Functional screening and in vitro analysis reveals thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Functional screening and \textit{in vitro} analysis reveals thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in \textit{Escherichia coli}

Matthew D. McMahon\textsuperscript{a} and Kristala L.J. Prather\textsuperscript{a,b,}\textsuperscript{#}

\textsuperscript{a}Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; \textsuperscript{b}Synthetic Biology Engineering Research Center (SynBERC), Massachusetts Institute of Technology, Cambridge, MA, USA.

Running Head: Thioesterases for short-chain fatty acid biosynthesis

#Address Correspondance to Kristala L.J. Prather, kljp@mit.edu
Abstract

Short-chain fatty acid (SCFAs) biosynthesis is pertinent to production of biofuels, industrial compounds, and pharmaceuticals from renewable resources. To expand on *Escherichia coli* SCFA products, we previously implemented a coenzyme A (CoA)-dependent pathway that condenses acetyl-CoA to a diverse group of short-chain fatty acyl-CoAs. To increase product titers and reduce premature pathway termination products, we describe *in vivo* and *in vitro* analyses to understand and improve the specificity of the acyl-CoA thioesterase enzyme, which releases fatty acids from CoA. A total of 62 putative bacterial thioesterases, including 23 from the cow rumen microbiome, were inserted into a pathway that condenses acetyl-CoA to an acyl-CoA molecule derived from exogenously provided propionic or isobutyric acid. Functional screening revealed thioesterases that increase production of saturated (valerate), unsaturated (trans-2-pentenoate) and branched (4-methylvalerate) SCFAs compared to overexpression of *E. coli* thioesterase *tesB* or native expression of endogenous thioesterases. To determine if altered thioesterase acyl-CoA substrate specificity caused the increase in product titers, six of the most promising enzymes were analyzed *in vitro*. Biochemical assays revealed that the most productive thioesterases rely on promiscuous activity, but have greater specificity for product-associated acyl-CoAs than for precursor acyl-CoAs. Here we introduce novel thioesterases with improved specificity for saturated, branched and unsaturated short-chain acyl-CoAs, thereby expanding the diversity of potential fatty acid products while increasing titers of current products. The growing uncertainty associated with protein database annotations denotes this study as a model for isolating functional biochemical pathway enzymes in situations where experimental evidence of enzyme function is absent.
1. Introduction

The potential for producing biofuels, industrial compounds, and pharmaceuticals from renewable resources has led to an increased interest in short-chain (C2 – C7) fatty acid (SCFA) biosynthesis. Developing recombinant strains to produce these molecules could lead to production of polymers such as polyhydroxyalkanoates and pharmaceuticals such as statins from fossil fuel alternatives (1, 2). Supporting the development of microbially-synthesized products is the reduced cost and increased prevalence of genomic sequencing. The resulting profusion of metabolic diversity provides a wealth of potential enzymes with known genetic sequences for improving biosynthetic pathways. For example, recent genomic sequencing has unveiled the metabolic diversity of important members from the cow rumen microbiome, an environment rich in SCFAs (3, 4). These genome sequences provide an opportunity to find enzymes that improve production and specificity in SCFA biosynthesis pathways.

One such pathway that would benefit from improved enzyme specificity is Coenzyme A (CoA)-dependent biosynthesis of SCFAs. In the CoA-dependent pathway, a thiolase enzyme condenses an acyl-CoA molecule with acetyl-CoA (Figure 1). The resulting 3-ketoacyl-CoA molecule is then sequentially reduced by reductase, crotonase, and enoyl-reductase enzymes before the 3-hydroxy, unsaturated, and saturated fatty acids, respectively, are cleaved from CoA using a thioesterase enzyme. Previous work from our lab has used this pathway with acetyl-CoA and glycolyl, propionyl, or isobutyryl-CoA as the condensed substrates to produce a variety of SCFAs using *E. coli* TesB including 3-hydroxyvalerate (3-hydroxypentanoate) (1), dihydroxybutyrate (5), 3-hydroxy-4-methylvalerate (5), and a variety of alcohols (6). One benefit of CoA-dependent biosynthesis of SCFAs is the diversity of potential products (7); however, with this diversity comes a need for selective enzymes that increase final product titers by
minimizing substrate flux to undesired byproducts. The selectivity of the final enzyme in the pathway, the thioesterase, is of particular importance because it influences the product profile by catalyzing fatty acid release from CoA at each step of the pathway (Figure 1) and is important for secretion of fatty acid products (8). While many acyl-ACP thioesterases have been investigated for improved fatty acid production, acyl-CoA thioesterases are not as well explored (8-10). Despite its preference for acyl-CoAs in the C14-C18 range (11), the *E. coli* acyl-CoA thioesterase TesB produces diverse SCFAs (12). However, locating more selective thioesterases may reduce byproduct formation and increase final product titers.

Selecting individual thioesterase enzymes for functional screening against short-chain fatty acyl-CoAs is challenging because much of their vast phylogenetic and functional diversity is poorly understood (13). While many thioesterases have been explored for long-chain fatty acid production (14-17), few studies have focused on those that prefer short-chain acyl-CoAs. Several broad specificity acyl-CoA thioesterases including *E. coli* TesB and *Saccharomyces cerevisiae* Pte1p can be used for SCFA production, but lack the specificity necessary for optimizing biosynthetic pathways (11, 18). One approach to selecting thioesterases for functional screening that improves the likelihood of finding enzymes with the desired specificity is to investigate those proteins with similarity to commonly used and effective enzymes. However, the absence of known selective short-chain acyl-CoA thioesterases restricts this approach. Sampling candidates more broadly will increase opportunities for finding enzymes with new substrate specificities, but will also increase the number of thioesterases with undesired activities. Some combination of these routes can be used to screen sufficient phylogenetic breadth while also increasing the sample size of TesB-like thioesterases to reveal enzymes with greater specificity for short-chain acyl-CoAs.
To address this need, we functionally screened 62 putative thioesterases in a pathway for production of the SCFAs 3-hydroxyvalerate, trans-2-pentenoate, valerate, and 4-methylvalerate in *E. coli*. Six enzymes were homologous to variants active on short-chain acyl-CoAs, while the remaining 56 represent all of the annotated thioesterases from five bacterial strains of interest. Based upon *in vivo* fatty acid titers of those recombinant strains, six enzymes were chosen for *in vitro* analysis on a broad range of acyl-CoAs to determine their substrate preferences. The combination of *in vivo* and *in vitro* data indicate that we have uncovered thioesterases with greater specificity for and production of unsaturated, saturated, and branched SCFAs in *E. coli* relative to TesB and endogenous thioesterase activity.

### 2. Methods

#### 2.1. Bacterial Strains

Rumen isolate *Prevotella ruminicola* 23 was obtained from Dr. Roderick Mackie of the University of Illinois, Urbana-Champaign (USA). Genomic DNA from rumen isolates *Fibrobacter succinogenes* S85 and *Ruminococcus albus* 7 were obtained from Dr. Paul Weimer of the US Dairy Forage Research Center, USDA-Agricultural Research Service, Madison (USA). *Alcanivorax borkumensis* SK2 (ATCC# 700651) and genomic DNA from *Pseudomonas aeruginosa* PAO1 (ATCC# 47085) were purchased from the ATCC. Genomic DNA of *Rhodopseudomonas palustris* CGA009 was obtained from Dr. Caroline Harwood of the University of Washington, Seattle (USA). *Pseudomonas syringae* pv. *maculicola* ES4326 was obtained from Dr. Fred Ausubel at Massachusetts General Hospital (Boston, MA, USA). *E. coli* MG1655 (DE3) Δ*endA* Δ*recA* was previously constructed in our lab (1).
2.2. Plasmid and Strain Construction

Plasmid pET/ter/bktB/pct was constructed previously (6) and plasmid pCDF/phaB/phaJ4 was constructed by subcloning phaB from pET/bktB/phaB (5) into multiple cloning site I by NdeI/XhoI restriction digest and phaJ4 into multiple cloning site II of pCDFDuet-1 (EMDmillipore) using primers listed in Table S1. Plasmid pET/bktB/pct was constructed from pET/ter/bktB/pct by BamHI/NotI restriction digest followed by treatment with Mung Bean Nuclease (New England Biolabs) and blunt ligation of the 8 kb fragment.

Genomic DNA was isolated from bacterial strains using the Wizard Genomic DNA Purification Kit (Promega). Custom oligonucleotides were purchased for the PCR amplification of all individual thioesterases and CoA ligases from purified genomic DNA (Integrated DNA Technologies). Primers used for amplification are listed in Table S1. Following amplification, individual genes were inserted into the expression vector pACYCDuet-1 (EMDmillipore) using polymerase incomplete primer extension (PIPE)-based cloning (19). The genes encoding R. palustris CoA ligases FcsA (Rpa4267) and VcsA (Rpa3299) (20) and the genes encoding the six thioesterases chosen for further analysis were inserted into vector pTEV5 for protein purification using PIPE-based methods (21). The pTEV5 construct produced an enzyme with an N-terminal hexahistidine tag removable by TEV protease. Due to solubility problems in pTEV5, the genes encoding thioesterases Pr1687 and Fs2108 were amplified from genomic DNA and cloned using restriction enzymes NdeI and EcoRI into pTYB22 (New England Biolabs) which produced an enzyme with an N-terminal chitin-binding domain removable by intein self-cleavage. Plasmid sequences were confirmed using PCR amplification and DNA sequencing (GENEWIZ).

Gene deletions of yciA, yigI, and tesB in E. coli MG1655 (DE3) ΔendA ΔrecA were made using P1 transduction with strains JW1245-1, JW5588-1, and JW0442-1, respectively, from the
Keio collection as donor cells (22). The kanamycin resistance gene was removed using FLP-mediated recombination as previously described (23).

2.3. Culture Conditions

Recombinant strains of *E. coli* MG1655 (DE3) Δ*endA* Δ*recA* were grown at 30°C in Luria-Bertani (LB) medium overnight in a shaking incubator at 250 rpm. 50 µL of the overnight culture was used to inoculate a 50 mL LB culture supplemented with 10 g/L glucose containing 100 mg/L ampicillin, 50 mg/L streptomycin, and, when pACYCDuet-1 was present, 16 mg/L chloramphenicol. Cultures were grown at 30°C until an optical density at 600 nM (OD$_{600}$) of 0.8 was reached, at which point isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM with either propionate or isobutyrate to a final concentration of 15 mM. Cultures were incubated at 30°C for 48 h post-induction prior to fatty acid titer determination.

2.4. Fatty Acid Analysis

Culture samples were centrifuged to pellet cells and culture supernatant was removed for HPLC analysis. A 5 µl sample of culture supernatant was injected into an Agilent 1100 series instrument equipped with refractive index detection (RID). Samples were processed through an Aminex® HPX-87H anion-exchange column (Bio-Rad Laboratories) with isocratic flow of 5 mM H$_2$SO$_4$ mobile phase at a rate of 0.6 mL/min and column and detector temperatures set to 35°C. Concentrations of valerate, 3-hydroxyvalerate, trans-2-pentenoate, and 4-methylvalerate were determined using linear regression of external standards.

2.5. Protein Purification
Acyl-CoA ligases FcsA and VcsA and thioesterases EcTesB, PpTesB, EcYdiI and Pr655 were overproduced using pTEV5 constructs in *E. coli* BL21Star (DE3) (Invitrogen). One liter of cells were grown at 30°C in LB medium containing 100 mg/L ampicillin until an OD₆₀₀ of 0.5 was reached, at which point IPTG was added to the cultures at a concentration of 100 mg/L. Post induction, the cells were grown for 15 h at 30°C, then harvested by centrifugation and resuspended in 2.5x v/w Buffer A (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% v/v glycerol). Protein purification followed previously described protocols for nickel chelate chromatography followed by cleavage with TEV protease (24). Proteins were flash frozen in liquid nitrogen and the concentration was determined using the Bradford assay with bovine serum albumin as a standard (25) (Bio-Rad).

Pr1687 and Fs2108 protein was purified using the Intein-Mediated Purification with an Affinity Chitin-binding Tag (IMPACT) expression vector pTYB22 (New England Biolabs). Cells were grown and induced using the same conditions as described for pTEV5 constructs and proteins were purified with chitin-affinity chromatography followed by intein cleavage mediated by dithiothreitol (DTT) using previously described conditions (26). Proteins were stored and quantified as described for nickel chelate chromatography purified proteins.

2.6. Enzymatic Synthesis of Acyl-CoAs

Acyl-CoA synthesis reactions were carried out in 50 mM HEPES pH 7.5, 1 mM DTT, 5 mM MgCl₂, 5 mM ATP, and 2 mM CoA. Fatty acid substrate was added at 7.5 mM for butyrate, 3-hydroxyvalerate, trans-2-pentenoate, valerate, 4-methylvalerate, hexanoate, octanoate, decanoate, and dodecanoate while 2 mM fatty acid substrate was used for tetradecanoate. To increase solubility, 1% and 3% w/v triton-X 100 was added to reactions with dodecanoate and
tetradecanoate respectively. CoA ligase enzymes were added at 500 nM for all reactions. VcsA was added to butyrate, 3-hydroxyvalerate, valerate, 4-methylvalerate, and hexanoate reactions, while FcsA was added to octanoate, decanoate, dodecanoate, and tetradecanoate reactions. Reactions were run overnight at room temperature for all substrates except trans-2-pentenoate, which was run at 30°C for 6 h because these conditions reduced the appearance of degradation products. Acyl-CoA ligases were precipitated from reactions at 95°C for 5 min, and then removed by centrifugation. Acyl-CoA products were purified from substrates using an Agilent 1200 series HPLC with diode array detection (DAD). A 100 µL reaction volume was injected onto an Agilent Eclipse XDB-C18 column and separation achieved using a mobile phase of 50 mM ammonium acetate, 0.1% m/v acetic acid (Solvent A) - 50 mM ammonium acetate, 0.1% m/v acetic acid, 70% v/v acetonitrile (Solvent B) gradient. The method began at 100% Solvent A from 0-5 min, followed by a 0-100% gradient of solvent B from 5-50 minutes, followed by an isocratic step of 100% solvent B from 50-55 min. The gradient was run at a flow rate of 1 mL/min and CoA was monitored by measuring absorbance at 258 nM. Fractions containing peaks corresponding to acyl-CoAs were collected, flash frozen in liquid nitrogen, and lyophilized. Dried acyl-CoAs were then resuspended in water and the concentration was determined by the absorbance at 258 nM using the molar extinction coefficient of CoA (14,328 M⁻¹ cm⁻¹) within the linear range of detection (27).

2.7. Thioesterase Activity Assays

Thioesterase activity was measured using 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB) which reacts with free CoA after it is released through thioesterase-mediated bond cleavage.
Reactions were carried out in 1 mM DTNB, 100 mM HEPES pH 8.0, 20 µM acyl-CoA and were run in the linear range of each thioesterase. Substrate concentrations matched those used previously for EcTesB analysis (11). Reaction progress was monitored through the change in absorbance at 412 nm using the molar extinction coefficient of 5-thio-2-nitrobenzoate (14,150 M$^{-1}$ cm$^{-1}$), which is formed when DTNB reacts with free CoA (28).

3. Results

3.1 In Vivo Functional Screening for Thioesterases that Alter Fatty Acid Production

Acyl-CoA thioesterase substrate specificity plays an important role in deciding the SCFA profile produced using a CoA-based pathway. For example, if the target product is a saturated fatty acid, a thioesterase with broad substrate specificity will release fatty acids at each step in the biosynthetic pathway, depleting substrate pools and reducing the final product titer (Figure 1). To find thioesterases that increase specificity for and production of 3-hydroxy, unsaturated, saturated and branched fatty acids in the pathway, 62 putative thioesterases were chosen for screening. A group of six thioesterases, three TesB and three TesB2 enzymes, were chosen that had 38-50% amino acid identity to *E. coli* MG1655 TesB (EcTesB), which makes 3-hydroxy and saturated fatty acids *in vivo* (1, 6, 29), and 40-41% amino acid identity to *A. borkumensis* TesB2 enzyme, which was previously described as having specificity for 3-hydroxy acyl-CoAs (30) (Table 1).

To incorporate greater phylogenetic and functional diversity of thioesterases, the remaining 56 proteins encompassed all of the annotated thioesterases from five bacterial strains (Table 1). *E. coli* MG1655 (Accession: NC_000913.2) thioesterases were chosen with the dual purpose of identifying those that increase product titers of desired SCFAs for use in heterologous
pathways and those that reduce product titers and should be targeted for deletion from our host
strain. The second source of thioesterases, *Pseudomonas putida* KT2440 (Accession:
NC_002947.3), was chosen for its phylogenetic similarity to *E. coli* and because it is a known
producer of polyhydroxyalkanoates, which may indicate the presence of thioesterases with
specificity to 3-hydroxy acyl-CoAs (31). The remaining three organisms, *Prevotella ruminicola*
23 (Accession: NC_014033.1), *Fibrobacter succinogenes* S85 (Accession: NC_017448.1), and
*Ruminococcus albus* 7 (Accession: NC_014833.1), were chosen because they are prevalent in the
cow rumen microbiome and contribute to the high concentrations of SCFAs found there (3, 4,
32).

Each of the 62 putative thioesterases was individually overexpressed in *E. coli* containing
all the necessary genes for CoA-dependent biosynthesis of valerate (Figure 1). Previous work
from our lab identified *Treponema denticola* Ter, *Megaphaera elsdenii* Pct, and *Cupriavidus
nector* (formerly *R. eutropha* H16) BktB, PhaB, and PhaJ4 as suitable upstream pathway
enzymes for valerate production (6, unpublished data). Cells were grown in LB containing
glucose until log phase growth, then pathway genes were induced and cultures were
supplemented with either propionate or isobutyrate for straight or branched SCFA biosynthesis,
respectively (Figure 1). LC-based analysis of the culture supernatant was used to screen for
enzymes that altered the fatty acid product profiles. Specifically, the resulting chromatograms
were examined for increases and decreases of 3-hydroxyvalerate, trans-2-pentenoate, valerate,
and 4-methylvalerate. Profiles produced by recombinant strains were compared to controls with
either no recombinant thioesterase or with overproduced EcTesB, a thioesterase used previously
in our group (1, 5, 6).
Overproduction of greater than 20% of the thioesterases functionally screened in this study resulted in observable changes in substrate, intermediate, and product titers relative to the control strains (Table S2). Four thioesterases, EcYciA, Pp1466, Pp3807, and Pp4975, were associated with significant reduction in valerate and 4-methylvalerate production combined with increased acetate and propionate titers (Table S2). Two thioesterases, Pr1510 and Fs368, were associated with a production phenotype involving reduced titers of measured substrates, intermediates, and final products combined with increased glucose consumption (Table S2). The most logical explanation for this phenotype is channeling of substrates to long-chain fatty acid biosynthesis. Because long-chain fatty acids could not be quantified with our HPLC system, additional experiments outside the scope of this work are required to determine the activity profile of these enzymes.

3.2 Host Strain Development and Identification of Thioesterases for In Vitro Analysis

From the 12 annotated E. coli thioesterases screened during this study, three were chosen for deletion from the host strain. The tesB gene was deleted because its overexpression resulted in increased titers of 3-hydroxyvalerate, which could be a final product, but also acts as a shunt product in the formation of trans-2-pentenoate or valerate; yciA was deleted because its overexpression increased final titers of the precursor-derived and exogenously supplied acids acetate and propionate, which correlates with previous work (33), while decreasing final product titers of valerate and 4-methylvalerate; and yigI was deleted because its overexpression resulted in decreased production of 4-methylvalerate and increased final acetate and isobutyrate titers. The resulting triple thioesterase deletion strain E. coli MG1655 (DE3) ΔendA ΔrecA ΔtesB ΔyciA ΔyigI was used for further in vivo analysis of several active thioesterases.
From the full set of 62 functionally screened thioesterases, EcTesB, PpTesB, EcYdiI, Fs2108, Pr655, and Pr1687 were chosen for more detailed \textit{in vivo} and \textit{in vitro} analysis because their overproduction resulted in increased titers of 3-hydroxyvalerate, trans-2-pentenoate, valerate, or 4-methylvalerate. Propionate and isobutyrate feeding experiments were performed in triplicate with these six thioesterases in both \textit{E. coli} MG1655 (DE3) \textit{ΔendA ΔrecA} and the triple thioesterase deletion strain containing pET/\textit{ter}/bktB/pct and pCDF/phaB/phaJ4 in an effort to distinguish recombinant thioesterase activity from background and potentially increase final product titers (Figure 2a, b). The greatest 3-hydroxyvalerate titers were produced in recombinant strains overproducing TesB enzymes from \textit{E. coli} and \textit{P. putida} when compared with the other 60 putative thioesterases (Figure 2). The maximum titer (2.163 g/L) resulted from overproduction of EcTesB (Table 2), and is similar to previously published titers from our lab (1). 3-hydroxyvalerate titers were 20% lower in the triple deletion strain overproducing EcTesB than in \textit{E. coli} with native thioesterases.

Observable titers of unsaturated SCFAs were uncommon among the recombinant strains with only Pr655 overproduction resulting in detectable trans-2-pentenoate during initial screening (Figure 2a, b). To determine whether strains containing the other five thioesterases selected for further analysis could produce detectable unsaturated SCFAs \textit{in vivo}, each was introduced into a strain lacking the downstream enoyl-reductase gene \textit{ter}, which does not produce the favored substrate valeryl-CoA (Figure 2c, d). In these recombinant strains, product titers of trans-2-pentenoate were greatest for EcYdiI, Pr655, and Pr1687, with EcYdiI overproduction resulting in the greatest trans-2-pentenoate titer, 695 mg/L (Table 2). To our knowledge, this is the largest published titer of an unsaturated SCFA from an engineered pathway in \textit{E. coli}. 

13
Three thioesterases, Fs2108, Pr1687, and EcYdiI, were chosen for further analysis because their overproduction increased titers of the saturated acids valerate and 4-methylvalerate while decreasing shunt product titers (Figure 2a, b). Overproduction of thioesterase Fs2108 achieves the greatest final titers for both the straight SCFA valerate and the branched SCFA 4-methylvalerate (Table 2) with 3-hydroxy and unsaturated fatty acid intermediate titers below the limit of detection. While overproduction of Fs2108 resulted in a modest 15% improvement in valerate titer over the EcTesB control, an improvement of greater than 200% was observed for 4-methylvalerate titer over the same control. This 200% improvement was possible in part because 4-methylvalerate titers were significantly lower than valerate titers for all 62 thioesterases tested, leaving more room for improvement. Overproduction of Fs2108 in the deletion strain resulted in approximately 10% more valerate and 30% more 4-methylvalerate than in the host containing all native thioesterases, which likely results from reduced hydrolysis of 3-hydroxyacyl-CoA and trans-2-acyl-CoA precursors. No 4-methylvalerate production was observed from the triple deletion strain when isobutyrate was supplied without thioesterase overexpression, which indicates that the remaining native thioesterases have poor activity on 4-methylvaleryl-CoA (Figure 2a, b).

3.3 Determination of In Vitro SCFA Substrate Specificity for Active Thioesterases

To determine the substrate preferences for the selected thioesterases, all six were overexpressed in E. coli BL21Star (DE3) and purified. Overexpression of Pr1687 and Fs2108 required the use of a vector that inserted an N-terminal chitin binding domain to obtain soluble protein. EcTesB, PpTesB, EcYdiI, and Pr655 were soluble with an N-terminal hexahistidine tag. Thioesterases were then tested for in vitro acyl-CoA hydrolysis activity on 20 μM acetyl-CoA,
butyryl-CoA, valeryl-CoA, hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, dodecanoyl-CoA and tetradecanoyl-CoA using an Ellman’s reagent-based assay described previously for determining specific activity (11) (Figure 3a). EcTesB, Fs2108 and Pr1687 showed a preference for longer-chain acyl-CoAs. PpTesB also showed a preference for longer acyl-CoAs, but the relationship was less linear than for EcTesB, Fs2108 and Pr1687. Pr655 showed a strong preference for butyryl and valeryl-CoAs and activity on acyl-CoAs longer than hexanoyl-CoA was below the limit of detection for this assay. EcYdiI had low activity for all substrates tested. A recent publication showing that EcYdiI has strong activity on the aromatic compound 1,4-dihydroxy-2-naphthoyl-CoA provides justification for the weak activity of this enzyme on the substrates provided in this study (34).

Specific activities were also measured for all six thioesterases on 3-hydroxy, unsaturated, saturated, and branched valeryl-CoAs (Figure 3b). For 3-hydroxyvaleryl-CoA, EcTesB and PpTesB had 6-fold greater specific activity than the next best thioesterase. Both EcYdiI and Pr655 showed low activity for 3-hydroxyvaleryl-CoA, while specific activity of Fs2108 and Pr1687 for the same substrate was below the limit of detection. No thioesterase had specific activity greater than 2 µM CoA min$^{-1}$ mg protein$^{-1}$ on trans-2-pentenoyl-CoA. Pr655, EcTesB, PpTesB, and EcYdiI displayed similar specific activities for this substrate; however, the small differences in specific acitivity translated into significant deviations in trans-2-pentenoate product titers when the thioesterases were overproduced in recombinant strains lacking the downstream enzyme Ter (Figure 3b). Fs2108 and Pr1687 specific activity on trans-2-pentenoyl-CoA was below the limit of detection, but in vivo product profiles of strains lacking ter reflect a slight preference for 3-hydroxyvaleryl-CoA over trans-2-pentenoyl-CoA (Figure 2d).
All six enzymes selected for further analysis had detectable activity on valeryl-CoA; however, EcTesB, PpTesB, Pr655 and Fs2108 had much greater activity relative to EcYdiI and Pr1687 (Figure 3b). Activity against the 4-methylvaleryl-CoA (branched) substrate was greater than or equal to activity on straight valeryl-CoA for EcTesB, PpTesB, Fs2108, and Pr1687. EcYdiI and Pr655 both showed weak activity against 4-methylvaleryl-CoA. In vitro analysis of the enzymes associated with the greatest titers of valerate revealed that Fs2108 had greater specific activity for both valeryl-CoA and 4-methylvaleryl-CoA than either Pr1687 or EcYdiI (Figure 3b). Further, Fs2108 specific activity for both 3-hydroxyvaleryl-CoA and trans-2-pentenyl-CoA precursors was below the limit of detection for in vitro assays. The greater specific activity of Fs2108 on 4-methylvaleryl-CoA and valeryl-CoA combined with reduced specific activity on precursor acyl-CoAs likely accounts for the increased 4-methylvalerate and valerate titers found for Fs2108.

4. Discussion

In this study, 62 putative thioesterases were screened for increased product titers of 3-hydroxy, unsaturated, saturated and branched products of interest. From these, six were chosen for more thorough in vivo and in vitro analysis. Relating the specific activity profiles of EcTesB, PpTesB, EcYdiI, Fs2108, Pr655 and Pr1687 with their in vivo product profiles of 3-hydroxyvalerate, trans-2-pentenoate, valerate, or 4-methylvalerate leads to several important conclusions. First, the enzymes with the greatest in vivo product titers always had lower specific activity for the associated acyl-CoA than for alternative CoA substrates. For example, Fs2108 has a six-fold higher specific activity for tetradecanyl-CoA than for valeryl-CoA indicating that valeryl-CoA is not the enzyme’s preferred substrate. The disparity between the specific activities
of Fs2108 for these two substrates indicates that our pathway relies on the promiscuous thioesterase activity of Fs2108. This is particularly evident for two thioesterases, EcYdiI and Pr1697, that were associated with increased valerate production despite having much higher specific activity for long-chain acyl-CoAs in the case of Pr1687, or aromatic acyl-CoAs for EcYdiI (34). The fact that all six thioesterases that were investigated in vitro prefer alternate substrates over those provided in our pathways suggests that both protein engineering and future bioprospecting efforts could further improve on the short-chain fatty acyl-CoA thioesterases discovered here.

Comparing the in vivo product titers with in vitro substrate specificities of the six chosen thioesterases also suggests that after a specific activity level of 1-5 µM CoA min⁻¹ mg protein⁻¹ is reached for a given acyl-CoA, the enzyme’s activity for the pathway precursor acyl-CoAs becomes an important factor influencing final product titer, illustrating the importance of selecting pathway thioesterases with reduced activity on precursor acyl-CoAs. For example, strains overproducing thioesterases Pr1687 and EcYdiI produce more valerate than those overproducing EcTesB despite EcTesB having 30-fold greater specific activity for valeryl-CoA than Pr1687 and EcYdiI (Figure 2b). Accounting for this difference is the observation that EcTesB also has strong specific activity on the precursor 3-hydroxyvaleryl-CoA that translates into 3-hydroxyvalerate production while Pr1687 does not have detectable activity on 3-hydroxyvaleryl-CoA. Further supporting the importance of reduced specific activity on pathway precursors is the observation that recombinant strains overproducing Pr655 produce much less valerate than strains overproducing EcYdiI and Pr1687 (Figure 2a, b) even though Pr655 maintains greater specific activity for valeryl-CoA than thioesterases EcYdiI and Pr1687 (Figure 3b). The increased specific activity of Pr655 on 3-hydroxyvaleryl-CoA and trans-2-pentenoyl-
CoA precursors relative to EcYdiI and Pr1687 suggests that the reduced valerate titer results from increased precursor acyl-CoAs hydrolysis.

Confounding our conclusion that thioesterases with low specific activity on precursor acyl-CoAs have improved final product titers is the observation that recombinant strains overexpressing ydiI produce high valerate titers with trans-2-pentenoate titers below the limit of detection despite EcYdiI exhibiting higher in vitro specific activity for 20 µM trans-2-pentenoyl-CoA than for 20 µM valeryl-CoA (Figures 2a, 3b). In our recombinant strains, EcYdiI and the other screened thioesterases are competing with the downstream pathway enzyme Ter for the substrate trans-2-pentenoyl-CoA. One plausible explanation for the absence of detectable trans-2-pentenoate from this recombinant strain is that EcYdiI has a $K_m$, a measure of substrate affinity, for both trans-2-pentenoyl-CoA and valeryl-CoA greater than the 20 µM concentration used in our in vitro assay, which is reasonable to assume because EcYdiI has evolved for specificity towards aromatic acids (34). If Ter has a lower $K_m$ for trans-2-pentenoyl-CoA than EcYdiI, then it could reduce the intracellular concentration of trans-2-pentenoyl-CoA below 20 µM, causing EcYdiI activity for this compound to become physiologically irrelevant. In this situation, the intracellular valeryl-CoA concentration could increase to a point where low-level specific thioesterase activity would result in significant valerate production. Pr655, on the other hand, could have a lower $K_m$ for both trans-2-pentenoyl-CoA and valeryl-CoA, which would compete more effectively for the trans-2-pentenoyl-CoA with downstream enzyme Ter and allow continued low level production of trans-2-pentenoate (Figure 2a). Determination of kinetic parameters for these non-cognate substrates is limited by the low concentrations of acyl-CoAs obtained from enzymatic synthesis and LC-based purification.
The thioesterase substrate preferences found by our in vitro experiments indicate the breadth of thioesterase functional diversity screened in this study and underline the importance of sampling in the selection of pathway enzymes (Figure 3a and b). The TesB enzymes of E. coli and P. putida represent the most phylogenetically (50% amino acid identity) and functionally (Figure 3a and b) similar enzymes, while the remaining four thioesterases appear both phylogenetically and functionally disparate. The TesB thioesterases showed both similar in vivo product profiles and in vitro specific activity profiles, which suggests that future screens aimed at acquiring more diverse TesB function should sample enzymes with lower similarity at the amino acid level than the E. coli and P. putida homologs. The diversity of the remaining 60 thioesterase phenotypes is an indication of the functional diversity inherent to bacterial thioesterases and suggests that many opportunities remain for isolating enzymes with improved specificity over those currently used in CoA-dependent biosynthetic pathways.

Our study also highlights the challenges associated with enzyme selection for metabolic pathways. A common route to selecting pathway enzymes is to rely on proposed function of known enzymes in databases and the literature or choose enzymes homologous to those with proposed functions. Unfortunately, enzyme annotation in public databases has degraded as functional analysis has not kept up with the rate of sequence deposition (35). The absence of credible studies on short-chain acyl-CoA thioesterases prompted us to implement a broad functional screen of diverse candidates. Our results demonstrate the power of a well-designed screen for isolating uncharacterized or poorly characterized enzymes that improve product titers. Our methods for screening and in vitro characterization of substrate specificity serve as a template for investigating other poorly characterized enzyme functions for pathway development. The E. coli thioesterase EcYdiI serves as an example of an enzyme whose
documented function may eliminate it from contention (34), but once included in our functional screen, maintains the appropriate levels of promiscuous activity in intracellular conditions to provide the highest specificity for unsaturated SCFAs discovered so far. It is unlikely that selecting enzymes based solely upon annotated function would have identified this level of activity because short-chain acyl-coA thioesterase activity remains poorly characterized.

During this study we isolated acyl-CoA thioesterases from diverse bacterial sources that increase production of saturated, unsaturated, and branched SCFAs through improvements in acyl-CoA substrate specificity. By comparing in vivo product profiles with in vitro specific activities of thioesterases that produced the greatest 3-hydroxyvalerate, trans-2-pentenoate, valerate, and 4-methylvalerate titers, we discovered that the most productive thioesterases found during functional screening: 1) use promiscuous activity to produce the SCFAs monitored in this study and 2) maintain low specific activity for pathway precursors relative to the preferred acyl-CoA. These findings indicate this study as a model for isolating enzymes for biochemical pathway functions that are poorly characterized in the literature. Further, the thioesterases we identified provide opportunities for increasing titers of desirable products as well as developing new pathways for the production of unsaturated SCFAs. We have also provided important in vivo and in vitro data on the production phenotypes and substrate specificities of poorly characterized acyl-CoA thioesterases for short-chain acyl-CoAs that are valuable for future bioprospecting and engineering studies.

Acknowledgements
We would like to thank the U.S. Department of Agriculture Agricultural Research Service for providing genomic DNA from cow rumen strains. This work was supported by the Institute for
Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.
References


Figures.

Figure 1. Biochemical pathway and LC-based *in vivo* screen for thioesterases with improved substrate specificity. The pathway in *E. coli* for the biosynthesis of SCFAs contains enzymes Pct (*Megaphaera elsdenii*), BktB (*Cupriavidus necator*), PhaB (*Cupriavidus necator*), PhaJ4 (*Cupriavidus necator*), and Ter (*Treponema denticola*) with varied thioesterases. The dashed bond indicates the additional carbon incorporated into the fatty acids with feeding of isobutyrate in place of propionate.
Figure 2. *In vivo* fatty acid production. Titers of fatty acids produced from (a,c) *E. coli* MG1655 (DE3) ΔendA ΔrecA and (b,d) *E. coli* MG1655 (DE3) ΔendA ΔrecA ΔtesB ΔyciA ΔyigI expressing CoA-based pathways for the biosynthesis of (a,b) valerate or 4-methylvalerate, (c,d) trans-2-pentenoate, with six different thioesterases. The product titers shown were obtained from three separate shake flasks for each thioesterase with error bars representing standard deviations. Each strain contained plasmids pCDF/phaB/phaJ4 with either (a,b) pET/ter/bktB/pct or (c,d) pET/bktB/pct. Each strain also contained plasmid pACYCduet-1 with the thioesterases displayed on the X-axis. Absence of product bars for a thioesterase indicates that titers were below the limit of detection for the HPLC RID detector.

*4-methylvalerate titers were determined using three separate cultures for each thioesterase fed 15 mM isobutyrate in the place of propionate.

Figure 3. Acyl-CoA specificity of six active thioesterases. (a) *In vitro* specific activity of six thioesterases on C2-C14 acyl-CoAs. Specific activities represent the average from three enzymatic reactions with error bars representing standard deviations. They were determined in the linear range for each thioesterase with acyl-CoA concentrations of 20 µM for C2, acetyl-CoA; C4, butyryl-CoA; C5, valeryl-CoA; C6, hexanoyl-CoA; C8, octanoyl-CoA; C10, decanoyl-CoA; C12, dodecanoyl-CoA; C14, tetradecanoyl-CoA. (b) *In vitro* specific activity of six thioesterases on saturated, valeryl-CoA; branched, 4-methylvaleryl-CoA; 3-hydroxy, 3-hydroxyvaleryl-CoA; unsaturated, trans-2-pentenoyl-CoA. Inset is enlarged activities of EcYdiI and Pr655 for comparison of low level activities. Specific activities were determined using the
same conditions described in (a). The absence of a bar for a given substrate/thioesterase specific activity indicates that the activity was below the limit of detection at absorbance of 412 nm.
Figure 1.
Table 1. Thioesterases screened for activity in a biosynthetic pathway for the production of short-chain fatty acids

<table>
<thead>
<tr>
<th>TesB &amp; TesB2 enzymes</th>
<th>E. coli MG1655</th>
<th>P. putida KT2440</th>
<th>P. ruminicola 23</th>
<th>R. albus</th>
<th>F. succinogenes S85</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli TesA</td>
<td>EcTesA</td>
<td>Pr244</td>
<td>Pr655</td>
<td>Ral865</td>
<td>Fsu266</td>
</tr>
<tr>
<td>P. putida TesB</td>
<td>EcPaaI</td>
<td>Pr254</td>
<td>Pr1075</td>
<td>Ral874</td>
<td>Fsu270</td>
</tr>
<tr>
<td>R. opacus TesB</td>
<td>EcEntH</td>
<td>Pr262</td>
<td>Pr1498</td>
<td>Ral880</td>
<td>Fsu368</td>
</tr>
<tr>
<td>P. syringae TesB2</td>
<td>EcFadM</td>
<td>Pr301</td>
<td>Pr1510</td>
<td>Ral1843</td>
<td>Fsu803</td>
</tr>
<tr>
<td>A. borkumensis TesB2</td>
<td>EcYbfF</td>
<td>Pr580</td>
<td>Pr1668</td>
<td>Ral1929</td>
<td>Fsu944</td>
</tr>
<tr>
<td>P. aeruginosa PAO1 TesB2</td>
<td>EcYbgC</td>
<td>Pr1218</td>
<td>Pr1687</td>
<td>Ral1938</td>
<td>Fsu1747</td>
</tr>
<tr>
<td></td>
<td>EcYbhC</td>
<td>Pr1466</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td>EcYciA</td>
<td>Pr1980</td>
<td>Pr1687</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td>EcYiiD</td>
<td>Pr2050</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td>EcYigI</td>
<td>Pr2308</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td>EcYdfI</td>
<td>Pr2318</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr3281</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr3807</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr4105</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr4180</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr4181</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr4975</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr5198</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr5331</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pr5356</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PrPhaJl</td>
<td>Pr2385</td>
<td>Ral2059</td>
<td>Fsu2108</td>
</tr>
</tbody>
</table>

\(^{a}(11)\)
\(^{b}(30)\)
Table 2. Maximum fatty acid titers observed for each product of the CoA-dependent pathway

<table>
<thead>
<tr>
<th>Strains</th>
<th>3-Hydroxyvalerate $^a$</th>
<th>Trans-2-pentenoate $^b$</th>
<th>Valerate $^c$</th>
<th>4-Methylvalerate $^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655 ΔendA ΔrecA</td>
<td>2.163±0.078</td>
<td>0.695±0.002</td>
<td>1.425±0.011</td>
<td>0.428±0.019</td>
</tr>
<tr>
<td>MG1655 ΔendA ΔrecA ΔtesB ΔyciA ΔyigI</td>
<td>1.940±0.099</td>
<td>0.609±0.015</td>
<td>1.540±0.034</td>
<td>0.570±0.041</td>
</tr>
</tbody>
</table>

$^a$ Thioesterase EcTesB
$^b$ Thioesterase EcYdiI
$^c$ Thioesterase Fs2108