1±0.2	12.7±0.8	16.2±0.8	0.0±0.0	1.7±0.1	3230±90	-	-
R4.RD2	*E. coli*	35±1	5.2±0.1	16.4±0.4	12.7±0.3	16.8±0.3	3.5±0.1	3.9±0.2	5.5±0.2	0.0±0.0	0.8±0.1	3800±60	-	-
R4.RD3	*E. coli*	39±1	5.0±0.2	10.8±0.5	8.4±0.2	19.5±0.7	3.5±0.1	6.1±0.5	7.2±0.4	0.0±0.0	1.0±0.1	3500±60	-	-
R4.RD4	*E. coli*	22±2	3.0±0.2	7.9±0.6	6.6±0.4	21±2	5.0±0.3	17±1	13.5±0.9	0.0±0.0	2.0±0.2	2690±90	-	-
R4.RD5	*E. coli*	34±2	5.6±0.3	7.7±0.3	7.7±0.3	15.4±0.5	4.3±0.1	12.6±0.6	9.3±0.3	0.0±0.0	1.5±0.1	2910±90	-	-
R4.RD6	*E. coli*	29.5±0.6	4.5±0.2	7.2±0.3	6.7±0.4	16.9±0.4	4.1±0.2	15.7±0.6	10.2±0.7	0.0±0.0	1.8±0.1	2400±30	-	-
R4.RD7	*E. coli*	3.4±0.4	1.6±0.2	13.3±0.5	0.8±0.1	12±1	1.1±0.2	10.0±0.4	43±1	0.7±0.0	3.7±0.4	540±10	-	-
R4.RD8	*E. coli*	9.1±0.6	2.5±0.1	21±1	6.5±0.3	31±2	5.0±0.3	11±1	12±1	0.0±0.0	2.0±0.3	4100±200	-	-
R4.RD9	*E. coli*	6.5±0.2	2.0±0.0	21.4±0.4	11.9±0.3	27.9±0.5	6.6±0.2	8.4±0.2	11.8±0.5	0.0±0.0	1.2±0.0	2420±30	-	-
R4.RD10	*E. coli*	4.5±0.2	1.3±0.0	27.1±0.6	10.6±0.2	28.3±0.8	8.4±0.2	6.6±0.3	9.8±0.4	0.0±0.0	1.2±0.1	2570±40	-	-
R4.RD11	*E. coli*	4.8±0.6	0.0±0.0	26±3	10±1	26±2	7.5±0.8	15±1	10.3±0.7	0.0±0.0	1.6±0.1	1670±80	-	-
RM.M1	*E. coli*	4.5	1.6	15.2	4.4	47.7	6.1	15.8	0.2	0.1	3.2	3100	a	-
RM.M2	*E. coli*	3.9	1.2	21.0	8.4	34.6	8.2	17.9	0.3	0.2	2.3	1700	a	-
RM.M3	*E. coli*	2.6	0.7	17.9	5.8	41.0	7.2	19.7	0.2	0.2	2.8	2500	a	-
RM.M4	*E. coli*	1.4	0.6	15.0	2.8	54.7	3.4	13.1	0.3	0.4	5.4	1100	a	-
RM.M5	*E. coli*	4.0	2.4	3.8	nd	21.9	nd	5.2	nd	2.1	3.2	120	a	-
RM.M6	*E. coli*	4.9	1.3	10.9	3.7	49.4	6.4	19.1	0.2	0.1	2.8	3100	a	-
RM.M7	*E. coli*	2.5	0.7	16.8	5.5	42.2	6.2	20.7	0.2	0.2	2.9	2300	a	-
RM.M8	*E. coli*	3.9	1.0	18.6	7.2	35.9	8.1	19.4	0.3	0.3	3.1	1500	a	-
RM.M9	*E. coli*	3.4	2.3	4.2	nd	24.6	nd	7.6	nd	1.8	2.6	130	a	-
RM.M10	*E. coli*	2.9	1.0	19.0	7.5	35.0	8.1	20.0	0.6	0.3	3.2	1200	a	-
RM.M11	*E. coli*	4.1	2.4	3.9	nd	23.8	nd	5.2	1.8	1.8	2.9	120	a	-
RM.M12	*E. coli*	5.5	1.2	18.6	3.5	48.9	3.7	12.7	0.1	0.2	4.4	3300	a	-
RM.M13	*E. coli*	3.7	nd	3.9	nd	32.0	nd	5.1	1.7	1.8	2.6	140	a	-
RM.M14	*E. coli*	2.6	1.1	19.9	8.4	33.2	8.3	17.2	0.8	0.4	2.7	850	a	-
RM.M15	*E. coli*	3.2	0.6	11.8	5.0	44.4	11.0	20.4	0.2	0.2	2.0	3300	a	-
RM.M16	*E. coli*	3.2	1.1	19.5	7.7	35.2	8.0	18.2	0.5	0.4	2.7	1100	a	-
RM.M17	*E. coli*	3.0	0.9	15.6	5.8	41.0	7.5	20.6	0.4	0.2	2.6	1600	a	-
RM.M18	*E. coli*	3.2	1.1	19.1	7.3	36.5	8.3	18.1	0.4	0.4	2.6	1200	a	-
RM.M19	*E. coli*	3.4	1.0	19.3	7.3	35.2	7.5	19.9	0.6	0.4	3.2	1300	a	-
RM.M20	*E. coli*	3.1	0.9	18.3	8.1	38.1	8.6	18.4	0.2	0.2	2.0	1600	a	-
RM.M21	*E. coli*	2.5	0.7	16.6	3.6	49.3	5.5	15.9	0.1	0.2	4.2	3200	a	-

RM.M22	*E. coli*	1.4	0.7	11.0	0.5	66.6	1.4	8.7	0.4	0.4	5.6	780	a	-
RM.M23	*E. coli*	4.6	2.7	4.1	nd	23.2	nd	6.9	nd	2.0	3.0	120	a	-
RM.M24	*E. coli*	3.8	1.6	7.5	nd	41.8	1.1	6.3	1.1	1.2	1.5	200	a	-
RM.M25	*E. coli*	2.7	0.9	23.4	4.1	45.9	4.4	14.2	0.1	0.1	2.9	3800	a	-
RM.M26	*E. coli*	3.9	2.5	3.6	nd	20.3	nd	6.2	nd	2.3	3.1	110	a	-
RM.M27	*E. coli*	3.4	0.8	19.1	4.8	46.6	6.9	13.7	0.1	0.1	3.5	3600	a	-
RM.M28	*E. coli*	3.9	2.4	3.1	nd	18.9	nd	6.2	1.8	2.2	3.0	110	a	-
RM.M29	*E. coli*	21.9	1.7	16.3	2.7	38.0	3.0	11.3	0.2	0.2	3.8	2500	a	-
RM.M30	*E. coli*	3.5	0.8	17.9	7.0	39.8	7.4	19.0	0.2	0.2	2.2	2000	a	-
RM.M31	*E. coli*	5.9	1.6	7.9	2.1	43.9	6.3	13.0	0.5	1.0	4.1	400	a	-
RM.M32	*E. coli*	2.9	0.9	17.6	6.3	40.3	7.0	19.1	0.5	0.3	2.6	1700	a	-
RM.M33	*E. coli*	4.9	2.3	4.0	nd	21.6	nd	6.5	nd	1.9	2.7	120	a	-
RM.M34	*E. coli*	3.1	1.0	20.9	7.8	36.4	8.2	17.8	0.2	0.3	2.2	1600	a	-
RM.M35	*E. coli*	3.5	1.0	16.4	7.5	38.1	7.4	18.0	0.6	0.5	2.7	920	a	-
RM.M36	*E. coli*	3.2	1.1	21.7	7.4	38.9	6.6	17.8	0.1	0.1	1.9	2500	a	-
RM.M37	*E. coli*	3.4	1.2	20.5	7.7	35.1	7.9	16.9	0.5	0.4	2.5	1000	a	-
RM.M38	*E. coli*	2.2	0.8	19.1	5.3	43.0	6.2	18.9	0.2	0.2	3.0	2700	a	-
RM.M39	*E. coli*	4.9	2.3	31.6	12.4	28.1	6.9	10.4	0.3	0.2	1.9	2100	a	-
RM.M40	*E. coli*	2.5	0.8	19.1	3.9	48.3	5.0	15.8	0.1	0.1	3.5	3500	a	-
RM.M41	*E. coli*	2.2	0.7	18.6	5.7	43.8	7.3	17.7	0.3	0.2	2.6	3200	a	-
RM.M42	*E. coli*	5.0	1.1	9.2	4.7	42.8	15.5	18.7	0.2	0.2	1.4	3500	a	-
RM.M43	*E. coli*	3.5	1.1	20.0	7.8	35.4	8.6	16.8	0.5	0.4	2.4	960	a	-
RM.M44	*E. coli*	3.2	1.0	20.0	7.6	37.3	8.2	17.7	0.3	0.3	2.3	1500	a	-
RM.M45	*E. coli*	4.3	2.4	3.9	nd	21.7	nd	6.7	1.8	2.3	2.7	120	a	-
RM.M46	*E. coli*	4.6	2.6	3.9	nd	22.0	nd	6.5	nd	2.0	2.7	110	a	-
RM.M47	*E. coli*	1.4	0.7	10.4	0.5	63.6	1.2	8.0	0.6	0.5	6.9	540	a	-
RM.M48	*E. coli*	1.9	0.7	18.5	5.4	43.6	7.0	18.6	0.3	0.2	2.7	2700	a	-
RM.M49	*E. coli*	2.6	1.5	8.8	nd	51.2	1.1	6.5	nd	1.4	1.5	210	a	-
RM.M50	*E. coli*	4.5	2.4	3.6	nd	21.4	nd	6.3	nd	2.0	2.6	120	a	-
RM.M51	*E. coli*	4.5	2.4	3.7	nd	21.8	nd	6.3	nd	2.1	2.7	120	a	-
RM.M52	*E. coli*	3.1	1.1	19.3	7.2	37.5	7.7	19.2	0.2	0.3	2.4	1400	a	-
RM.M53	*E. coli*	3.1	1.4	18.4	7.4	32.9	8.1	17.2	0.4	0.6	3.5	550	a	-
RM.M54	*E. coli*	3.9	2.1	3.9	nd	30.3	nd	6.7	nd	1.7	2.4	130	a	-
RM.M55	*E. coli*	2.9	1.1	20.6	8.0	35.3	8.1	17.6	0.5	0.4	2.5	1100	a	-
RM.M56	*E. coli*	4.3	1.6	20.1	7.8	27.1	7.0	16.9	0.4	0.7	4.0	540	a	-
RM.M57	*E. coli*	4.4	3.7	23.0	19.3	27.9	5.5	12.9	0.1	0.2	1.6	1900	a	-
RM.M58	*E. coli*	2.8	1.3	21.9	8.5	33.0	7.6	17.6	0.3	0.5	2.8	930	a	-
RM.M59	*E. coli*	4.3	1.1	12.7	5.1	42.8	7.7	21.4	0.2	0.2	2.3	1600	a	-

RM.M60	*E. coli*	2.8	0.9	23.0	6.8	38.5	6.6	17.5	0.3	0.1	2.4	2900	a	-
RM.M61	*E. coli*	4.3	1.6	20.2	8.2	29.8	7.4	15.6	0.6	0.7	3.7	570	a	-
AAC49179	*C. palustris*	97.5±0.2	0.9±0.1	0.6±0.1	0.2±0.0	0.2±0.1	0.2±0.0	nd	0.1±0.1	nd	nd	710±50	-	[30]
AAB71731	*U. americana*	44±3	23±1	3.7±0.4	7.9±0.9	9.8±0.8	1.5±0.2	1.4±0.1	4.4±0.7	nd	nd	1100±60	-	[30]
AAG43857	*I. germanica*	3.3±0.4	0.5±0.1	0.7±0.1	1.0±0.1	30±3	0.3±0.1	20±3	44±4	nd	nd	260±20	-	[30]
AAG43858	*I. germanica*	8±4	nd	1±1	nd	30±10	nd	20±10	40±20	nd	nd	15±5	-	[30]
EER87824	*S. bicolor*	4.8±0.7	0.2±0.0	1.6±0.3	0.4±0.1	46±5	nd	13±3	35±6	nd	nd	130±10	-	[30]
EER88593	*S. bicolor*	6±1	0.7±0.1	3.2±0.5	0.6±0.3	45±5	3.3±0.9	11±2	31±4	nd	nd	91±8	-	[30]
AEM72519	*C. nucifera*	14±3	1.1±0.3	1.4±0.3	1.1±0.1	44±5	0.4±0.3	6±1	31±5	nd	nd	130±10	-	[30]
AEM72520	*C. nucifera*	1.7±0.2	0.1±0.0	1.0±0.2	0.3±0.1	36±3	0.7±0.1	16±2	44±3	nd	nd	570±30	-	[30]
AEM72521	*C. nucifera*	11.1±0.8	1.2±0.1	34±2	6.1±0.6	14±2	23±3	2±1	9±3	nd	nd	200±10	-	[30]
AEM72522	*C. viscosissima*	52±6	26±5	7±2	6±2	7±2	0.9±0.6	nd	nd	nd	nd	79±10	-	[30]
AEM72523	*C. viscosissima*	4±1	0.5±0.2	1.0±0.3	0.9±0.4	47±2	0.7±0.5	26±2	19±2	nd	nd	249±9	-	[30]
AEM72524	*C. viscosissima*	7±5	5±3	2±2	2±2	84±6	nd	nd	nd	nd	nd	19±2	-	[30]
AAD42220	*E. guineensis*	14±4	1.9±0.8	2±1	2±1	47±5	3±3	nd	26±5	nd	nd	37±4	-	[30]
EDQ65090	*P. patens*	9±1	0.4±0.1	0.4±0.1	0.2±0.0	42±4	0.2±0.1	16±3	32±4	nd	nd	380±30	-	[30]
EER96252	*S. bicolor*	5.8±0.9	1.8±0.8	1.2±0.3	1.1±0.4	34±3	0.9±0.2	18±3	37±3	nd	nd	180±10	-	[30]
EES11622	*S. bicolor*	5±3	nd	nd	nd	50±10	nd	15±8	29±9	nd	nd	9±2	-	[30]
EEH52851	*M. pusilla*	4±2	nd	0±1	0.6±0.1	65±4	8±4	nd	23±3	nd	nd	16±2	-	[30]
ACL08376	*D. vulgaris*	29±1	3.5±0.2	7.9±0.3	24±1	6.0±0.4	24±1	1.2±0.4	2.6±0.3	nd	nd	330±9	-	[30]
CAH09236	*B. fragilis*	20±2	2.7±0.2	3.6±0.4	19±1	5.1±0.3	25±1	2.2±0.3	5.4±0.4	nd	nd	215±6	-	[30]
ABR43801	*P. distasonis*	18±5	6.3±0.4	16±1	9.3±0.8	21±2	27±2	nd	nd	nd	nd	70±4	-	[30]
AAO77182	*B. thetaiotaomicron*	13.4±0.8	2.1±0.2	4.6±0.7	16.7±0.9	6±1	26±1	nd	2.6±0.3	nd	nd	60±3	-	[30]
ABG82470	*C. perfringens*	70±4	3.0±0.5	nd	1.1±0.2	nd	9±2	nd	nd	nd	nd	72±10	-	[30]
EEG55387	*C. asparagiforme*	26±6	6±1	7±2	1.6±0.8	35±9	18±6	nd	1.3±0.9	nd	nd	26±4	-	[30]
EET61113	*M. formatexigens*	31.8±0.3	5.1±0.1	4.3±0.2	8.9±0.5	1.9±0.2	10.5±0.2	0.4±0.2	1.2±0.1	nd	nd	381±3	-	[30]
EDV77528	*G. sp.*	9±4	2±1	7±3	30±10	11±3	32±8	0±1	2±1	nd	nd	60±10	-	[30]
BAH81730	*S. dysgalactiae*	30±1	5.0±0.2	5.7±0.3	13.5±0.7	4.4±0.3	20±1	0.3±0.1	3.3±0.1	nd	nd	620±10	-	[30]
ABJ63754	*L. brevis*	55.5±0.7	2.6±0.1	3.8±0.1	7.9±0.2	1.9±0.1	6.3±0.2	nd	0.7±0.1	nd	nd	710±10	-	[30]
CAD63310	*L. plantarum*	68.0±0.8	1.2±0.1	2.8±0.2	4.6±0.2	1.9±0.1	6.9±0.4	nd	0.5±0.1	nd	nd	440±10	-	[30]
EEI82564	*A. tetradius*	87±2	2.2±0.3	1.1±0.2	2.8±0.5	1.2±0.2	3.0±0.5	0.1±0.1	0.7±0.1	nd	nd	1400±100	-	[30]
CAE80300	*B. bacteriovorus*	37±3	3.3±0.5	6.7±0.7	7.6±0.6	8.2±0.7	28±2	1.6±0.2	6.6±0.5	nd	nd	330±20	-	[30]
ABN54268	*R. thermocellum*	8.4±0.4	4.5±0.2	2.7±0.2	7.9±0.4	9.8±0.8	60±1	0.8±0.7	4.7±0.4	nd	nd	98±3	-	[30]
Q9SJE2	*A. thaliana*	nd	nd	nd	nd	19.4	nd	53.0	14.9	1.0	10.9	100	a,c	[181]
AAX51637	*M. longifolia*	nd	nd	nd	nd	7±4	nd	66±5		15±2	10±3	nd	b,c,d	[182]
AAX51636	*D. butyracea*	nd	nd	nd	nd	17.8±0.6	nd	71±4	0.6±0.1	2.6±0.1	5.2±0.4	nd	b	[183]

AHF72806	*L. communis*	nd	6±2	17±6	nd	16±2	nd	17±2	20±3	1.5±0.2	15±2	nd	b	[184]
EST75919	*E. coli*	2.7	0.8	12.8	2.2	49.5	2.6	11.9	11.9	0.9	4.0	1400	a,b	[54]
AAC49151	*C. camphorum*	nd	nd	5.8	0.6	27.6	12.4	24.2	16.9	1.2	11.3	1600	a	[67]
Q41635	*U. californica*	nd	nd	63.2	13.7	3.0	10.2	1.6	4.3	nd	4.0	1000	a,b	[29]
ADB79567	*A. hypogaea L.*	nd	nd	4±1	nd	8±2	nd	45±3	14±3	2.3±0.3	5.8±0.9	nd	b	[185]
AAX51636	*D. butyracea*	nd	nd	8.3	nd	6.8	nd	43.6	5.1	29.7	6.5	580±40	a,b	[60]
Q9SQI3	*G. hirsutum*	nd	nd	0.1	nd	35.6	nd	21.3	36.6	1.4	4.9	3100±200	a,b	[60]
XP002515564	*R. communis*	nd	nd	0.1	nd	44.4	nd	17.0	32.9	1.2	4.4	9200±100	a,b	[60]
ABU96744	*J. curcas*	nd	nd	0.0	nd	44.4	nd	17.5	32.5	1.2	4.4	8700±200	a,b	[60]
WP004921669	*A. baylyi*	20±7	9±2	15±2	8.0±0.3	17±1	6±1	8±2	10±4	2.3±0.1	1.6±0.4	18000±2000	b	[186]
CAA57794	*A. brasilense*	nd	nd	nd	nd	5.6±0.3	nd	53±3	9.2±0.6	0.9±0.0	29±2	nd	b	[187]
EEC51251	*P. tricornutum*	nd	nd	nd	nd	2.8±0.2	nd	43.0±0.2	14.4±0.3	2.2±0.3	14.2±0.0	nd	b	[188]
AAL79361	*H. annuus L.*	nd	nd	nd	nd	nd	nd	39.9	3.6	21.6	10.9	nd	a	[189]
ADA64920	*L. lactis*	3.6±0.6	1.7±0.2	2.3±0.3	nd	4.6±0.6	nd	51±3	6.6±0.2	26±2	4.5±0.5	2800±100	-	[190]
AAC72882	*C. hookeriana*	93.5	2.2	0.9	0.2	0.0	0.1	1.2	0.0	nd	0.0	930	a	[191]
rTE3	*C. viscosissima*	42.5	26.6	4.1	9.5	1.4	1.6	1.3	0.8	nd	nd	510	a	[191]
rTE4	*C. viscosissima*	24.3	36.0	4.1	12.2	3.6	3.8	0.0	3.0	nd	nd	490	a	[191]
rTE8	*C. viscosissima*	17.9	10.2	3.4	5.3	31.6	6.4	0.0	17.9	nd	nd	87	a	[191]
rTE12	*C. viscosissima*	18.5	43.0	2.8	15.6	9.9	0.0	0.0	0.0	nd	nd	50	a	[191]
rTE15	*C. viscosissima*	4.1	21.0	14.7	34.3	5.1	10.0	0.0	0.0	nd	nd	56	a	[191]
rTE16	*C. viscosissima*	38.9	15.3	9.3	16.2	2.7	7.5	0.0	1.8	nd	nd	820	a	[191]
rTE20	*C. viscosissima*	16.1	19.9	12.4	24.3	2.8	13.2	0.0	2.2	nd	nd	390	a	[191]
rTE24	*C. viscosissima*	19.8	4.1	6.2	7.3	27.8	6.2	1.0	23.9	nd	nd	280	a	[191]
rTE28	*C. viscosissima*	4.8	1.0	1.3	1.7	41.4	1.8	2.3	44.8	nd	nd	280	a	[191]
rTE32	*C. viscosissima*	25.5	24.6	11.0	16.1	4.8	4.2	0.0	4.3	nd	nd	490	a	[191]
rTE36	*C. viscosissima*	6.9	18.5	13.1	17.6	11.7	17.4	0.0	9.2	nd	nd	190	a	[191]
rTE40	*C. viscosissima*	16.7	3.1	6.0	4.7	32.0	10.6	0.0	20.1	nd	nd	140	a	[191]
rTE44	*C. viscosissima*	6.8	1.7	1.3	1.1	46.4	5.8	0.0	31.3	nd	nd	47	a	[191]
rTE48	*C. viscosissima*	23.8	17.8	8.9	18.3	1.7	5.8	0.8	0.8	nd	nd	810	a	[191]
rTE51	*C. viscosissima*	0.0	20.9	2.4	30.0	21.8	2.1	0.0	0.0	nd	nd	6.4	a	[191]
rTE52	*C. viscosissima*	31.0	11.4	5.3	12.3	1.6	4.6	0.0	1.0	nd	nd	870	a	[191]
rTE56	*C. viscosissima*	24.9	5.6	7.5	8.2	26.2	4.6	0.8	14.8	nd	nd	410	a	[191]
rTE60	*C. viscosissima*	3.6	1.0	1.1	0.9	54.9	0.1	10.1	27.4	nd	nd	140	a	[191]
CvB2MT1	*C. viscosissima*	34.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	nd	nd	2.4	a	[191]
CvB2MT2	*C. viscosissima*	8.9	4.6	10.0	9.7	24.5	19.9	0.1	19.5	nd	nd	82	a	[191]
CvB2MT3	*C. viscosissima*	17.2	4.1	2.1	4.6	40.1	4.3	0.0	21.6	nd	nd	90	a	[191]
CvB2MT4	*C. viscosissima*	3.7	0.6	0.0	0.2	60.9	0.0	9.3	24.5	nd	nd	200	a	[191]

CvB2MT5	*C. viscosissima*	1.5	0.2	0.0	0.0	81.5	0.0	2.9	13.9	nd	nd	460	a	[191]
CvB2MT6	*C. viscosissima*	38.2	5.8	21.7	9.1	3.4	13.9	0.0	0.0	nd	nd	31	a	[191]
CvB2MT7	*C. viscosissima*	6.1	0.6	0.0	0.0	62.2	0.0	0.9	28.8	nd	nd	47	a	[191]
CvB2MT8	*C. viscosissima*	3.9	0.3	0.0	0.0	58.0	0.0	4.4	32.8	nd	nd	74	a	[191]
CvB2MT9	*C. viscosissima*	1.5	0.3	0.3	0.4	52.6	0.6	10.4	33.7	nd	nd	410	a	[191]
CvB2MT10	*C. viscosissima*	21.6	9.6	8.8	14.1	9.8	16.4	0.0	9.9	nd	nd	170	a	[191]
CvB2MT11	*C. viscosissima*	4.3	1.5	3.8	4.0	45.5	4.9	11.9	23.5	nd	nd	120	a	[191]
CvB2MT12	*C. viscosissima*	12.5	7.2	16.4	15.4	15.3	20.0	0.0	7.8	nd	nd	150	a	[191]
CvB2MT13	*C. viscosissima*	12.7	2.1	0.9	1.5	50.7	0.5	3.1	26.0	nd	nd	180	a	[191]
CvB2MT14	*C. viscosissima*	3.4	0.8	1.6	0.8	52.2	0.7	14.1	26.0	nd	nd	110	a	[191]
CvB2MT15	*C. viscosissima*	16.1	7.5	15.6	19.8	9.2	19.8	0.7	4.0	nd	nd	150	a	[191]
CvB2MT16	*C. viscosissima*	15.4	5.6	3.9	9.6	22.8	30.9	0.0	7.0	nd	nd	310	a	[191]
CvB2MT17	*C. viscosissima*	26.2	6.5	17.6	13.5	8.6	15.8	5.0	1.6	nd	nd	110	a	[191]
CvB2MT18	*C. viscosissima*	21.9	8.4	13.6	19.9	6.1	15.3	2.3	3.2	nd	nd	300	a	[191]
CvB2MT19	*C. viscosissima*	18.2	6.3	10.3	22.0	8.5	18.4	0.6	6.3	nd	nd	170	a	[191]
CvB2MT20	*C. viscosissima*	26.2	16.9	6.6	19.1	2.4	6.2	0.3	1.4	nd	nd	690	a	[191]
CvB2MT21	*C. viscosissima*	17.8	3.4	4.6	2.7	43.8	4.5	2.0	17.5	nd	nd	200	a	[191]
CvB2MT22	*C. viscosissima*	20.7	9.2	13.0	15.6	8.9	16.7	0.0	10.0	nd	nd	85	a	[191]
CvB2MT23	*C. viscosissima*	5.6	1.0	0.8	0.8	54.2	0.9	6.6	28.7	nd	nd	380	a	[191]
CvB2MT24	*C. viscosissima*	3.6	0.4	0.0	0.3	73.6	0.2	1.4	20.0	nd	nd	400	a	[191]
CvB2MT25	*C. viscosissima*	28.8	9.3	13.0	15.9	5.2	13.7	2.0	0.6	nd	nd	290	a	[191]
CvB2MT26	*C. viscosissima*	31.7	1.2	4.8	0.0	17.9	3.9	11.3	5.7	nd	nd	41	a	[191]
CvB2MT27	*C. viscosissima*	20.8	9.9	10.3	21.5	5.3	17.6	0.3	3.2	nd	nd	230	a	[191]
CvB2MT28	*C. viscosissima*	17.6	5.0	7.5	3.7	40.1	4.1	3.4	14.3	nd	nd	330	a	[191]
CvB2MT29	*C. viscosissima*	11.4	7.2	15.0	18.1	12.1	20.5	0.4	8.9	nd	nd	290	a	[191]
CvB2MT30	*C. viscosissima*	29.2	17.2	6.1	15.7	1.5	4.7	0.4	0.6	nd	nd	1000	a	[191]
CvB2MT31	*C. viscosissima*	17.9	5.9	9.0	19.6	12.1	17.3	0.7	8.1	nd	nd	390	a	[191]
CvB2MT32	*C. viscosissima*	21.5	7.3	4.8	14.4	9.5	30.6	0.0	3.4	nd	nd	330	a	[191]
CvB2MT33	*C. viscosissima*	27.2	7.9	9.6	20.4	5.1	12.5	0.0	3.6	nd	nd	570	a	[191]
CvB2MT34	*C. viscosissima*	31.8	9.6	9.2	18.4	2.8	11.2	0.3	0.4	nd	nd	460	a	[191]
CvB2MT35	*C. viscosissima*	43.1	2.5	3.6	4.1	3.0	8.8	0.0	0.1	nd	nd	120	a	[191]
CvB2MT36	*C. viscosissima*	40.3	0.0	0.0	0.0	2.4	0.0	0.0	0.0	nd	nd	47	a	[191]
CvB2MT37	*C. viscosissima*	40.2	3.4	2.8	1.0	1.1	2.0	0.0	0.2	nd	nd	330	a	[191]
CvB2MT38	*C. viscosissima*	37.2	3.1	5.7	4.2	1.6	8.1	0.0	0.0	nd	nd	1000	a	[191]
CvB2MT40	*C. viscosissima*	61.4	12.5	3.3	6.2	0.6	0.9	0.0	0.0	nd	nd	1100	a	[191]
CvB2MT41	*C. viscosissima*	36.2	3.8	1.4	0.7	0.0	0.0	0.0	0.0	nd	nd	93	a	[191]
CvB2MT42	*C. viscosissima*	27.0	15.0	6.9	18.9	2.6	6.1	1.5	1.1	nd	nd	820	a	[191]
CvB2MT43	*C. viscosissima*	4.1	1.2	1.5	1.4	47.0	0.7	11.5	31.0	nd	nd	170	a	[191]

CvB2MT44	*C. viscosissima*	22.7	23.6	7.9	24.2	1.8	4.0	0.0	0.2	nd	nd	710	a	[191]
CvB2MT45	*C. viscosissima*	22.5	15.8	13.1	24.4	2.6	11.4	0.0	0.0	nd	nd	570	a	[191]
CvB2MT47	*C. viscosissima*	18.5	12.0	17.9	17.5	1.4	4.5	22.6	0.0	nd	nd	190	a	[191]
CvB2MT48	*C. viscosissima*	13.7	38.6	6.7	17.5	0.0	3.9	0.2	0.0	nd	nd	140	a	[191]
TEGm157	Hybrid	12.0	0.7	4.0	0.5	30.9	4.6	19.1	19.2	nd	7.6	600	a	[191]
TEGm162	Hybrid	38.5	5.3	11.7	2.2	19.4	11.6	3.1	0.6	nd	1.5	1700	a	[191]
TEGm169	Hybrid	33.7	4.6	11.7	2.8	18.8	14.6	3.8	1.9	nd	2.7	1400	a	[191]
TEGm171	Hybrid	17.6	0.8	6.4	3.1	26.2	5.9	12.1	23.0	nd	2.0	1100	a	[191]
TEGm173	Hybrid	2.8	0.4	3.5	0.5	35.9	3.1	12.0	36.2	nd	4.9	1200	a	[191]
TEGm181	Hybrid	3.7	0.7	4.8	0.5	35.5	4.6	15.1	29.2	nd	5.0	830	a	[191]
TEGm183	Hybrid	12.7	4.2	9.2	6.6	23.2	11.7	10.0	18.4	nd	2.4	1200	a	[191]
TEGm198	Hybrid	1.3	0.1	1.4	0.1	36.4	0.7	23.0	36.6	nd	0.0	710	a	[191]
TEGm200	Hybrid	1.8	0.4	4.7	3.7	19.2	8.3	16.6	41.5	nd	3.5	940	a	[191]
TEGm201	Hybrid	1.3	0.1	1.2	0.1	19.3	1.3	28.4	41.1	nd	7.0	590	a	[191]
TEGm202	Hybrid	2.8	0.3	4.0	0.8	30.1	2.9	17.9	36.4	nd	4.0	1000	a	[191]
TEGm203	Hybrid	0.6	0.1	2.4	0.0	45.1	2.2	24.1	19.6	nd	5.7	530	a	[191]
TEGm204	Hybrid	1.3	0.1	1.6	0.3	38.0	1.5	19.8	33.6	nd	3.4	1200	a	[191]
TEGm205	Hybrid	0.3	0.0	0.7	0.0	35.9	1.1	23.9	34.1	nd	3.9	890	a	[191]
TEGm215	Hybrid	9.0	0.7	5.6	1.7	30.6	6.5	15.2	24.3	nd	3.7	980	a	[191]
TEGm219	Hybrid	2.4	0.3	4.4	0.2	54.7	3.8	16.1	17.5	nd	0.0	530	a	[191]
TEGm245	Hybrid	1.0	0.0	0.7	0.0	32.1	0.0	27.7	38.5	nd	0.0	1100	a	[191]
TEGm250	Hybrid	20.0	4.3	3.8	2.0	19.8	3.5	19.1	17.1	nd	6.2	1600	a	[191]
TEGm258	Hybrid	50.4	6.4	4.7	1.8	9.1	1.3	10.2	1.9	nd	2.3	1600	a	[191]
TEGm288	Hybrid	30.9	5.5	9.8	8.5	12.9	7.1	12.1	8.3	nd	2.2	1300	a	[191]
TEGm413	Hybrid	14.8	6.6	13.6	7.6	14.2	5.9	14.7	16.9	nd	3.9	1100	a	[191]
TEGm419	Hybrid	13.2	7.6	13.7	12.2	15.2	8.0	8.9	15.2	nd	3.6	1400	a	[191]
TEGm492	Hybrid	2.5	1.5	8.6	0.0	21.9	2.8	25.7	24.0	nd	12.5	600	a	[191]
TEGm501	Hybrid	0.4	0.4	9.1	1.0	38.7	4.9	17.1	23.7	nd	4.6	870	a	[191]
TEGm520	Hybrid	2.4	3.1	8.8	10.3	24.0	15.0	11.4	20.6	nd	2.8	1000	a	[191]
TEGm546	Hybrid	1.3	5.2	21.1	7.1	22.1	11.7	14.9	12.2	nd	3.4	610	a	[191]
UcFatB-M1	*U. californica*	nd	nd	nd	nd	29±3	nd	24±3	32±2	10±1	5.3±0.8	140±10	b	[65]
UcFatB-M2	*U. californica*	nd	nd	nd	nd	33±4	nd	24±3	27±2	9.9±0.8	5.6±0.5	105±10	b	[65]
UcFatB-M3	*U. californica*	nd	nd	nd	nd	33±1	nd	32.7±0.6	21±1	8.4±0.4	5.8±0.4	111±4	b	[65]
UcFatB-M4	*U. californica*	nd	nd	nd	nd	15.4±0.6	nd	21±4	46±1	16±4	1.0±0.4	53±4	b	[65]
UcFatB-M5	*U. californica*	nd	nd	nd	nd	11.1±0.9	nd	42±6	6±2	41±4	0.4±0.4	24±2	b	[65]
UcFatB-M6	*U. californica*	nd	nd	nd	nd	12.9±0.9	nd	44±4	8.9±0.6	34±3	0.9±0.9	33±3	b	[65]
UcFatB-M7	*U. californica*	nd	nd	nd	nd	12±2	nd	45±6	10±1	30±4	2.6±0.5	38±4	b	[65]
UcFatB-M8	*U. californica*	nd	nd	nd	nd	12±2	nd	41±5	7±1	39±3	0.6±0.6	33±3	b	[65]

[a] Standard deviation not determined or provided; total FFA values with unknown standard deviation was reported to two significant Figures
[b] Data approximated using image processing software
[c] Mole-based values not provided and incapable of being calculated; mass-based values used instead[d] 16:0 and 16:1 fatty acid compositions combined

Appendix IV: Sequences of all variants considered in this study.
Mutant names match the formatting provided for Appendix III. “Changes” indicates the number of amino acid changes, or whether a frameshift or truncation occurred for the mutant. Amino acid substitutions are provided in the format *XyZ*, where *X* is the WT residue at position *y* and *Z* is the new amino acid. “stop” indicates a stop codon replaces the existing amino acid, resulting in a shorter peptide (a truncation). A frameshift indicates that a WT amino acid was eradicated and not replaced by a different residue.

Name	Changes	Mutations
R1.M1	1	Y145K
R1.M2	1	L11G
R1.M3	3	L11G, I107K, R108K
R1.M4	4	L11G, L76K, I107K, R108K
R1.M5	5	L11G, L76K, I107K, R108K, F139E
R1.M6	6	L11G, L76K, I107K, R108K, F139E, Y145R
R1.M7	6	L11G, L76K, I107K, R108K, F139E, Y145K
R1.M8	7	L11G, L76K, I107K, R108K, A111E, F139E, Y145R
R1.M9	5	I107K, R108K, A111E, F139E, Y145R
R1.M10	7	L11G, L76K, I107K, R108K, A111E, F139E, Y145K
R1.M11	5	I107K, R108K, A111E, F139E, Y145K
R1.M12	5	I107K, R108K, A111E, F139H, Y145K
R2.M1	3	G72F, F139W, Y145F
R2.M2	7	G72L, F139P, M141W, E142L, Y145F, G155A, I156L
R2.M3	5	L11W, G72F, F139W, E142P, Y145F
R2.M4	7	G72L, F139P, M141W, E142P, Y145P, G155W, I156L
R2.M5	2	E142P, Y145F

R2.M6	4	L11F, G72F, F139M, Y145F
R2.M7	4	G72F, Y145F, G155W, I156L
R2.M8	7	G72L, F139P, M141W, E142P, Y145P, G155A, I156L
R2.M9	5	L11F, G72F, F139W, E142P, Y145F
R2.M10	6	G72L, F139P, M141W, E142P, G155A, I156L
R2.M11	2	L11P, Y145F
R2.M12	4	L11F, G72F, F139W, Y145F
R2.M13	2	G72F, Y145F
R2.M14	3	G72A, E142P, Y145P
R2.RD1	1	G72A
R2.RD2	1	G72L
R2.RD3	1	L11P
R2.RD4	2	L11P, G72F
R2.RD5	1	L11W
R3.M1	3	S122K, Y145K, L146K
R3.M2	1	M141L
R3.M3	4	I107L, R108K, Y145K, L146K
R3.M4	3	M141L, Y145K, L146K
R3.M5	5	I107K, R108F, L109I, Y145K, L146K
R3.M6	1	R108K
R3.M7	2	Y145K, L146K
R3.M8	2	S122K, Y145K
R3.M9	3	S122L, Y145K, L146K
R3.M10	4	S122K, M141K, Y145K, L146K
R3.RD1	2	I107L, R108K
R3.RD2	3	I107K, R108F, L109I

R3.RD3	1	S122K
R3.RD4	1	S122L
R3.RD5	2	E142P, Y145K
R4.M1	3	I107M, R108F, E142P
R4.M2	3	M141L, E142P, Y145F
R4.M3	2	S122M, E142P
R4.M4	1	E142P
R4.M5	4	S122M, M141L, E142P, Y145F
R4.M6	2	E142P, Y145F
R4.M7	1	I107G
R4.M8	1	M141G
R4.M9	1	I107K
R4.M10	3	R108F, L109F, E142P
R4.M11	4	R108F, L109F, S122I, E142P
R4.M12	4	R108F, L109F, M141G, Y145F
R4.M13	5	R108F, L109F, S122M, Y145F, L146P
R4.M14	3	R108F, Y145F, L146P
R4.M15	3	R108F, L109F, Y145F
R4.M16	3	R108K, L109I, L146K
R4.M17	3	R108F, L109F, S122M
R4.M18	2	R108F, L109F
R4.RD1	2	M141L, E142P
R4.RD2	2	M141L, Y145K
R4.RD3	4	M141L, E142P, Y145K, L146K
R4.RD4	3	R108F, L109F, M141L
R4.RD5	6	R108F, L109F, M141L, E142P, Y145K, L146K

R4.RD6	5	R108F, L109F, M141L, Y145K, L146K
R4.RD7	1	H157A
R4.RD8	2	I107M, R108F
R4.RD9	1	R108F
R4.RD10	1	S122I
R4.RD11	1	S122M
RM.M1	2	S36T, P110T
RM.M2	1	L146I
RM.M3	2	A19S, A98P
RM.M4	3	Q49H, M105K, D161N
RM.M5	truncation	G8C, M105L, Y117stop
RM.M6	1	P110S
RM.M7	1	L57P
RM.M8	1	A123T
RM.M9	3	G72V, D74N, F79Y
RM.M10	1	Q93H
RM.M11	truncation	Q106stop
RM.M12	1	P158H
RM.M13	2	S41N, A162V
RM.M14	1	P148S
RM.M15	2	F121Y, P126L
RM.M16	1	R64H
RM.M17	2	P102A, A111V
RM.M18	2	F79Y, Q82H
RM.M19	1	L57Q
RM.M20	2	A22T, A129T

RM.M21	1	L136P
RM.M22	3	T3M, M17I, E69G
RM.M23	truncation	W33stop
RM.M24	frameshift	Q90single nucleotide deletion
RM.M25	2	G44C, A56T
RM.M26	frameshift	A25T, frameshift after K127
RM.M27	1	S12I
RM.M28	3	W23C, V68L, G75S
RM.M29	3	V38F, F139Y, Y145C
RM.M30	2	A22V, A171V
RM.M31	4	L109P, Y117N, A123V, K147E
RM.M32	2	N28K, A123V
RM.M33	frameshift	G8V frameshift resulting in stop codon at 10^{th} amino acid position
RM.M34	1	P55L
RM.M35	2	S41G, L92M
RM.M36	2	S47P, M170I
RM.M37	1	P135S
RM.M38	1	A25P
RM.M39	2	T35S, E142D
RM.M40	4	L92S, V95I, E101D, W169L
RM.M41	1	L92S
RM.M42	5	I42V, G75A, Q106R, A111G, P138S
RM.M43	1	L67M
RM.M44	1	A120V
RM.M45	truncation	E69stop
RM.M46	truncation	E119stop

RM.M47	1	S41N
RM.M48	1	L58R
RM.M49	truncation	A19T, W31stop
RM.M50	3	M105K, R115C, P135del
RM.M51	truncation	truncation one nucleotide after A40 resulting also in frameshift
RM.M52	2	M170I, A171V
RM.M53	1	W150L
RM.M54	2	A25T, S41I
RM.M55	2	P135T, D161N
RM.M56	1	D2V
RM.M57	2	A111T, M141I
RM.M58	1	G155D
RM.M59	1	R108C
RM.M60	1	T46I
RM.M61	1	E101V

Appendix V: Fatty acid production profiles for C12-specific random variants and their constituent point mutations.

This table provides the mole fractions of C8-C18 FFAs and total FFA in the cell lysate for RM.M39 and RM.M57 (the two C12-specific RM variants) as well as the individual amino acid substitutions (see Appendix IV).

	Fatty Acid Composition (mol %)										
Name	**8:0**	**10:0**	**12:0**	**12:1**	**14:0**	**14:1**	**16:0**	**16:1**	**18:0**	**18:1**	**Total FFA (μM)**
WT	5.9±0.3	1.09±0.06	20±1	7.5±0.5	29±2	7.1±0.4	14.2±0.8	11.6±0.8	0.22±0.01	2.1±0.1	1900±50
T35S	2.7±0.9	0.9±0.1	16±4	6±2	28±7	6±2	21±7	17±2	0.3±0.1	2.3±0.4	1600±200
E142D	5.5±0.7	2.5±0.2	33±3	12±1	26±2	6.1±0.6	5.7±0.8	8.1±0.8	0.14±0.04	1.2±0.1	3000±100
RM.M39	6±2	2.5±0.4	32±6	12±2	26±6	6±1	6.4±0.6	8±2	0.10±0.01	1.2±0.3	2800±200
A111T	4±2	1.0±0.1	15±3	6±1	25±5	6±1	23±3	17±3	0.35±0.06	2.2±0.4	1140±80
M141I	4±2	2.5±0.3	17±4	14±3	21±4	3.8±0.9	20±2	14±2	0.32±0.04	2.2±0.3	1200±80
RM.M57	5±1	2.9±0.7	17±4	14±4	21±4	3.8±1.0	20±2	13±1	0.33±0.05	2.1±0.3	1400±100

Appendix VI: TesA residues sorted by distance from acyl-ACP for design position selection.

Design positions were selected from each round using the eight closest residues that were not excluded from consideration. The method used to sort by distance varied between R1-R2 and R3-R4. The sorted list is provided in the table. Residues are labeled by their WT amino acid followed by the position of the residue. Residues that were excluded from design position consideration were annotated. The distance used to sort the residues is provided within parentheses. Thus, the eight highest ranked residues that were not annotated (i.e., excluded) formed the set of design positions employed within IPRO.

Rank	**R1[a]**	**R2[a]**	**R3-R4[b]**
1	Y145 (1.83 Å)	Y145 (1.83 Å)	P110[k] (1.83 Å)
2	L109[f,h] (2.70 Å)	L109[f] (2.70 Å)	Y145 (2.70 Å)
3	P110[e] (2.77 Å)	P110[e] (2.77 Å)	R108 (2.77 Å)
4	R108 (3.02 Å)	R108[i] (3.02 Å)	M141 (3.02 Å)
5	N73[d,g,h] (3.08 Å)	N73[d,g] (3.08 Å)	A111[l] (3.08 Å)
6	S10[c,d,g] (3.25 Å)	S10[c,d,g] (3.25 Å)	L109 (3.25 Å)
7	G44[d,g,h] (3.33 Å)	G44[d,g] (3.33 Å)	I156[k] (3.33 Å)
8	D9[g] (3.47 Å)	D9[g] (3.47 Å)	E142 (3.47 Å)
9	L11 (3.71 Å)	L11 (3.71 Å)	G155[k] (3.71 Å)
10	A111 (3.72 Å)	A111[j] (3.72 Å)	F139[k] (3.72 Å)
11	G72 (3.81 Å)	G72 (3.81 Å)	N112[k,l] (3.81 Å)
12	L76 (3.86 Å)	L76[j] (3.86 Å)	Y113[k,l] (3.86 Å)
13	H157[c,h] (3.99 Å)	H157[c] (3.99 Å)	V144[k] (3.99 Å)
14	I156[h] (4.21 Å)	I156 (4.21 Å)	N118[l] (4.21 Å)
15	F139 (4.37 Å)	F139 (4.37 Å)	I107 (4.37 Å)
16	S43[g] (4.37 Å)	S43[g] (4.37 Å)	M151[k,l] (4.37 Å)
17	D45[g] (4.82 Å)	D45[g] (4.82 Å)	L76[k] (4.82 Å)
18	I107 (4.83 Å)	I107[i] (4.83 Å)	G72[k] (4.83 Å)

19	G155 (5.01 Å)	G155 (5.01 Å)	L146 (5.01 Å)
20	I42[g,h] (5.14 Å)	I42[g] (5.14 Å)	H157[k] (5.14 Å)
21	E142[h] (5.53 Å)	E142 (5.53 Å)	S122 (5.53 Å)
22	M141 (5.56 Å)	M141 (5.56 Å)	D154[k] (5.56 Å)
23	E69 (5.62 Å)	E69 (5.62 Å)	P158[k] (5.62 Å)
24	Y15 (5.96 Å)	Y15 (5.96 Å)	L11[k] (5.96 Å)
25	N112 (5.98 Å)	N112[j] (5.98 Å)	E143 (5.98 Å)
26	G71 (6.15 Å)	G71 (6.15 Å)	F121[k] (6.15 Å)
27	M151 (6.22 Å)	M151 (6.22 Å)	G75[k] (6.22 Å)
28	N118 (6.23 Å)	N118[j] (6.23 Å)	F140[k] (6.23 Å)
29	S12 (6.43 Å)	S12 (6.43 Å)	G114[k,l] (6.43 Å)
30	F121 (6.66 Å)	F121 (6.66 Å)	Q106 (6.66 Å)
31	Y113 (6.67 Å)	Y113[j] (6.67 Å)	P148[l] (6.67 Å)
32	G75 (6.97 Å)	G75[j] (6.97 Å)	Q152[k,l] (6.97 Å)
33	G8 (6.99 Å)	G8 (6.99 Å)	W150[l] (6.99 Å)
34	P158[e] (7.02 Å)	P158[e] (7.02 Å)	K147[l] (7.02 Å)
35	D74 (7.06 Å)	D74 (7.06 Å)	Y117[k,l] (7.06 Å)
36	T46[h] (7.12 Å)	T46 (7.12 Å)	P138[k] (7.12 Å)
37	G14[h] (7.56 Å)	G14 (7.56 Å)	G71[k] (7.56 Å)
38	A13 (7.56 Å)	A13 (7.56 Å)	R115[l] (7.56 Å)
39	L70[h] (7.67 Å)	L70 (7.67 Å)	N73[k] (7.67 Å)
40	D154[c] (7.87 Å)	D154[c] (7.87 Å)	A162[k] (7.87 Å)
41	S122 (8.22 Å)	S122 (8.22 Å)	S10[k] (8.22 Å)
42	R77[h] (8.22 Å)	R77[j] (8.22 Å)	M105 (8.22 Å)
43	V144 (8.32 Å)	V144 (8.32 Å)	E119[l] (8.32 Å)
44	Q49 (8.38 Å)	Q49 (8.38 Å)	Y15[k] (8.38 Å)

45	Q106 (8.41 Å)	Q106 (8.41 Å)	Y125[k] (8.41 Å)
46	F140[h] (8.41 Å)	F140 (8.41 Å)	D9[k] (8.41 Å)
47	S41 (8.51 Å)	S41 (8.51 Å)	D153[l] (8.51 Å)
48	Y125 (8.55 Å)	Y125 (8.55 Å)	E69[k] (8.55 Å)
49	L146[h] (8.61 Å)	L146 (8.61 Å)	G78[k] (8.61 Å)
50	E143[h] (8.76 Å)	E143 (8.76 Å)	D74[k] (8.76 Å)
51	M105 (9.31 Å)	M105 (9.31 Å)	I166[k] (9.31 Å)
52	S47 (9.35 Å)	S47 (9.35 Å)	N159[k] (9.35 Å)
53	R16 (9.41 Å)	R16 (9.41 Å)	R77[k] (9.41 Å)
54	G114[h] (9.51 Å)	G114[j] (9.51 Å)	I124 (9.51 Å)
55	Q152 (9.75 Å)	Q152 (9.75 Å)	A120 (9.75 Å)
56	Y117 (9.76 Å)	Y117[j] (9.76 Å)	Q149[l] (9.76 Å)
57	R53 (9.87 Å)	R53 (9.87 Å)	A123 (9.87 Å)
58	W23 (10.13 Å)	W23 (10.13 Å)	L70[k] (10.13 Å)
59	G78 (10.15 Å)	G78[j] (10.15 Å)	L137 (10.15 Å)
60	E119[h] (10.41 Å)	E119[j] (10.41 Å)	S12[k] (10.41 Å)
61	L7[h] (10.46 Å)	L7 (10.46 Å)	P81[k] (10.46 Å)
62	K147[h] (10.68 Å)	K147 (10.68 Å)	R116[l] (10.68 Å)
63	R115[h] (10.78 Å)	R115[j] (10.78 Å)	G44[k] (10.78 Å)
64	G50 (10.89 Å)	G50 (10.89 Å)	W169 (10.89 Å)
65	T84 (11.10 Å)	T84 (11.10 Å)	G14[k] (11.10 Å)
66	M17 (11.11 Å)	M17 (11.11 Å)	T84 (11.11 Å)
67	A40[h] (11.16 Å)	A40 (11.16 Å)	F165 (11.16 Å)
68	Q163 (11.17 Å)	Q163 (11.17 Å)	F79 (11.17 Å)
69	W150 (11.31 Å)	W150 (11.31 Å)	G8[k] (11.31 Å)
70	I166 (11.32 Å)	I166 (11.32 Å)	S43[k] (11.32 Å)

71	D153[h] (11.33 Å)	D153 (11.33 Å)	P126 (11.33 Å)
72	N39 (11.41 Å)	N39 (11.41 Å)	W23[k] (11.41 Å)
73	N159 (11.42 Å)	N159 (11.42 Å)	L136[k] (11.42 Å)
74	F79 (11.44 Å)	F79[j] (11.44 Å)	S47 (11.44 Å)
75	A22 (11.51 Å)	A22 (11.51 Å)	Q163[k] (11.51 Å)
76	V68 (11.54 Å)	V68 (11.54 Å)	T46[k] (11.54 Å)
77	P148[e] (11.59 Å)	P148[e] (11.59 Å)	R16[l] (11.59 Å)
78	P138[e] (11.67 Å)	P138[e] (11.67 Å)	M170 (11.67 Å)
79	Q48 (11.69 Å)	Q48 (11.69 Å)	D161 (11.69 Å)
80	I6 (11.72 Å)	I6 (11.72 Å)	L104[k] (11.72 Å)
81	L88 (12.02 Å)	L88 (12.02 Å)	A13[k] (12.02 Å)
82	P24[e] (12.16 Å)	P24[e] (12.16 Å)	D45[k] (12.16 Å)
83	I124 (12.31 Å)	I124 (12.31 Å)	Q173 (12.31 Å)
84	R116[h] (12.66 Å)	R116[j] (12.66 Å)	S41[k] (12.66 Å)
85	A123[h] (12.68 Å)	A123 (12.68 Å)	A22 (12.68 Å)
86	A120 (12.69 Å)	A120[j] (12.69 Å)	V68 (12.69 Å)
87	L137 (13.07 Å)	L137 (13.07 Å)	L88[k] (13.07 Å)
88	L67 (13.19 Å)	L67 (13.19 Å)	R160 (13.19 Å)
89	L51 (13.27 Å)	L51 (13.27 Å)	M17[l] (13.27 Å)
90	A162 (13.28 Å)	A162 (13.28 Å)	E85 (13.28 Å)
91	L104 (13.37 Å)	L104 (13.37 Å)	L67 (13.37 Å)
92	S18[h] (13.41 Å)	S18 (13.41 Å)	A129[k] (13.41 Å)
93	W169 (13.43 Å)	W169 (13.43 Å)	K127 (13.43 Å)
94	P81[e] (13.54 Å)	P81[e] (13.54 Å)	Q80 (13.54 Å)
95	A19 (13.56 Å)	A19 (13.56 Å)	A167[k] (13.56 Å)
96	Q149[h] (13.56 Å)	Q149 (13.56 Å)	L128[k] (13.56 Å)

97	M170 (13.69 Å)	M170 (13.69 Å)	P24[k] (13.69 Å)
98	I91 (13.76 Å)	I91 (13.76 Å)	I6 (13.76 Å)
99	R160[h] (13.90 Å)	R160 (13.90 Å)	I42 (13.90 Å)
100	L136 (14.23 Å)	L136 (14.23 Å)	D168 (14.23 Å)
101	Q80[h] (14.28 Å)	Q80[j] (14.28 Å)	P164 (14.28 Å)
102	P126[e] (14.42 Å)	P126[e] (14.42 Å)	L7[k] (14.42 Å)
103	A21 (14.45 Å)	A21 (14.45 Å)	L103[k] (14.45 Å)
104	T87 (14.46 Å)	T87 (14.46 Å)	Q83 (14.46 Å)
105	E85 (14.64 Å)	E85 (14.64 Å)	T87 (14.64 Å)
106	A52 (15.04 Å)	A52 (15.04 Å)	K172 (15.04 Å)
107	L103 (15.26 Å)	L103 (15.26 Å)	G50[k] (15.26 Å)
108	L57 (15.37 Å)	L57 (15.37 Å)	Q49 (15.37 Å)
109	Q83 (15.58 Å)	Q83 (15.58 Å)	Q82 (15.58 Å)
110	L128 (15.59 Å)	L128 (15.59 Å)	A21 (15.59 Å)
111	A25 (15.64 Å)	A25 (15.64 Å)	P135 (15.64 Å)
112	V38[h] (15.71 Å)	V38 (15.71 Å)	L174[k] (15.71 Å)
113	S20[h] (15.81 Å)	S20 (15.81 Å)	Q48 (15.81 Å)
114	L5 (16.02 Å)	L5 (16.02 Å)	L27[k] (16.02 Å)
115	A167[h] (16.03 Å)	A167 (16.03 Å)	N39[k] (16.03 Å)
116	Q82[h] (16.04 Å)	Q82 (16.04 Å)	L26[k] (16.04 Å)
117	L54 (16.11 Å)	L54 (16.11 Å)	I91[k] (16.11 Å)
118	D161 (16.25 Å)	D161 (16.25 Å)	A171 (16.25 Å)
119	V66 (16.52 Å)	V66 (16.52 Å)	V134[l] (16.52 Å)
120	F165 (16.57 Å)	F165 (16.57 Å)	K130 (16.57 Å)
121	L92 (16.80 Å)	L92 (16.80 Å)	A19[l] (16.80 Å)
122	P164[e,h] (16.88 Å)	P164[e] (16.88 Å)	Q86 (16.88 Å)

123	Q173 (17.03 Å)	Q173 (17.03 Å)	L51[k] (17.03 Å)
124	V37 (17.04 Å)	V37 (17.04 Å)	A25[k] (17.04 Å)
125	Q86[h] (17.06 Å)	Q86 (17.06 Å)	L92 (17.06 Å)
126	R89 (17.23 Å)	R89 (17.23 Å)	A40[k] (17.23 Å)
127	L26 (17.38 Å)	L26 (17.38 Å)	P102 (17.38 Å)
128	P135[e] (17.51 Å)	P135[e] (17.51 Å)	R89 (17.51 Å)
129	K127 (17.53 Å)	K127 (17.53 Å)	V66[k] (17.53 Å)
130	L27 (17.66 Å)	L27 (17.66 Å)	L177 (17.66 Å)
131	L174 (17.80 Å)	L174 (17.80 Å)	S18[l] (17.80 Å)
132	L4 (18.13 Å)	L4 (18.13 Å)	E131 (18.13 Å)
133	A129 (18.20 Å)	A129 (18.20 Å)	V37[k] (18.20 Å)
134	D168 (18.62 Å)	D168 (18.62 Å)	F132[l] (18.62 Å)
135	P55[e] (18.88 Å)	P55[e] (18.88 Å)	L4 (18.88 Å)
136	A171[h] (18.95 Å)	A171 (18.95 Å)	R53[k] (18.95 Å)
137	V95 (19.03 Å)	V95 (19.03 Å)	L5[k] (19.03 Å)
138	P102[e] (19.14 Å)	P102[e] (19.14 Å)	Q175 (19.14 Å)
139	Q90 (19.20 Å)	Q90 (19.20 Å)	L54[k] (19.20 Å)
140	V134 (19.34 Å)	V134 (19.34 Å)	P176 (19.34 Å)
141	N28 (19.46 Å)	N28 (19.46 Å)	Q90 (19.46 Å)
142	H61 (19.60 Å)	H61 (19.60 Å)	S20[l] (19.60 Å)
143	W65 (19.67 Å)	W65 (19.67 Å)	W65 (19.67 Å)
144	L177 (20.06 Å)	L177 (20.06 Å)	A52 (20.06 Å)
145	Q60 (20.07 Å)	Q60 (20.07 Å)	V95 (20.07 Å)
146	D94 (20.41 Å)	D94 (20.41 Å)	N28 (20.41 Å)
147	A56 (20.42 Å)	A56 (20.42 Å)	K30[l] (20.42 Å)
148	L58 (20.50 Å)	L58 (20.50 Å)	V38 (20.50 Å)

149	K172 (20.68 Å)	K172 (20.68 Å)	V178 (20.68 Å)
150	K130 (20.74 Å)	K130 (20.74 Å)	D133[l] (20.74 Å)
151	E101 (20.91 Å)	E101 (20.91 Å)	D29[l] (20.91 Å)
152	W31 (20.95 Å)	W31[j] (20.95 Å)	E101 (20.95 Å)
153	F132 (21.00 Å)	F132 (21.00 Å)	L57 (21.00 Å)
154	S36[h] (21.31 Å)	S36 (21.31 Å)	W31[l] (21.31 Å)
155	Q93 (21.32 Å)	Q93 (21.32 Å)	D94 (21.32 Å)
156	P63[e] (21.42 Å)	P63[e] (21.42 Å)	Q93 (21.42 Å)
157	D29 (21.54 Å)	D29 (21.54 Å)	P63[k] (21.54 Å)
158	T3 (21.67 Å)	T3 (21.67 Å)	A100[k] (21.67 Å)
159	E131 (21.97 Å)	E131 (21.97 Å)	T35 (21.97 Å)
160	K30 (22.18 Å)	K30 (22.18 Å)	K96 (22.18 Å)
161	Q175 (22.37 Å)	Q175 (22.37 Å)	H180 (22.37 Å)
162	V178 (22.79 Å)	V178 (22.79 Å)	T3 (22.79 Å)
163	A100 (23.19 Å)	A100 (23.19 Å)	L58[k] (23.19 Å)
164	K59 (23.20 Å)	K59 (23.20 Å)	H61[k,l] (23.20 Å)
165	D133[h] (23.23 Å)	D133 (23.23 Å)	S36 (23.23 Å)
166	K96 (23.32 Å)	K96 (23.32 Å)	R64 (23.32 Å)
167	R64 (23.54 Å)	R64 (23.54 Å)	P55[k] (23.54 Å)
168	P176[e] (23.57 Å)	P176[e] (23.57 Å)	A98[l] (23.57 Å)
169	T35 (23.61 Å)	T35[j] (23.61 Å)	A56 (23.61 Å)
170	Q62[h] (24.10 Å)	Q62 (24.10 Å)	K34[l] (24.10 Å)
171	A98 (24.21 Å)	A98 (24.21 Å)	Q32[l] (24.21 Å)
172	D2 (24.93 Å)	D2 (24.93 Å)	D2 (24.93 Å)
173	H180 (25.52 Å)	H180 (25.52 Å)	A97 (25.52 Å)
174	A97 (25.70 Å)	A97 (25.70 Å)	Q60[l] (25.70 Å)

175	Q32 (25.71 Å)	Q32[j] (25.71 Å)	K59 (25.71 Å)
176	N99[h] (26.55 Å)	N99 (26.55 Å)	N99 (26.55 Å)
177	K34 (27.24 Å)	K34[j] (27.24 Å)	Q62 (27.24 Å)
178	A1 (27.76 Å)	A1 (27.76 Å)	A1 (27.76 Å)
179	S33[h] (28.10 Å)	S33[j] (28.10 Å)	S33[l] (28.10 Å)

[a] Distance was calculated by finding the closest contact between the residue's heavy atoms and the acyl moiety carbon atoms of 'TesA:tetradecanoyl-ACP

[b] Distance was calculated between the residue's Cβ atom (Cα for glycine, Cγ for H180) and the ω-1 atom of dodecanoyl-ACP

[c] Residue excluded because it is part of the 'TesA catalytic triad [58]

[d] Residue excluded because they form the 'TesA oxyanion hole [58]

[e] Residue excluded because legacy version of IPRO could not handle mutations from proline [24]

[f] Residue excluded because it was considered important for 'TesA function based on mutagenesis studies [20]

[g] Residue excluded because it was within 4.5 Å of the thioester sulfur atom

[h] Residue excluded because the Cα atom was over 0.75 Å closer to the ω-1 atom than the Cβ atom (glycines not considered)

[i] Residue excluded based on R1 results that abolished 'TesA activity

[j] Residue excluded because it is belongs to one of the three 'TesA flexible loops [20]

[k] Residue excluded because it has ≥40% sequence conservation in lysophospholipase L1-like subgroup alignment [65]

[l] Residue excluded because sequence alignment included at least one gap

Appendix VII: Crystallography data collection statistics.
The data collection statistics for four structures (WT and R3.M4 under three different conditions) are provided.

	WT	**R3.M4 (pH 5.0)**	**R3.M4 (pH 7.5)**	**R3.M4 (pH 7.5 "soaked")**
Resolution limits (Å)	50-1.65 (1.75-1.65)[b]	50-1.2 (1.3-1.2)[b]	50-1.15 (1.25-1.15)[b]	50-0.97 (1.07-0.97)[b]
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell				
a (Å)	40.9	40.7	40.6	40.5
b (Å)	82.1	55.2	55.3	55.0
c (Å)	53.9	42.3	42.2	42.0
β (°)	90.4	105.2	104.9	104.9
no. of independent reflections	42019 (6561)	56148 (11626)	62723 (12626)	102836 (24792)
completeness (%)	98.0 (94.9)	99.4 (97.3)	98.2 (91.8)	97.8 (92.8)
redundancy	3.7 (2.6)	3.7 (2.2)	2.3 (1.3)	4.1 (2.3)
avg I/avg σ(I)	10.6 (1.9)	13.2 (3.0)	11.0 (2.7)	11.8 (2.4)
R_{sym} (%)[a]	6.8 (37.8)	5.6 (30.2)	4.5 (21.0)	6.0 (27.8)

[a] $R_{sym} = \left(\sum \left| \sum I - \bar{I} \right| / \sum I \right) \times 100$

[b] Statistics for the highest resolution bin

Appendix VIII: Crystallography refinement statistics.

The refinement statistics for the crystallography experiments summarized in Appendix VI are provided.

	WT	**R3.M4 (pH 5.0)**	**R3.M4 (pH 7.5)**	**R3.M4 (pH 7.5 "soaked")**
PDB code	5TIC	5TID	5TIE	5TIF
resolution limits (Å)	50-1.65	50-1.2	50-1.15	50-0.97
R-factor[a] (overall)%/ no. reflections	18.8/42019	15.9/56148	13.1/62723	12.3/102836
R-factor[a] (working)%/ no. reflections	18.6/39926	15.8/53269	12.9/59641	12.2/97699
R-factor[a] (free)%/ no. reflections	22.2/2093	17.7/2879	16.3/3082	13.7/5137
no. of protein atoms	2839	1448	1471	1542
no. of heteroatoms	347	344	391	449
average B values				
protein ($Å^2$)	20.1	10.4	10.7	13.7
ligand ($Å^2$)	-	21.5	20.6	27.6
solvent ($Å^2$)	28.1	22.2	24.1	32.7
weighted RMSDs from ideality				
bond lengths (Å)	0.012	0.012	0.017	0.015
bond angles (°)	1.77	1.74	1.91	2.01
planar groups (Å)	0.008	0.009	0.010	0.010
Ramachandran regions (%)[b]				
most favored	97.7	97.2	97.2	97.8
additionally allowed	2.3	2.8	2.8	2.2
generously allowed	0.0	0.0	0.0	0.0

[a] $R\text{-factor} = \left(\sum |F_o - F_c| / \sum |F_o|\right) \times 100$

[b] Distribution of Ramachandran angles according to PROCHECK [87]

Appendix IX: Fatty acid profile of 90 putative mutants.

Each of the colonies selected was grown to isolate its plasmid. Isolated plasmids were transformed into RL08ara (Δ*fadD*) strain and grown in LB 0.4% glycerol and 20 µM IPTG (mutants M1-M10) and 1 mM IPTG (mutants M3, M4, and M11-M90).

mutant	# residue changes	mutations	C8:0 (mg/L)	C10:0 (mg/L)	C10:1 (mg/L)	C12:0 (mg/L)	C12:1 (mg/L)	C14:0 (mg/L)	C14:1 (mg/L)	C16:0 (mg/L)	C16:1 (mg/L)	C18:0 (mg/L)	C18:1 (mg/L)	Total FFA (mg/L)	induction level tested (µM)	Fold Increase over CpFatB1.2
CpFatB1.2	0	n/a	12.4 20.7 ± 1.75	5.6 1.7 ± 0.48	0.9 1.5 ± 0.07	0.0 0.0	1.5 1.7 ± 0.12	0.6 0.7 ± 0.1	9.4 8.4 ± 1.1	10.1 13.2 ± 1.2	52.6 56.0 ± 2.9	0.0 0.0	0.0 1.3 ± 0.04	93.2 107.1	20 1000	1.0 1.0
CpFatB1.2-M1	0	n/a	20.7	4.7	1.6	0.0	1.8	0.8	9.8	12.8	58.2	0.0	0.0	110.2	20	1.7
CpFatB1.2-M2	3	M29T, T117S, Q163L	15.0	2.8	1.3	0.0	1.6	0.8	10.5	17.5	69.7	0.0	1.5	120.7	20	1.2
CpFatB1.2-M3	2	A59S, K296R	49.5 82.9 ± 5.86	2.7 1.8 ± 0.34	4.8 7.0 ± 0.26	0.6 1.0 ± 0.04	3.6 4.9 ± 0.09	1.1 1.4 ± 0.04	9.4 8.6 ± 0.4	11.7 10.1 ± 0.3	51.8 51.1 ± 0.8	0.0 0	1.4 1.1 ± 0.9	136.7 169.9	20 1000	4.0 4.0
CpFatB1.2-M4	2	ΔA54 (new start codon at M19), N28S, I65M	66.8 155.3 ± 16.2	2.9 2.1 ± 0.18	4.8 10.8 ± 0.96	0.6 1.3 ± 0.12	3.8 6.1 ± 0.26	0.9 1.6 ± 0.1	8.3 8.8 ± 0.4	12.6 9.1 ± 0.8	46.9 47.8 ± 1.6	0.0 0	1.4 1.2 ± 1.0	149.0 243.9	20 1000	5.4 7.5
CpFatB1.2-M4-287	N/A	287 truncation of M4	468.0	0.5	4.1	10.1	2.7	10.0	2.9	40.4	8.5	0.0	1.9	575.4	1000	22.6
CpFatB1.2-M4-288	N/A	288 truncation of M4	17.9	0.5	1.9	1.3	0.0	11.5	0.8	64.3	10.4	0.0	1.5	110.4	1000	0.9
CpFatB1.2-M4-289	N/A	289 truncation of M4	11.7	0.5	0.5	0.7	0.0	3.9	0.0	47.9	3.9	0.0	1.3	69.9	1000	0.6
CpFatB1.2-M4-290	N/A	290 truncation of M4	11.7	0.5	2.0	1.1	0.0	16.4	1.0	69.6	12.5	1.3	1.4	117.0	1000	0.6
CpFatB1.2-M4-291	N/A	291 truncation of M4	11.9	0.5	3.1	1.3	0.0	14.2	0.9	68.5	10.3	0.0	1.5	112.2	1000	0.6

CpFatB1.2-M5	2	W17R, T204S	24.9	4.3	1.9	0.0	2.0	0.7	8.4	12.1	52.8	0.0	0.0	107.0	20	2.01
CpFatB1.2-M6	2	L251M, L265I,	17.4	4.2	1.5	0.0	1.6	0.7	8.5	12.8	58.4	0.0	0.0	105.0	20	1.40
CpFatB1.2-M7	0	n/a	20.5	5.9	1.3	0.0	1.8	0.6	7.7	12.6	56.4	0.0	1.3	108.0	20	1.65
CpFatB1.2-M8	2	K15E, S207T	16.4	5.5	1.1	0.0	1.6	0.7	8.6	10.8	54.3	0.0	0.0	99.0	20	1.32
CpFatB1.2-M9	1	R261S	28.6	4.4	2.4	0.0	2.5	0.8	8.4	11.4	53.3	0.0	0.0	111.8	20	2.31
CpFatB1.2-M10	0	n/a	12.4	5.6	0.9	0.0	1.5	0.6	9.4	10.1	52.6	0.0	0.0	93.2	20	1.27
CpFatB1.2-M11	ND	ND	26.4	0.0	1.0	0.0	1.7	0.0	7.6	12.1	69.0	0.0	14.5	132.2	1000	1.3
CpFatB1.2-M12	1	M136V	61.0	1.3	2.8	0.0	3.8	0.0	7.9	11.8	60.2	0.0	21.4	170.2	1000	2.9
CpFatB1.2-M13	ND	ND	13.9	0.0	0.0	0.0	0.8	0.0	3.3	3.7	47.1	0.0	11.9	80.7	1000	0.7
CpFatB1.2-M14	ND	ND	16.5	0.0	0.9	0.0	1.4	0.0	5.3	8.9	68.5	0.0	15.3	116.8	1000	0.8
CpFatB1.2-M15	ND	ND	15.7	0.0	0.7	0.0	1.1	0.0	3.1	3.7	45.5	0.0	12.2	81.9	1000	0.8
CpFatB1.2-M16	1	K296R	43.2	0.8	2.3	0.0	3.0	0.0	6.9	10.7	61.8	0.0	19.2	147.8	1000	2.1
CpFatB1.2-M17	ND	ND	26.9	0.6	1.5	0.0	1.8	0.0	9.4	14.2	68.0	0.0	18.0	140.3	1000	1.3
CpFatB1.2-M18	ND	ND	27.2	1.1	1.5	0.0	2.2	0.0	10.0	17.4	86.7	0.0	24.2	170.2	1000	1.3
CpFatB1.2-M19	ND	ND	18.5	1.1	0.9	0.0	1.5	0.0	6.9	12.2	65.5	0.0	16.6	123.1	1000	0.9
CpFatB1.2-M20	2	W17STOP D293V	61.5	0.7	3.4	0.0	3.6	0.0	7.0	8.7	57.2	0.0	18.3	160.3	1000	3.0
CpFatB1.2-M21	ND	ND	17.4	0.0	0.9	0.0	1.9	0.0	10.1	13.2	63.5	1.1	14.5	122.7	1000	0.8
CpFatB1.2-M22	ND	ND	38.9	0.0	2.0	0.0	3.2	0.0	9.1	13.3	60.0	1.0	19.1	146.6	1000	1.9
CpFatB1.2-M23	ND	ND	21.5	0.0	1.2	0.0	2.3	0.0	9.2	12.7	60.4	0.9	15.5	123.7	1000	1.0
CpFatB1.2-M24	ND	ND	17.4	0.0	1.0	0.0	2.0	0.0	10.6	13.8	69.2	1.5	13.6	129.1	1000	0.8

CpFatB1.2-M25	ND	ND	17.5	0.0	1.0	0.0	2.0	0.0	11.9	14.3	63.3	1.0	14.9	125.8	1000	0.8
CpFatB1.2-M26	ND	ND	20.3	0.0	1.1	0.0	1.9	0.0	7.6	12.6	68.1	1.2	14.4	127.3	1000	1.0
CpFatB1.2-M27	ND	ND	15.8	0.0	0.9	0.0	2.2	0.0	11.7	13.4	64.4	1.2	13.2	122.9	1000	0.8
CpFatB1.2-M28	ND	ND	16.9	0.0	0.9	0.0	2.0	0.0	10.8	14.1	68.4	1.4	13.5	128.1	1000	0.8
CpFatB1.2-M29	ND	ND	21.9	0.0	1.2	0.0	2.0	0.0	5.7	9.4	65.5	1.4	13.1	120.1	1000	1.1
CpFatB1.2-M30	5	M19T, G35D, T117S, T121A, M138I	28.4	0.0	1.8	0.0	4.0	0.0	9.7	12.7	66.2	1.1	18.8	142.7	1000	1.4
CpFatB1.2-M31	ND	ND	15.2	0.0	1.0	0.0	2.0	0.0	10.1	13.5	64.0	3.5	11.7	121.0	1000	0.7
CpFatB1.2-M32	1	R22H	31.4	0.0	1.8	0.0	2.7	0.0	9.2	13.8	64.3	3.1	19.1	145.3	1000	1.5
CpFatB1.2-M33	ND	ND	12.8	0.0	0.8	0.0	1.6	0.0	6.1	10.8	63.5	3.5	8.7	107.9	1000	0.6
CpFatB1.2-M34	ND	ND	8.0	0.0	0.5	0.0	0.9	0.0	3.1	1.1	36.9	2.5	3.1	56.2	1000	0.4
CpFatB1.2-M35	ND	ND	14.2	0.0	0.9	0.0	1.7	0.0	7.9	12.8	65.6	3.8	10.6	117.6	1000	0.7
CpFatB1.2-M36	ND	ND	11.8	0.0	0.8	0.0	2.2	0.0	8.7	11.5	64.0	3.4	10.2	112.7	1000	0.6
CpFatB1.2-M37	ND	ND	11.4	0.0	0.7	0.0	1.3	0.0	5.4	6.9	63.9	3.6	6.8	100.0	1000	0.6
CpFatB1.2-M38	ND	ND	12.1	0.0	0.7	0.0	1.2	0.0	4.6	6.1	61.1	3.3	9.2	98.5	1000	0.6
CpFatB1.2-M39	ND	ND	17.3	0.0	1.1	0.0	1.9	0.0	6.5	12.3	76.2	4.6	11.2	131.0	1000	0.8
CpFatB1.2-M40	1	D293V	43.1	0.0	2.9	0.0	3.7	0.0	7.4	10.0	56.0	2.8	17.1	143.1	1000	2.1
CpFatB1.2-M41	ND	ND	8.6	0.0	0.0	0.0	1.0	0.0	3.3	2.4	39.8	0.7	6.8	62.6	1000	0.4
CpFatB1.2-M42	ND	ND	8.3	0.0	0.5	0.0	1.0	0.0	3.6	3.8	48.0	0.8	9.6	75.7	1000	0.4
CpFatB1.2-M43	ND	ND	13.8	0.0	0.8	0.0	1.6	0.0	5.5	8.9	71.5	1.2	15.5	118.8	1000	0.7

CpFatB1.2-M44	ND	ND	16.0	0.0	0.9	0.0	2.0	0.0	6.6	12.8	62.2	1.2	13.3	115.0	1000	0.8
CpFatB1.2-M45	ND	ND	8.4	0.0	0.5	0.0	1.0	0.0	3.2	3.5	42.6	0.7	9.0	69.0	1000	0.4
CpFatB1.2-M46	ND	ND	14.2	0.0	0.8	0.0	1.6	0.0	5.1	9.2	63.5	1.2	11.2	106.7	1000	0.7
CpFatB1.2-M47	3	N146K, D293V, N309D	69.2	0.0	4.3	0.0	4.6	0.0	6.1	8.9	62.7	1.2	21.0	178.0	1000	3.3
CpFatB1.2-M48	ND	ND	24.7	0.0	1.3	0.0	2.2	0.0	7.4	11.1	58.8	1.0	12.7	119.3	1000	1.2
CpFatB1.2-M49	ND	ND	8.3	0.0	0.5	0.0	1.0	0.0	3.4	3.6	45.0	0.8	9.9	72.5	1000	0.4
CpFatB1.2-M50	ND	ND	18.1	0.0	1.0	0.0	1.3	0.0	4.7	7.6	62.7	1.2	9.1	105.7	1000	0.9
CpFatB1.2-M51	ND	ND	13.2	0.0	0.0	0.0	1.5	0.0	6.1	0.0	61.9	0.0	11.8	94.5	1000	0.6
CpFatB1.2-M52	ND	ND	7.9	0.0	0.0	0.0	0.9	0.0	3.0	0.0	39.5	0.0	5.8	57.2	1000	0.4
CpFatB1.2-M53	ND	ND	14.7	0.0	0.0	0.0	1.5	0.0	4.7	0.0	60.0	0.0	9.3	90.2	1000	0.7
CpFatB1.2-M54	ND	ND	9.9	0.0	0.0	0.0	1.2	0.0	4.6	0.0	60.5	0.0	12.6	88.8	1000	0.5
CpFatB1.2-M55	ND	ND	23.5	0.0	0.0	0.0	2.0	0.0	5.9	0.0	55.3	0.0	13.1	99.9	1000	1.1
CpFatB1.2-M56	ND	ND	16.0	0.0	0.0	0.0	2.0	0.0	7.7	0.0	56.9	0.0	12.6	95.2	1000	0.8
CpFatB1.2-M57	ND	ND	19.8	0.0	0.0	0.0	2.0	0.0	7.1	0.0	60.5	0.0	14.4	103.8	1000	1.0
CpFatB1.2-M58	ND	ND	20.2	0.0	0.0	0.0	2.3	0.0	9.3	0.0	57.5	0.0	15.1	104.3	1000	1.0
CpFatB1.2-M59	ND	ND	18.6	0.0	0.0	0.0	1.7	0.0	6.0	0.0	64.6	0.0	13.4	104.4	1000	0.9
CpFatB1.2-M60	ND	ND	19.1	0.0	0.0	0.0	1.8	0.0	6.4	0.0	55.8	0.0	12.7	95.7	1000	0.9
CpFatB1.2-M61	ND	ND	13.6	0.0	0.0	0.0	1.0	0.0	3.2	1.4	39.6	0.7	4.3	63.8	1000	0.7
CpFatB1.2-M62	ND	ND	12.6	0.0	0.0	0.0	1.0	0.0	3.2	1.7	40.4	0.8	4.9	64.6	1000	0.6
CpFatB1.2-M63	ND	ND	19.7	0.0	0.0	0.0	1.6	0.0	6.7	12.7	66.2	1.3	10.4	118.6	1000	1.0

CpFatB1.2-M64	ND	ND	22.1	0.0	0.0	0.0	2.1	0.0	12.3	16.5	69.2	1.5	13.9	137.5	1000	1.1
CpFatB1.2-M65	ND	ND	20.5	0.0	0.0	0.0	1.7	0.0	6.6	10.3	73.9	1.3	10.0	124.3	1000	1.0
CpFatB1.2-M66	1	D293V	54.3	0.0	1.3	0.0	3.7	0.0	7.8	10.9	58.6	1.1	15.2	152.9	1000	2.6
CpFatB1.2-M67	ND	ND	12.7	0.0	0.0	0.0	0.9	0.0	3.2	1.0	39.4	0.8	3.6	61.6	1000	0.6
CpFatB1.2-M68	1	T245N	28.4	0.0	0.0	0.0	2.0	0.0	8.1	15.4	64.8	1.2	14.3	134.3	1000	1.4
CpFatB1.2-M69	ND	ND	21.2	0.0	0.0	0.0	1.7	0.0	7.9	14.8	65.2	1.1	12.0	123.7	1000	1.0
CpFatB1.2-M70	ND	ND	21.8	0.0	0.0	0.0	1.6	0.0	7.3	14.2	67.9	1.2	12.3	126.4	1000	1.1
CpFatB1.2-M71	1	M136I	46.7	0.0	1.0	0.0	2.8	0.0	5.2	3.9	85.2	1.8	18.0	164.6	1000	2.3
CpFatB1.2-M72	ND	ND	22.9	0.0	0.0	0.0	1.8	0.0	5.2	8.5	61.5	1.3	11.1	112.3	1000	1.1
CpFatB1.2-M73	3	V268I, R279H, D293V	61.2	0.0	1.4	0.0	4.2	0.0	7.8	8.5	56.0	0.9	16.4	156.5	1000	3.0
CpFatB1.2-M74	ND	ND	21.6	0.0	0.0	0.0	2.1	0.0	9.1	12.8	60.1	0.9	14.1	120.6	1000	1.0
CpFatB1.2-M75	ND	ND	18.0	0.0	0.0	0.0	1.4	0.0	4.3	7.3	50.2	1.0	9.2	91.2	1000	0.9
CpFatB1.2-M76	2	K23E, L86Q	46.6	0.0	0.6	0.0	2.5	0.0	6.6	11.9	58.6	0.9	16.9	144.6	1000	2.2
CpFatB1.2-M77	ND	ND	21.4	0.0	0.0	0.0	1.9	0.0	8.2	14.4	62.9	1.1	15.5	125.4	1000	1.0
CpFatB1.2-M78	ND	ND	28.2	0.0	0.0	0.0	2.4	0.0	8.7	12.6	66.9	1.2	15.6	135.7	1000	1.4
CpFatB1.2-M79	3	V9M, K15E, E236A	42.0	0.0	0.6	0.0	3.3	0.0	8.5	12.6	59.1	1.0	17.9	144.9	1000	2.0
CpFatB1.2-M80	ND	ND	19.9	0.0	0.0	0.0	1.3	0.0	6.3	10.4	59.8	1.1	12.4	111.2	1000	1.0
CpFatB1.2-M81	ND	ND	8.5	0.0	0.4	0.0	1.2	0.0	5.0	3.2	72.0	1.6	8.3	100.1	1000	0.4
CpFatB1.2-M82	ND	ND	11.3	0.0	0.6	0.0	1.8	0.0	4.9	6.4	71.8	1.6	15.2	113.6	1000	0.5
CpFatB1.2-M83	ND	ND	18.8	0.0	1.2	0.0	2.6	0.0	5.6	4.4	69.2	1.5	14.6	117.8	1000	0.9

CpFatB1.2-M84	ND	ND	9.2	0.0	0.4	0.0	1.2	0.0	4.6	2.4	66.2	1.5	7.4	93.0	1000	0.4
CpFatB1.2-M85	ND	ND	9.4	0.0	0.5	0.0	1.4	0.0	4.6	5.0	69.1	1.6	10.3	101.8	1000	0.5
CpFatB1.2-M86	ND	ND	16.2	0.0	1.0	0.0	2.1	0.0	5.3	6.6	71.7	1.4	16.1	120.3	1000	0.8
CpFatB1.2-M87	ND	ND	27.2	0.0	1.6	0.0	1.9	0.0	4.9	2.7	65.6	1.6	9.4	114.9	1000	1.3
CpFatB1.2-M88	ND	ND	12.0	0.0	0.7	0.0	1.8	0.0	5.1	6.7	74.5	1.6	16.4	118.8	1000	0.6
CpFatB1.2-M89	ND	ND	8.6	0.0	0.4	0.0	1.2	0.0	4.3	2.4	63.9	1.5	7.7	90.0	1000	0.4
CpFatB1.2-M90	ND	ND	11.1	0.0	0.6	0.0	1.5	0.0	5.1	5.2	72.8	1.5	12.7	110.6	1000	0.5

Appendix X: Nucleotide sequence of CpFatB1 and key derivatives

Protein name	Nucleotide Sequence	reference
CpFatB1	ATGGTAGCTGCTGCGGCGAGTAGCGCTTGCTTTCCAGTGCCGTCTCCGGGCGCAAGCCCGAAACCGGGGAAACTGGGTAACTGG AGCTCTTCGCTGAGTCCGAGCCTGAAACCGAAATCTATTCCGAACGGCGGCTTTCAGGTTAAAGCAAACGCGTCCGCACATCCGAA GGCAAACGGCTCCGCGGTTACTCTGAAAAGCGGTAGCCTGAACACGCAGGAAGACACACTGTCCAGCAGCCCTCCGCCGCGTGC TTTCTTCAACCAGCTGCCGGACTGGAGCATGCTGCTCACCGCGATTACCACCGTGTTCGTTGCGCCGGAAAAACGTTGGACTATGT TCGACCGCAAATCTAAACGTCCGAACATGCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAA TCCTTCTCCATTCGTAGCTACGAAATCTGTGCGGACCGCACGGCGAGCATCGAGACCGTGATGAACCACGTTCAGGAAACCAGCC TGAACCAGTGTAAGTCCATCGGCCTGCTGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTA ACCCGTATGAAAATTATGGTAAACCGTTATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAAT TGGCATGGGTCGCGACTGGCTGATTAGCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACC AGAAAACGCGTCGCTTCTCTAAACTGCCCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGA AGACAACGACGGTAAACTGCAGAAATTCGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTG GACGTAAACCAGCACGTTTCCAATGTCAAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTT TGCAGCCTGACTCTCGAATACCGTCGCGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTG GTGATCGTTTTCAGTATCGTCATCTGCTGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAAAAAC GCGGGCACTAATGGTGCTATCTCCACTGGGAAAACAtaa	136
CpFatB1.2	ATGCTGCTCACCGCGATTACCACCGTGTTCGTTGCGCCGGAAAAACGTTGGACTATGTTCGACCGCAAATCTAAACGTCCGAACAT GCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAATCCTTCTCCATTCGTAGCTACGAAATCT GTGCGGACCGCACGGCGAGCATCGAGACCGTGATGAACCACGTTCAGGAAACCAGCCTGAACCAGTGTAAGTCCATCGGCCTGC TGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTAACCCGTATGAAAATTATGGTAAACCGT TATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAATTGGCATGGGTCGCGACTGGCTGATTA GCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACCAGAAAACGCGTCGCTTCTCTAAACTGC CCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGAAGACAACGACGGTAAACTGCAGAAATT CGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTGGACGTAAACCAGCACGTTTCCAATGTC AAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTTTGCAGCCTGACTCTCGAATACCGTCG CGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTGGTGATCGTTTTCAGTATCGTCATCTGC TGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAAAAACGCGGGCACTAATGGTGCTATCTCCAC TGGGAAAACAtaa	31
CpFatB1.3	ATGAAATCTAAACGTCCGAACATGCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAATCCTTC TCCATTCGTAGCTACGAAATCTGTGCGGACCGCACGGCGAGCATCGAGACCGTGATGAACCACGTTCAGGAAACCAGCCTGAACC AGTGTAAGTCCATCGGCCTGCTGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTAACCCG TATGAAAATTATGGTAAACCGTTATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAATTGGCA TGGGTCGCGACTGGCTGATTAGCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACCAGAAA ACGCGTCGCTTCTCTAAACTGCCCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGAAGACA ACGACGGTAAACTGCAGAAATTCGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTGGACGT AAACCAGCACGTTTCCAATGTCAAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTTTGCAG CCTGACTCTCGAATACCGTCGCGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTGGTGATC GTTTTCAGTATCGTCATCTGCTGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAAAAACGCGGG CACTAATGGTGCTATCTCCACTGGGAAAACAtaa	This Work
CpFatB1.4	ATGGGTCTGGTTTTCCGCCAATCCTTCTCCATTCGTAGCTACGAAATCTGTGCGGACCGCACGGCGAGCATCGAGACCGTGATGAA CCACGTTCAGGAAACCAGCCTGAACCAGTGTAAGTCCATCGGCCTGCTGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAA CGCGACCTGATTTGGGTTGTAACCCGTATGAAAATTATGGTAAACCGTTATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTG	This Work

	GCTGTCCCAGTCCGGTAAAATTGGCATGGGTCGCGACTGGCTGATTAGCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCT CAGTTTACGCCATGATGAACCAGAAAACGCGTCGCTTCTCTAAACTGCCCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTA GACTCCCCGCCGGCCATTGAAGACAACGACGGTAAACTGCAGAAATTCGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGA CCCCGGGCTGGTACGACCTGGACGTAAACCAGCACGTTTCCAATGTCAAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAA GTTCTGGAAACCCAAGAACTTTGCAGCCTGACTCTCGAATACCGTCGCGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTC CATGGACCCGTCTAAAGTTGGTGATCGTTTTCAGTATCGTCATCTGCTGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTA CCGAATGGCGTCCGAAAAACGCGGGCACTAATGGTGCTATCTCCACTGGGAAAACAtaa	
CpFatB1.2-M3	ATGCTGCTCACCGCGATTACCACCGTGTTCGTTGCGCCGGAAAAACGTTGGACTATGTTCGACCGCAAATCTAAACGTCCGAACAT GCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAATCCTTCTCCATTCGTAGCTACGAAATCT GTTCGGACCGCACGGCGAGCATCGAGACCGTGATGAACCACGTTCAGGAAACCAGCCTGAACCAGTGTAAGTCCATCGGCCTGCT GGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTAACCCGTATGAAAATTATGGTAAACCGTT ATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAATTGGCATGGGTCGCGACTGGCTGATTAG CGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACCAGAAAACGCGTCGCTTCTCTAAACTGCC CCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGAAGACAACGACGGTAAACTGCAGAAATTC GATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTGGACGTAAACCAGCACGTTTCCAATGTCA AATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTTTGCAGCCTGACTCTCGAATACCGTCGC GAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTGGTGATCGTTTTCAGTATCGTCATCTGCT GCGTCTGGAAGATGGTGCGGACATTATGAGGGGCCGTACCGAATGGCGTCCGAAAAACGCGGGCACTAATGGTGCTATCTCCACT GGGAAAACAtaa	This Work
CpFatB1.2-M4	ATGCTGCTCACCGCGATTACCACCGTGTTCGTTGCGCCGGAAAACGTTGGACTATGTTTGACCGCAAATCTAAACGTCCGAGCATG CTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAATCCTTCTCCATTCGTAGCTACGAAATCTGT GCGGACCGCACGGCGAGCATGGAGACCGTGATGAACCACGTTCAGGAAACCAGCCTGAACCAGTGTAAGTCCATCGGCCTGCTG GACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTAACCCGTATGAAAATTATGGTAAACCGTTA TCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAATTGGCATGGGTCGCGACTGGCTGATTAGC GATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACCAGAAAACGCGTCGCTTCTCTAAACTGCCC CACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGAAGACAACGACGGTAAACTGCAGAAATTCG ATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTGGACGTAAACCAGCACGTTTCCAATGTCAAA TATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTTTGCAGCCTGACTCTCGAATACCGTCGCGA ATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTGGTGATCGTTTTCAGTATCGTCATCTGCTGC GTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAAAAACGCGGGCACTAATGGTGCTATCTCCACTGG GAAAACAtaa	This Work
CpFatB1.2-M4-287	ATGTTTGACCGCAAATCTAAACGTCCGAGCATGCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCG CCAATCCTTCTCCATTCGTAGCTACGAAATCTGTGCGGACCGCACGGCGAGCATGGAGACCGTGATGAACCACGTTCAGGAAACC AGCCTGAACCAGTGTAAGTCCATCGGCCTGCTGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGG TTGTAACCCGTATGAAAATTATGGTAAACCGTTATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGT AAAATTGGCATGGGTCGCGACTGGCTGATTAGCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGAT GAACCAGAAAACGCGTCGCTTCTCTAAACTGCCCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCC ATTGAAGACAACGACGGTAAACTGCAGAAATTCGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGA CCTGGACGTAAACCAGCACGTTTCCAATGTCAAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAG AACTTTGCAGCCTGACTCTCGAATACCGTCGCGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAA GTTGGTGATCGTTTTCAGTATCGTCATCTGCTGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAA AAACGCGGGCACTAATGGTGCTATCTCCACTGGGAAAACAtaa	This Work
Repair template for chromosomal integration of	GCATTTATGCCGATGTGAACGGTTTTCTGGACCCGGTGGATTTCCGTGGACAACTGGTTACGGTAGTCGGGCCAATCACTGGTGC GGTTGACGGCAAAATCGGCAATACGCCCTATAAATTTATGGTGATGCAAGTAACGGGTTACAAACGTTGGCATTTAACCCAGCAGG TGATTATGCCGCCTCAGCCGATTGATCCATGGTTTTATGGCGGTCGTGGCTGGCCCTATGGCTACGGCGGATGGGGCTGGTATAA TCCCGGCCCCGCGAGAGTACAAACAGTTGTAACTGAATAATTGCTTGTTTTTAAAGAAAAAGAAACAGCGGCTGGTCCGCTGTTTCT	This Work

CpFatB1.2-M4-287 on the *fadD* loci. This creates NHL17 strain	GCATTCTTACGGTAAAGATAAAAATAAATAGTGACGCGCTTCGCAACCTTTTCGTTGGGTAATTATCAAGCTGGTATGATGAGTTAAT ATTATGTTAACGGCATGTATATCATTTGGGGTTGCGATGACGACGAACACGCATTTTAGAGGTGAAGAAcgaggacaccatcgaatggtgcaa aacctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgt ggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgct gattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtc gaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgtta tttcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccatt aagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgc aaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtggg atacgacgataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtg aagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcc cgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctggtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaa gctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttctggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttg acaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagaccatgaggaggattataaaATGTTTGACCGCAAATCTAAACGTCC GAGCATGCTGATGGATAGCTTCGGTCTGGAACGTGTTGTCCAGGATGGTCTGGTTTTCCGCCAATCCTTCTCCATTCGTAGCTACG AAATCTGTGCGGACCGCACGGCGAGCATGGAGACCGTGATGAACCACGTTCAGGAAACCAGCCTGAACCAGTGTAAGTCCATCGG CCTGCTGGACGACGGTTTCGGTCGCTCACCGGAAATGTGTAAACGCGACCTGATTTGGGTTGTAACCCGTATGAAAATTATGGTAA ACCGTTATCCGACCTGGGGCGACACCATTGAAGTCTCCACCTGGCTGTCCCAGTCCGGTAAAATTGGCATGGGTCGCGACTGGCT GATTAGCGATTGTAATACCGGTGAAATACTGGTTCGTGCAACCTCAGTTTACGCCATGATGAACCAGAAAACGCGTCGCTTCTCTAA ACTGCCCCACGAAGTACGCCAGGAGTTCGCGCCGCACTTCTTAGACTCCCCGCCGGCCATTGAAGACAACGACGGTAAACTGCAG AAATTCGATGTGAAAACCGGTGATTCCATCCGCAAAGGTCTGACCCCGGGCTGGTACGACCTGGACGTAAACCAGCACGTTTCCAA TGTCAAATATATCGGTTGGATCCTGGAATCCATGCCGACCGAAGTTCTGGAAACCCAAGAACTTTGCAGCCTGACTCTCGAATACC GTCGCGAATGCGGTCGCGACTCAGTACTTGAGAGCGTAACCTCCATGGACCCGTCTAAAGTTGGTGATCGTTTTCAGTATCGTCAT CTGCTGCGTCTGGAAGATGGTGCGGACATTATGAAGGGCCGTACCGAATGGCGTCCGAAAAACGCGGGCACTAATGGTGCTATCT CCACTGGGAAAACCAtaacttGgctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtag cgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaagg ctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcggatttgaacgttgcgaagcaacggcccggagggtggcgggcag gacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggccttttgcgtttctacaaactctttGCGTTAAGTCAGTCGTCAGACGCCGGTTA ATCCGGCGTTTTTTTTGACGCCCACTAAAGAGAAAACAATTTGAATTACCAAATGATTACCACGGACGATGCGCTGGCTTCTTTGTG TGAAGCCGTCCGTGCCTTTCCGGCGATAGCCCTGGATACTGAATTTGTTCGTACGCGCACTTATTACCCGCAGCTGGGGTTGATTC AACTTTTCGATGGCGAGCATCTGGCGCTAATCGATCCACTCGGGATCACCGACTGGTCACCGCTGAAAGCGATCCTGCGCGATCC GTCCATCACAAAATTTCTCCATGCAGGCAGTGAAGATCTGGAAGTGTTCCTCAATGTCTTTGGCGAATTACCACAACCCTTGATTGA CACGCAAATCCTTGCTGCCTTCTGCGGACGCCCGATGTCATGGGGTTTCGCTTCCATGGTGGAAGAGTATTCCGGCGTTACGCTG GACAAGAGTGAATCGCGCACCGACTGGCTGGCCAGACCGCTGACCGAACGTCAGTGTGAATAC	
Repair template for chromosomal deletion of *lipB* gene	TATCAGGATAAAATTCTTGTCCGCCAGCTCAATCCGGACTTCGAATATATTACCGCTta a	This Work

Appendix XI: Protein sequence of CpFatB1 and key derivatives

Protein name	Protein Sequence	reference
CpFatB1	MVAAAASSACFPVPSPGASPKPGKLGNWSSSLSPSLKPKSIPNGGFQVKANASAHPKANGSAVTLKSGSLN TQEDTLSSSPPPRAFFNQLPDWSMLLTAITTVFVAPEKRWTMFDRKSKRPNMLMDSFGLERVVQDGLVFR QSFSIRSYEICADRTASIETVMNHVQETSLNQCKSIGLLDDGFGRSPEMCKRDLIWVVTRMKIMVNRYPTWG DTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVRATSVYAMMNQKTRRFSKLPHEVRQEFAPHFLDSPPAIE DNDGKLQKFDVKTGDSIRKGLTPGWYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYRRECGRD SVLESVTSMDPSKVGDRFQYRHLLRLEDGADIMKGRTEWRPKNAGTNGAISTGKT*	136
CpFatB1.2	MLLTAITTVFVAPEKRWTMFDRKSKRPNMLMDSFGLERVVQDGLVFRQSFSIRSYEICADRTASIETVMNHV QETSLNQCKSIGLLDDGFGRSPEMCKRDLIWVVTRMKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLIS DCNTGEILVRATSVYAMMNQKTRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRKGLTPG WYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYRRECGRDSVLESVTSMDPSKVGDRFQYRHLL RLEDGADIMKGRTEWRPKNAGTNGAISTGKT*	31
CpFatB1.3	MKSKRPNMLMDSFGLERVVQDGLVFRQSFSIRSYEICADRTASIETVMNHVQETSLNQCKSIGLLDDGFGRS PEMCKRDLIWVVTRMKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVRATSVYAMMNQK TRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRKGLTPGWYDLDVNQHVSNVKYIGWILE SMPTEVLETQELCSLTLEYRRECGRDSVLESVTSMDPSKVGDRFQYRHLLRLEDGADIMKGRTEWRPKNA GTNGAISTGKT*	This Work
CpFatB1.4	MGLVFRQSFSIRSYEICADRTASIETVMNHVQETSLNQCKSIGLLDDGFGRSPEMCKRDLIWVVTRMKIMVN RYPTWGDTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVRATSVYAMMNQKTRRFSKLPHEVRQEFAPHFL DSPPAIEDNDGKLQKFDVKTGDSIRKGLTPGWYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYR RECGRDSVLESVTSMDPSKVGDRFQYRHLLRLEDGADIMKGRTEWRPKNAGTNGAISTGKT*	This Work
CpFatB1.2-M3	MLLTAITTVFVAPEKRWTMFDRKSKRPNMLMDSFGLERVVQDGLVFRQSFSIRSYEICSDRTASIETVMNHV QETSLNQCKSIGLLDDGFGRSPEMCKRDLIWVVTRMKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLIS DCNTGEILVRATSVYAMMNQKTRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRKGLTPG WYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYRRECGRDSVLESVTSMDPSKVGDRFQYRHLL RLEDGADIMRGRTEWRPKNAGTNGAISTGKT*	This Work
CpFatB1.2-M4	YAAHRDYHRVRCAGKRWTMFDRKSKRPSMLMDSFGLERVVQDGLVFRQSFSIR SYEICADRTASMETVMNHVQETSLNQCKSIGLLDDGFGRSPEMCKRDLIWVVTR MKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVRATSVYAMM NQKTRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRKGLTPG WYDLDVNQHVSNVKYIGWILESMPTEVLETQELCSLTLEYRRECGRDSVLESVTS MDPSKVGDRFQYRHLLRLEDGADIMKGRTEWRPKNAGTNGAISTGKT*	This Work
CpFatB1.2-M4-287	MFDRKSKRPSMLMDSFGLERVVQDGLVFRQSFSIRSYEICADRTASMETVMNHVQETSLNQCKSIGLLDDG FGRSPEMCKRDLIWVVTRMKIMVNRYPTWGDTIEVSTWLSQSGKIGMGRDWLISDCNTGEILVRATSVYAM MNQKTRRFSKLPHEVRQEFAPHFLDSPPAIEDNDGKLQKFDVKTGDSIRKGLTPGWYDLDVNQHVSNVKYI GWILESMPTEVLETQELCSLTLEYRRECGRDSVLESVTSMDPSKVGDRFQYRHLLRLEDGADIMKGRTEWR PKNAGTNGAISTGKT*	This Work

Appendix XII: Protein sequence of FadM and homologs

Organism	Protein name	Protein Sequence	reference
Escherichia coli	Ec-FadM	MQTQIKVRGY HLDVYQHVNN ARYLEFLEEA RWDGLENSDS FQWMTAHNIA FVVVNININYRRPAVLSDLL TITSQLQQLN GKSGILSQVI TLEPEGQVVA DALITFVCID LKTQKALALEGELREKLEQM VK	This Work
Salmonella enterica	Se-FadM	mqtqikvrgy hldvyqhvnn arylefleea rwdglensds fqwmtarnia fvvvnininy rrpavlsdll tvtsqvqqln gksgvlsqti tlepegqvva dalitfv	This Work
Neisseria gonorrhoeae	Ng-FadM	mkltvrnyhl dgyghvnnar ylefleearw affekrglmh elagliliva ridirysrpa vegdvlqfsc rlktpgmrri vltqtitlpn gktaaeadit lmpvhaatqr tvslpatlar alealse	This Work
Providencia sneebia	Ps-FadM	matqikvygh hidvfnhvnn arylefyeea rwawlenhnl lnfllknnlg mvavnininy cqgavlfdql tvisrlerig tksascyqqi ireknggktl isdvtvtfvf velatnksvv isgellehle pllqgesdqf ievn	This Work
Marinomonas ushuaiensis	Mu-FadM	miniiqikvr gyhldifqhv nnarylefle egrwaffdef gagtelmeqg lawavvnini dfraegnfad vlevhtqfsk lgnrsvtmkq rimnaktdtl vaeadvtyvc fskekkaavp lpddykmkiq kaleesd	This Work
Serratia odorifera	So-FadM	mqtlikvrgy hldvyqhvnn arylefleea rwewldnlda frwmtennia fivvnininy rkpavlgdrl ridsrmeqln gksgvlsqkv tlepqgsava dalltfvcvd lktqralpie gelrehlmsl qqpvinatgd grl	This Work
Cucumis sativus	Cs-FldM	mqtqikvrgy hldvyqhvnn arylefleea rwdglenses fqwltahnia fvvvnininy rrpavlgdvl tvtsevqqln gksgvlsqvv tlepegqvva dalitfvcid lktqkalple gelreklelm ia	This Work
Beauveria bassiana	Bb-FudM	MQTHIKVRGYHMDVYQHVNNARYLEFLEEARWEGLEKTTGFQWMTEHNIAFIVVNINISYRRPAVLGDLLRI ESSLQQLNGKSGVLSQVVTLEPEGEAVADALLTFVCIDLKTQKAVPLEGELRDKLEQMMNA	This Work
Buttiauxella agrestis	Ba-FadM	mqthikvrgy hmdvyqhvnn arylefleea rweglentag fqwmtehnia fivvnininy rrpavlgdll riesslqqln gksgvlsqvv klepegeiva dalltfvcid lktqksvple gelreklvqm me	This Work
Cronobacter turicensis	Ct-FadM	mqtqikvrgy hidvyqhvnn arylefleea rwdglendes fkwmmannia fivvnininy rrpavlgdll tvtsqvkqln gksgilsqii tlepegevva dalitfvcid lktqkalpie gelreklekl tg	This Work
Pseudomonas aeruginosa	Pa-FadM	mssehkilar ipisvrwrdm dsmghvnnak yisyleearv rwmlgvdgvs mtdriapvvaatnvnyrlpi vwpndivvel fverlgnssv tighrivdqq dasklysdgn vvvvwmdtqtgksaplpeai rnast	This Work
Mycobacterium tuberculosis	Mt-FadM1	msttikvigy hidafghvnn arylefleaa rwdwlesyda yrwfkqmdia vvvvnininyrlpvyvgerl vidsylqhag qksgvlkqii trqedkqvva daevtfvfin mktgkaipiegeirdkfell tapkeg	This Work
Mycobacterium tuberculosis	Mt-FadM2	mtfsvpvtvr gyeldvnghl nqavyhqyae harwevlraa glvpdkmrls glgpvvlestvkyrrelhlg deitvtcecr wgegkafwmd qqirkldgtv saefsvvlgl mdldarklvsnpgkrflela espdvlgl	This Work
Mycobacterium tuberculosis	Mt-FadM3	msgtgfvtpv avrwsdidmy qhinhatmvt lleearipfl spvfgaeiit tglliaevrisykgqlrlld splqvtitvd rlravdftlg yevrsvnaap dsrpaviaet qlaafdideqklvrlsaehr eylgrfvr	This Work
Mycobacterium tuberculosis	Mt-FadM4	msagfvapvp vrwsdidmyq hinhatmvtl leearvpflk epfaaditti glliadvrvtykdqlrlads plqvtiwtkr lravdftlgy evrsvaadpe skpavvaesq laavhieeqrlvrlspqhre ylqrwmr	This Work
Mycobacterium tuberculosis	Mt-FadM5	mtvgfvaqvp vrwsdidmyq hvnhatmvtl leearvpflk pafevdilei glliadvrvtykaqlklvds plqvtvwtkq lrtvdftlgy evrsvgadpq skpaviaesq laavhiqeqrlvrlgpqhre ylqrwlr	This Work