Metabolic Engineering 77 (2023) 21–31

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Evaluation of strategies to narrow the product chain-length distribution of microbially synthesized free fatty acids

Michael A. Jindra [a,e], Kisurb Choe [b,e], Ratul Chowdhury [c], Ryan Kong [a], Soodabeh Ghaffari [d], Jonathan V. Sweedler [b,e], Brian F. Pfleger [a,e,*]

[a] Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, USA
[b] Department of Chemistry, University of Illinois, Urbana, IL, USA
[c] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA, USA
[d] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
[e] DOE Center Advanced Bioenergy and Bioproducts Innovation, USA

ARTICLE INFO

Keywords:
Acyl-ACP thioesterase
Medium-chain fatty acids
Enzyme selectivity
Fatty acid biosynthesis
High-throughput enzyme assay
MALDI-MS

ABSTRACT

The dominant strategy for tailoring the chain-length distribution of free fatty acids (FFA) synthesized by heterologous hosts is expression of a selective acyl-acyl carrier protein (ACP) thioesterase. However, few of these enzymes can generate a precise (greater than 90% of a desired chain-length) product distribution when expressed in a microbial or plant host. The presence of alternative chain-lengths can complicate purification in situations where blends of fatty acids are not desired. We report the assessment of several strategies for improving the dodecanoyl-ACP thioesterase from the California bay laurel to exhibit more selective production of medium-chain free fatty acids to near exclusivity. We demonstrated that matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS) was an effective library screening technique for identification of thioesterase variants with favorable shifts in chain-length specificity. This strategy proved to be a more effective screening technique than several rational approaches discussed herein. With this data, we isolated four thioesterase variants which exhibited a more selective FFA distribution over wildtype when expressed in the fatty acid accumulating *E. coli* strain, RL08. We then combined mutations from the MALDI isolates to generate BTE-MMD19, a thioesterase variant capable of producing free fatty acids consisting of 90% of C_{12} products. Of the four mutations which conferred a specificity shift, we noted that three affected the shape of the binding pocket, while one occurred on the positively charged acyl carrier protein landing pad. Finally, we fused the maltose binding protein (MBP) from *E. coli* to the N – terminus of BTE-MMD19 to improve enzyme solubility and achieve a titer of 1.9 g per L of twelve-carbon fatty acids in a shake flask.

Author contributions

Michael Jindra: Conceptualization, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Validation, Formal analysis, Project administration, Kisurb Choe: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing, Ratul Chowdhury: Conceptualization, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Ryan Kong: Investigation, Data curation, Visualization, Validation, Soodabeh Ghaffari: Investigation, Methodology, Validation, Jonathan Sweedler: Funding Acquisition, Supervision, Writing – Original Draft, Writing – Review & Editing, Brian Pfleger: Funding Acquisition, Supervision, Writing – Original Draft, Writing – Review & Editing.

1. Introduction

Oleochemicals service a broad array of markets and applications, yet the limited availability of the eight-to-twelve-carbon products in this chemical class motivate the pursuit of alternatives to the conventional routes of procurement (Sarria et al., 2017). Medium-chain fatty acids, particularly dodecanoic acid, are predominantly obtained from palm kernel oil and confer desirable properties in cosmetic and surfactant applications (Parsons et al., 2020). While the refining process is a mature technology, palm kernel oil represents about 8% of the mass of the oil in

* Corresponding author. 1415 Engineering Drive Madison, WI, 53706, USA.
E-mail address: brian.pfleger@wisc.edu (B.F. Pfleger).

https://doi.org/10.1016/j.ymben.2023.02.012
Received 3 November 2022; Received in revised form 29 January 2023; Accepted 28 February 2023
Available online 1 March 2023

the fruit, and the kernel oil is composed of only about 50% dodecanoic acid (Rupilius and Ahmad, 2007), (Amira et al., 2014). Production of the longer chain fractions by palm oil refiners ensures that enough palm kernel oil is harvested to satisfy most of the global demand for dodecanoic acid. However, if the palm oil producing countries experience a shortage in supply, the resulting strain could leave the medium-chain oleochemical space underserved. Furthermore, the conversion of tropical ecosystems to palm oil plantations remains a significant impediment to maintaining biodiversity and controlling greenhouse gas emissions in southeastern Asia (Parsons et al., 2020), (Wilcove and Pin, 2010).

In order to enter the oleochemical market as an improvement over the incumbent technology, microbial platforms must demonstrate selective production of medium-chain oleochemicals, industrial-scale product titers with a small land-use requirement, and utilization of low-cost, low-footprint feedstocks (Pfleger et al., 2015), (Parsons et al., 2018). This work focuses on developing the first value proposition: selectivity. Expressing an engineered or natively selective acyl-ACP thioesterase has served as the predominant strategy for redirecting flux of growing acyl-ACPs from the fatty acid biosynthesis cycle to accumulate medium-chain FFA in oleaginous bacteria and plants (Fig. 1) (Grisewood et al., 2017), (Hernández Lozada et al., 2018). The selectivity imparted to the FFA pool has been extended to fatty acid derivatives such as medium-chain fatty alcohols, methyl ketones, and polyhydroxylalkanoates (Youngquist et al., 2013), (Hernández Lozada et al., 2020), (Agnew et al., 2012), (Yan et al., 2021). Thus, highly selective thioesterases which can deliver a precise product profile, particularly the medium-chain length building blocks, are necessary to the development of viable oleochemical bioprocesses.

Mutagenesis efforts have succeeded in tuning free fatty acid product profiles and identifying the key features driving substrate specificity and activity of the acyl-ACP thioesterase (Grisewood et al., 2017), (Sarria et al., 2018), (Jing et al., 2018), (Yuan et al., 1995). Hernández et al. demonstrated the engineering of a highly active and nearly exclusive octanoyl-ACP thioesterase by leveraging a growth selection which facilitated identification of high C_8-producing variants (Hernández Lozada et al., 2018). However, this highly tailored product distribution, 90% of a single chain length in the product pool, remains elusive for C_{10} and C_{12} thioesterases (Table 1). We approached this engineering need by enhancing the selectivity of the well-studied BTE (a.k.a. ^{Uc}FatB1), an acyl-ACP thioesterase from the California bay laurel with high native specificity for dodecanoyl-ACP substrates (Voelker et al., 1992). To meet our objective, we first used the BTE crystal structure to design a library of variants *in silico* using the Iterative Protein Redesign and Optimization (IPRO) tool (Saraf et al., 2006), (Pantazes et al., 2015). This approach was previously effective in enhancing the substrate specificity of 'TesA, a truncated variant of the native acyl-ACP thioesterase from *E. coli*, resulting in several variants with about 50% medium-chain selectivity in their product pools (Grisewood et al., 2017). We hypothesized that the availability of the high-resolution structural data, the native specificity of BTE for C_{12} substrates, and the previous success with the IPRO algorithm would facilitate the creation of a thioesterase with 90% selectivity for dodecanoyl-ACP. Ultimately, we found that mutagenesis campaigns guided by structure and sequence-structure data were unable to produce any variants with enhanced specificity. We consulted previous work which resulted in the engineering of a highly active and highly selective octanoyl-ACP thioesterase, which combined a randomly mutagenized library of CupTE from *Cuphea palustris* (CupTE) with a growth selection (Hernández Lozada et al., 2018). Since we did not have a growth selection for our product of interest, we sought a high-throughput screening method capable of processing a randomly mutagenized library of the BTE gene. We pivoted to a structure agnostic approach which leveraged colony screening capabilities of a matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF MS). This technology enabled screening of *E. coli* colonies transformed with a randomly mutagenized BTE library for shifts in FFA product distribution compared to colonies expressing the wildtype. The novel acyl-ACP thioesterase design-build-test cycle introduced herein led to the discovery of several residue changes which were productive in meeting our engineering target as well as reinforcing the fundamental understanding of the primary modes of selectivity in this enzyme class.

Table 1
The medium-chain acyl-ACP thioesterase selecitivity data for each product chain-length when expressed in E. coli. The columns represent the origin of the gene, selectivity enhancing mutations, if any, the selectivity for the main product in the FFA pool, and the titer of the main FFA product. Experimental results were compared across trials performed in LB with 4 g per L of glycerol. References under the first column indicate the first study to express the WT gene in E. coli, and references under the second column indicate the study which engineered the more selective variant.

Summary of most selective medium-chain acyl-ACP thioesterases for each product chain-length when expressed in *E. coli*

	Gene Source	Mutations	Selectivity	Titer (mg/L)
C_8	*Cuphea palustris* (Jing et al., 2011)	Truncation to M19, N28S, I65M (Hernández Lozada et al., 2018)	90%	470
C_{10}	*Cuphea lanceolata* (Banerjee et al., 2022)	Truncation to M25, D34S (Banerjee et al., 2022)	70%	390
C_{12}	*Umbellularia californica* (Voelker and Davies, 1994)	WT	85%	560

2. Materials and methods

2.1. Media and molecular biology materials

Media components were ordered from IBI Scientific (Dubuque, IA) and Fisher Scientific (Waltham, MA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Enzymes for DNA cloning were purchased from New England Biolabs (Ipswich, MA).

2.2. DNA cloning methods

Various cloning techniques were employed to incorporate mutations into BTE expression vectors. All point mutations and site saturation

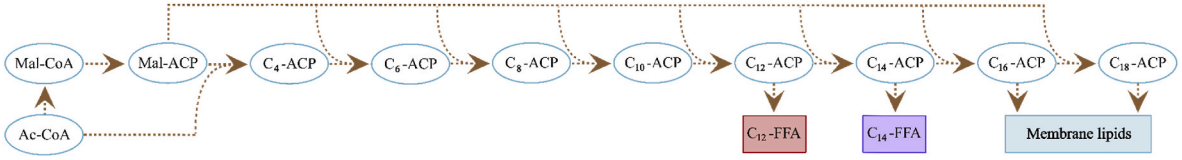


Fig. 1. Schematic of the effects of expression of the acyl-ACP thioesterase BTE in E. coli. Carbon sources (not shown) are catabolized into acetyl-CoA (Ac-CoA), the primary metabolite responsible for fueling the elongation process in the fatty acid biosynthesis cycle. Acyl-ACP substrates otherwise destined for cell membrane components are intercepted from the fatty acid biosynthesis cycle and rerouted to the fatty acid pool. The WT BTE enzyme has activity on C_{12} and C_{14} substrates, leading to the accumulation of dodecanoic, dodecenoic, tetradecanoic, and tetradecenoic acid in the cytosol. Malonyl-CoA and malonyl-ACP are abbreviated as Mal-CoA and Mal-ACP.

mutations were incorporated using Gibson cloning techniques described previously (Banerjee et al., 2022). The randomly mutagenized library screened by MALDI-MS analysis was generated by amplifying the entire *BTE* gene with Genemorph error-prone polymerase from Agilent (Santa Clara, CA), adjusting the template DNA concentration and the number of PCR cycles to target 2–4 mutations per kilobase. Primers for mutagenesis contained cut sites for *Xba*I and *Hin*dIII such that the mutagenized DNA could be subsequently digested and ligated into the pTrc99a vector backbone digested with the same enzymes (Amann et al., 1988). After Sanger sequencing of a sample population, we measured a mutational frequency of 4.8 per kilobase was incorporated into the library.

The library of alignment mutants (denoted as AM) was constructed via several batches of nickase mutagenesis described by Wrenbeck et al. (2016a). The cut site for the *Bbv*CI nickase was cloned into the pTrc99a backbone to facilitate this technique for mutagenizing the *BTE* gene.

2.3. In vivo characterization of acyl-ACP BTE variants

Each BTE construct was transformed into RL08*ara*, an *E. coli* MG1655 derivative engineered for free fatty acid accumulation (see Table S1 for genotype) (Lennen et al., 2010). Single colonies of the transformants were grown overnight at 37 °C and 250 r.p.m. in 5 mL of LB media supplemented with 100 mg per L of carbenicillin. Production flask experiments commenced with the inoculation of 25 mL of media with 275 μL of the overnight culture. The cultures were induced via addition of IPTG to a final concentration of 20 μM once OD reached about 0.2 and continued to incubate for 24 h. Fatty acids were extracted from 2.5 mL of bulk cell culture, derivatized into fatty acid methyl esters, and quantified as previously described (Banerjee et al., 2022).

The IPRO, AM, MM, and MMD libraries were all screened using standard LB media supplemented with 4 g per L of glycerol. The rich media used for experiments in Fig. 6 was an undefined buffered media supplemented with 20 g per L of glycerol and 2.5 mM calcium pantothenate, as prepared by Clomburg et al. (2012). The C_{12} to C_{14} ratio was calculated by dividing the titer of dodecanoic and dodecenoic acid by the titer of tetradecanoic and tetradecenoic acid. The percent of C_{12} products in the FFA distribution was calculated by dividing the titer of dodecanoic and dodecanoic acid by the titer of all other species, excluding membrane components (C_{16} and C_{18} products). The summary of these results in Table S3, however, show the percent of each individual FAME in the overall product distribution, with C_{16} and C_{18} products included in the denominator.

2.4. Nile red analysis for activity of BTE variants

The activity of BTE variants from the site saturation and the alignment-informed mutagenesis libraries was prescreened using the Nile red fluorescence assay (Greenspan et al., 1985). A 96-well plate containing 140 μL of LB and carbenicillin (100 mg per L) was inoculated with colonies of RL08 transformants. The preculture plate was incubated overnight at 37 °C and 250 r.p.m. 25 μL of the preculture was inoculated into 500 μL of LB supplemented with 4 g per L of glycerol, carbenicillin, and 20 μM IPTG in a deep-well plate. After 24 h, the deep well plate was spun for 10 min at 4000 *x* G and 200 μL of supernatant was added to a black-walled 96 well assay plate (Corning, NY) and mixed with 2 μL of 1 mM Nile red dye dissolved in DMSO and incubated in the dark for 15 min. Absorbance and fluorescence were measured in a Tecan M1000 plate reader (Männedorf, Switzerland) at 600 nm and with an excitation at 550 nm and emission at 630 nm, respectively. Members of the AM library were tested in biological triplicate. For the site saturation mutagenesis libraries, 48 single isolates of each library were tested. Variants were characterized as active if the FU measurement was within the range of the measurements recorded for the WT. Identified active variants were fully characterized for their effect on FFA distribution.

2.5. In vitro characterization of BTE kinetic parameters

The expression constructs for WT BTE, BTE MMD19, and *E. coli* ACP were designed with an N-terminal 6x-histidine tag for purification by immobilized metal affinity chromatography. Constructs encoding BTE, Sfp from *Bacillus subtilis*, and AasS from *Vibrio harveyi* were transformed into BL21 (DE3) and the construct the *E. coli* acyl carrier protein was transformed into BAP1 (to ensure full phosphopantetheinylation) (Pfeifer et al., 2001). The BL21 (DE3) cells were cultured in 500 mL of LB media. Culturing, lysis, and purification conditions for the thioesterases were identical to those previously described in Hernández-Lozada et al. (Hernández Lozada et al., 2018). BAP1 cells for overproduction of ACP were cultured in 500 mL of Terrific broth and induced at 250 μM IPTG. Concentration of the BTE proteins was quantified using an extinction coefficient of 46,200 $M^{-1}cm^{-1}$ at 280 nm absorption. Sfp concentration and AasS concentration was determined using extinction coefficients of 30,620 $M^{-1}cm^{-1}$ and 67,520 $M^{-1}cm^{-1}$, respectively.

Prior to separation via chromatography, the lysis suspension was centrifuged 3 times for clarification of the supernatant as filtration was not feasible. The lysate was loaded onto a HisTrap™ HP 5 mL chromatography column (Marlborough, MA) and eluted with buffer containing 20 mM Na_2HPO_4, 500 mM NaCl, and 500 mM imidazole and was adjusted to a pH of 8. After collection of the relevant fractions from the eluant, the chromatography was repeated with the column flowthrough to recover as much ACP as possible. The ACP was then buffer exchanged into a protease buffer solution containing 20 mM Tris-HCl and 50 mM NaCl with a pH of 7.5. Commercial His-tagged TEV protease from Genscript (Piscataway, NJ) was added and the cleavage reaction mixture was dialyzed in protease buffer solution overnight at 20 °C with gentle stirring. The TEV protease was removed via column chromatography and the cleaved apo/holo-ACP product was collected in the flow-through. The apo/holo ACP was purified and buffer exchanged into a buffer containing 50 mM Na_2HPO_4 and 10% glycerol (w/v) with a pH of 8. The derivatization to dodecanoyl-ACP and tetradecanoyl-ACP was accomplished with the Sfp and the AasS proteins according to the methods described previously. The derivatization to tetradecanoyl-ACP required 48 h to reach full conversion, as confirmed by HPLC analysis. Addition of an equimolar ratio of DTNB and buffer exhange with a PD-10 desalting column was crucial prior to performing the kinetic assay. The concentration of acyl-ACP was determined via BCA assay (Thermo Fisher Scientific).

The *in vitro* analysis was carried out by mixing dodecanoyl-ACP or tetradecanoyl-ACP in concentrations ranging from 0 to 350 μM with 40 nM of thioesterase, 250 μM DTNB, 100 mM Na_2HPO_4 in a 100 μL reaction volume. The absorbance at 412 nm was monitored every 5 s for 2 min on a NanoDrop 2000c (Thermo Fisher Scientific). All concentrations except 350 μM were tested in triplicate.

2.6. Colony screening of randomly mutagenized library with MALDI-MS

Overnight culture of *E. coli* holding the mutant library was diluted 1E6 in LB, and 500 μL of the diluted culture was spread onto an LB plate with 100 mg per L carbenicillin, 10 g per L glucose, and 0.1 mM IPTG. After overnight culture at 37 °C in an incubator, membrane filter paper (EMD Millipore, Burlington, MA) was laid on the colonies on the Petri dish. The colonies transferred onto the paper were imprinted onto to a 105 mm × 75 mm MALDI target plate by applying pressure. The sample plate with the colonies was imaged by a flatbed scanner (EPSON V300 Perfection, Nagano, Japan) at 800 dpi, and a custom-made spraying device was used to apply MALDI solution (PNA dissolved in acetonitrile at 5 mg per mL) onto the MALDI target plate. Coordinates of the colonies and custom geometry file were obtained from image analysis by the macroMS tool (Choe et al., 2021). For MALDI-MS screening of colonies, an ultrafleXtreme mass spectrometer (Bruker Corporation, Billerica, MA) was used. From each colony, 4000 laser shots were collected at 1000 Hz and the laser size was set to "Ultra"(~70 μm footprint) at 50%

power in negative ion reflectron mode. The operating parameters were 60 ns for the pulsed ion extraction time, 20 kV for the accelerating voltage, 18.75 kV for the extraction voltage, 5 kV for the lens voltage, and 26.4 kV for the reflector voltage. Random walk was used with 50 shots per raster point. The acquired data, the image file, and the text file listing the *m/z* ranges of the peaks for C_{12}, C_{14}, and C_{16} fatty acids and membrane lipids, including unsaturated species, were uploaded to macroMS generating an Excel file listing ion counts for these peaks for all colonies. Colonies showing the top 10% peak signals for C_{12} FFA were selected, and C_{14} FFA resulting from fragmentation of the membrane lipid was removed by subtracting 6–10% of the C_{16} product signals from the original C_{14} FFA signal. The ratio between C_{12} FFA and C_{14} FFA was then calculated, and the colonies showing changes in the ratio were isolated from the original LB plate using the macroMS image viewer tool.

2.7. BTE model construction and structure-based redesign and analysis

Experimentally determined atomic coordinates of the 12:0-ACP thioesterase from *Umbellularia californica* (BTE) (PDB ID: 5X04) were used as the structural prior for structure-based redesign (Feng et al., 2017). To assess and address the molecular mechanism behind the active variants, we first predicted their corresponding structure using RGN2 and subsequently performed flexible substrate (and cofactor) docking at the active site (using a combination of AutoDock Vina and AutoLigand) and force-field energy minimization (using Amberff19SB – for proteins, and Ambergaff2 – for ligands) (Harris et al., 2008), (Trott and Olson, 2010), (He et al., 2020), (Tian et al., 2020). We then constructed a Catalytic Contact Cascade (C^3) map to explain the biophysical underpinnings of the active variants, as shown in Fig. 6. This map helped us discern the catalytic effects of point mutations to covalently and non-covalently linked residues far away from the substrate binding pocket where effects of molecular forces (van der Waals and electrostatics) are insignificant (Hernández Lozada et al., 2018).

3. Results and discussion

3.1. Computationally guided library generation

Our engineering goal was to deliver a bacteria and enzyme system capable of producing a FFA pool comprised of at least 90% C_{12} FFA products by engineering the BTE enzyme. We began a rationally guided mutagenesis campaign of the BTE enzyme by using the IPRO algorithm to redesign the binding pocket to exclude non-C12-chain length acyl-ACP substrates (Saraf et al., 2006). The availability of the BTE crystal structure enabled the selection of 14 residues as design positions for *in silico* mutagenesis (Feng et al., 2017). Twenty-four of the variants calculated to have a more favorable interaction energy for medium-chain substrates were cloned into the pTrc99a vector and subsequently tested *in vivo* in the production host, RL08*ara*. Of these 24 variants, only 2 retained activity, albeit with less activity than wildtype (Fig. 2). No variants were discovered which improved the specificity for dodecanoyl-ACP *in vivo*, indicated by the distribution of FFA being similar to the distribution observed from the WT. The dataset revealed that some of the selected design positions may have been crucial to the activity or stability of the BTE enzyme. Namely, any variant made with a substitution at N288 yielded an inactive thioesterase.

To evaluate our selection of design positions and the mutability of the BTE enzyme, we generated a site saturation mutagenesis library of 5 of the 14 design positions: R168, T169, H170, M197, and N288. After screening the site saturation mutagenesis libraries with a Nile red assay, we observed that some of the design positions were more tolerant of mutations than others (Fig. S1). For example, we observed the H170*X* library could accommodate aromatic, aliphatic, positive, negative, and

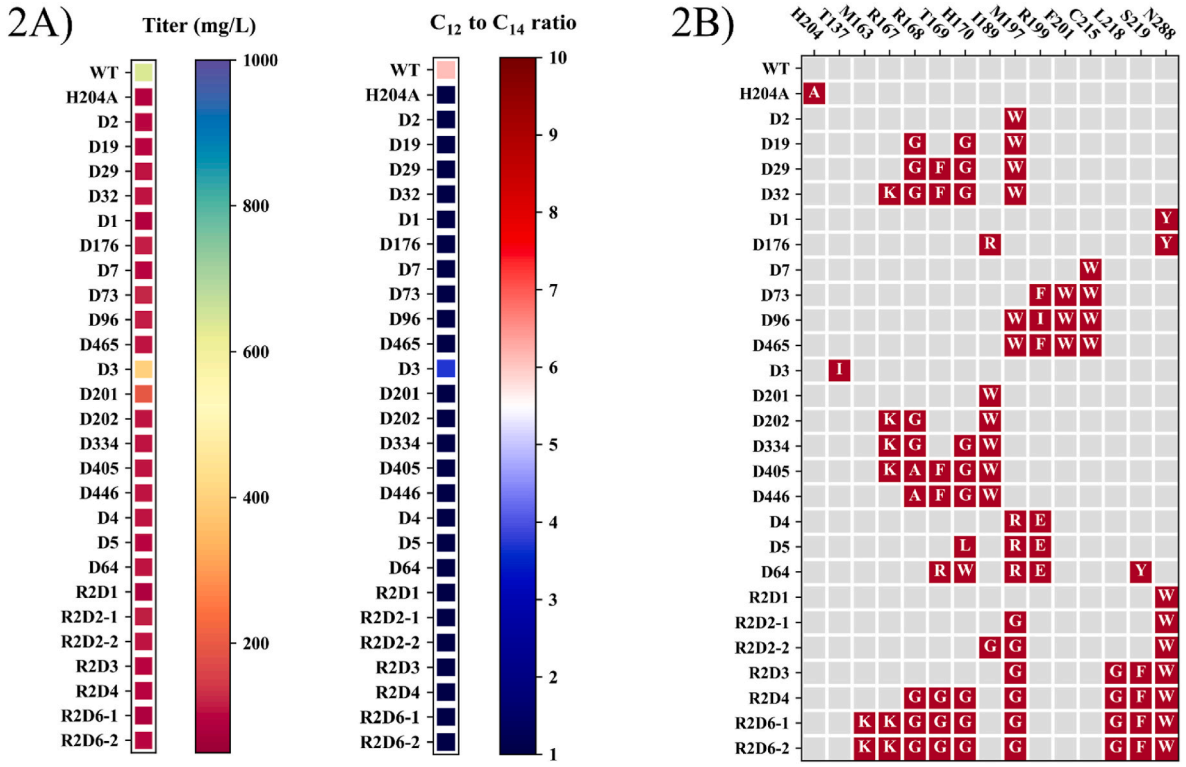


Fig. 2. (A) Heatmaps showing *in vivo* results of IPRO-informed BTE variants, with total FFA titer on the left and the C_{12} to C_{14} ratio within the product distribution on the right. Titers represent methyl ester derivatives of bulk cell cultures, including FFA and membrane lipids. Of 25 variants tested, 2 retained activity and none were more specific than the WT for dodecanoyl-ACP. WT and BTE-H204A are included as positive and negative controls, respectively. (B) Grid of mutations incorporated into IPRO library. Red squares indicate a substitution along with the new residue incorporated at its respective design position. Gray squares indicate no change in the given variant at the design position.

uncharged polar substitutions at that position. In contrast, only isolates which expressed the WT sequence were active in the N288*X* library. This experiment elucidated why many of the IPRO variants were inactive; with the intent to occlude longer chain substrates from the binding pocket, aromatic amino acids were substituted at position N288 in 9 of the redesigns. While this led to more favorable binding energies in the IPRO calculation, these aromatic amino acids likely destabilized the protein or ablated catalytic activity. The results from the Nile red study suggested that BTE was mutable, however, a more judicious selection of design positions would be necessary to ensure preservation of activity and increased likelihood of discovering variants with enhanced medium-chain specificity. However, we resolved to proceed with BTE as our mutagenesis template because of its natively high preference for lauroyl-ACPs and the potential for meeting our design goal of exclusive specificity. We next compared elements of our mutagenesis approach to other successful plant acyl-ACP thioesterase engineering campaigns.

A common factor among successful plant acyl-ACP TE engineering studies was the use of a sequence alignment among closely related homologs (Jing et al., 2018), (Yuan et al., 1995), (Banerjee et al., 2022). Most notably, BTE was successfully engineered to possess exclusive C_{14} substrate preference with 3 amino acid substitutions. These modifications were informed by sequence alignment to a myristoyl-ACP TE homolog from the Lauraceae family (Yuan et al., 1995). The authors also reported nonadditive combinatorial effects, as one of the mutations did not confer a specificity shift on its own but resulted in a product specificity shift from ~60% to ~90% when combined with the double mutant, BTE M197R/R199H. We performed a multiple sequence alignment of 7 well-characterized plant acyl-ACP TE to the BTE amino acid sequence. Interestingly, the alignment revealed that N288 was conserved across all homologs, suggesting that this design position may be crucial for stability of the plant acyl-ACP thioesterase. In summary, while IPRO was able to identify amino acid changes that resulted in more favorable interaction energies with medium-chain substrates, the design algorithm was not able to detect destabilizing mutations *a priori*.

We thus reasoned that we required a different strategy that maximized a chance of finding active variants, allowed exploration of epistatic effects of multiple mutations, and maintained a feasible design space for our testing throughput. Subsequently, we adjusted our design process to be guided by a multiple sequence alignment of various acyl-ACP thioesterases from the plant kingdom.

3.2. Multiple-sequence-alignment guided mutagenesis

To execute our multiple sequence alignment mutagenesis strategy, we curated a list of 4 pairs of acyl-ACP thioesterases from the *Lauraceae* and *Lythraceae* families. Each pair consisted of two thioesterases from the same plant species: one with long-chain specificity and one with medium-chain specificity. We analyzed the multiple sequence alignment with the following criteria. 1) Does the residue at the same position differ between the medium-chain and the long-chain homolog from the same host? 2) Does BTE possess the residue of the long-chain homolog at that position? If both criteria were met, then the position was noted and added to a list of candidate mutations for the build and test phase. The results of this alignment analysis are shown in Table S4. A number of residues fit the above criteria, however, only 9 residues fit this criteria that were located within the binding pocket secondary structures (Feng et al., 2017). We designed the alignment-informed library, termed AM, to contain variants with multiple substitutions while still maintaining a manageable library size. Thus, we prioritized the 8 binding pocket residues that emerged from this analysis and included 4 other residues that were near the binding pocket in the generation of the AM library (Table S5). T137A, termed AM0, was cloned and tested as a proof-of-concept since all homologs which preferred long-chain substrates possessed an alanine at the 137th position. Variant AM0 exhibited an increase in tetradecanoic acid production and a decrease in dodecanoic acid production. This was corroborated by a similar shift in product profile observed by Feng et al. upon testing the effects of a glycine substitution at the same residue (Feng et al., 2017). We then proceeded to construct the AM library with a nickase-assisted mutagenesis technique and screened the resulting variants using the Nile red assay (Wrenbeck et al., 2016b). Positive hits from the Nile red screen were subsequently tested in production cultures for characterization of the effect on FFA distribution (Fig. 3). This strategy was effective in increasing the proportion of active BTE variants in the library, as over 5 out of the 14 members of the library retained activity. However, we were unsuccessful in identifying variants with improved C_{12} specificity. At this point, we concluded the rational mutagenesis campaigns were not productive towards our engineering goal of achieving near-exclusive selectivity for medium-chain substrates. The rapid advance of GPU-accelerated protein structure determination has revolutionized the scope of structural biology and it is only now that, within the same study we can see how physics-based tools like IPRO (and Rosetta) need to be replaced with AI-guided variant predictors like RGN2 for sequence design. IPRO or the multiple sequence alignment, in principle, might have led to a variant with the desired performance, but neither strategy was able to narrow our library to a size commensurate with the throughput of the testing methods. We sought a more rapid testing pipeline to meet our goal.

3.3. MALDI-guided screening of randomly mutagenized library

We sought a technique which would allow us to screen for changes in fatty acid distribution and could accommodate rationally designed or randomly generated libraries. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectroscopy emerged as a promising option, as several demonstrations have achieved rapid detection of oleochemicals and lipids from biological samples (Wang et al., 2015). More recently, MALDI-ToF MS was shown to facilitate the detection of *Saccharomyces cerevisiae* colonies with medium-chain fatty acid production capability at a rate of about 2 s per sample (Xue et al., 2020). Upon demonstrating successful MALDI detection of dodecanoic acid from RL08 colonies expressing the BTE enzyme, we prepared a randomly mutagenized library of the entire coding sequence and screened for changes in the C_{12} to C_{14} ratio (Fig. 4A). The screen presented 18 candidate variants after analysis of 3236 colonies. Of these variants, 4 displayed an increase in the C_{12} FFA proportion observed in the FFA distribution and all maintained activity when tested for *in vivo* performance in liquid cultures (Fig. 4B).

A final library, termed the MALDI mutant designs (MMD), was created based on the residues which conferred the desired change in FFA production observed in MM19, MM21, MM75, and MM375 (Fig. S2). Namely, the substition at R109C and positions M133, M222, and R227 were tested individually, in combination, and with various substitutions to further probe the response of the FFA distribution to changes in these specific locations. We found that variant MMD19, which included M133L, M222K, and R227H, exhibited a C_{12} to C_{14} ratio of 11.4 and met the target of producing 90% C_{12} in the FFA distribution (Fig. 5). Variants with the R109C substitution (MMD29 and MMD30) were more selective than WT, however, these variants were less active and less selective than the MMD19 variant.

3.4. Structural analysis of MMD19 variant

Despite dramatic, recent increases in prediction accuracy in protein structure prediction using tools such as AlphaFold2 and RosettaFold, the predicted structures bear strong affiliation with the multiple sequence alignment (MSA) of the target protein (Ronneberger et al., 2021), (Baek et al., 2021). Thus, for protein design endeavors such as this, where a point mutation could drastically alter protein function with only marginal changes to the sequence alignment, the success of AlphaFold2 would depend on prior exposure to variants of our target protein in its training set. Therefore, we used our latest end-to-end differentiable

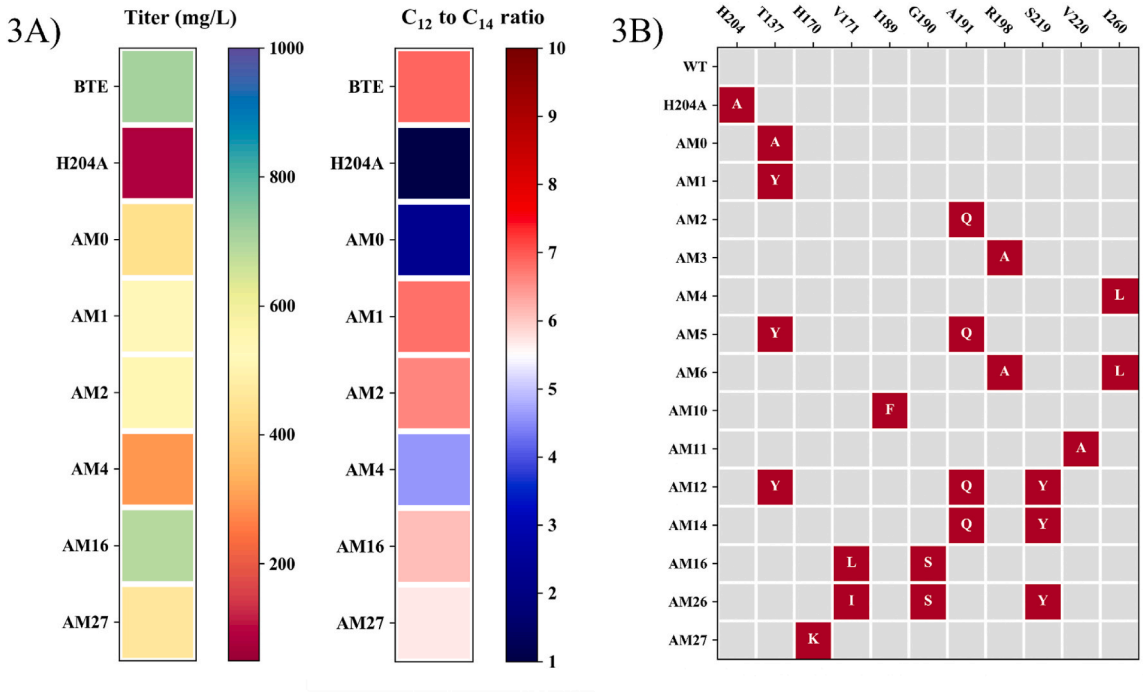


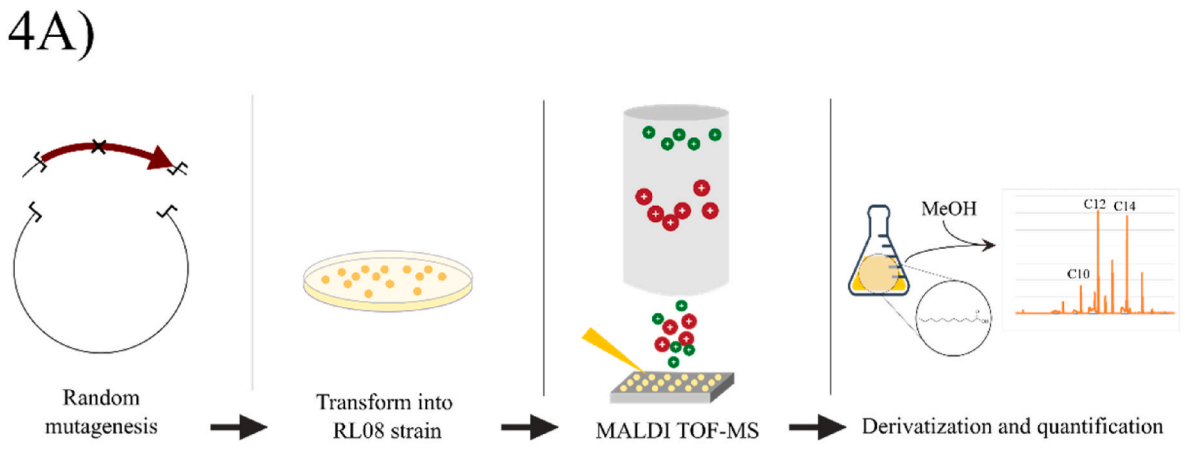
Fig. 3. (A) Heatmaps showing *in vivo* results of all active BTE variants informed by multiple sequence alignment, with total FFA titer on the left and the C_{12} to C_{14} ratio within the product distribution on the right. Titers represent methyl ester derivatives of bulk cell cultures, including FFA and membrane lipids. Inactive variants were not tested due to prescreening with the Nile red assay. (B) Grid of mutations incorporated into the AM library, including inactive variants. Red squares indicate a substitution along with the new residue incorporated at its respective position. Gray squares indicate no change in the given variant at the design position.

recurrent geometric network (RGN2) instead, which can predict protein structure from single protein sequences without use of MSAs (Chowdhury et al., 2022). This deep learning protein language model encodes latent structural information from millions of unaligned proteins and represents Cα backbone geometry and outperforms AlphaFold2 on single-sequence prediction tasks such as non-homologous proteins (i.e., no MSA present) (Michaud et al., 2022). Thus, RGN2 was used to analyze the structure of WT BTE and the MMD19 variant.

The BTE structure follows the canonical HotDog fold pattern typical of plant acyl-ACP thioesterases: an α helix surrounded by antiparallel β sheets (Cantu et al., 2010). The three mutations in the triple mutant, even though apparently disconnected, work in concert as they are connected by a 16-residue fragment of the β5 sheet in the HotDog fold and a neighboring turn i.e., contact map (Fig. 6). The contact map consists of 12 covalent bonds via the β5 sheet of the HotDog fold and one sulfide interaction. Two of these three residues (R227 and M222) are responsible for ACP recognition and acyl-chain entry at the pocket entrance by electrostatic interaction with the D56 of the ACP moiety. The third position, M133, is at the base of the pocket which controls substrate length (i.e., proximity to the omega carbon of the substrate). This third residue, M133, is what drives the major selection of medium-chain (C_{12}) vs. long-chain (C_{14}) selectivity. At a molecular level amino acid substitution M133L sterically pushes the acyl part of tetradecanoyl-ACP out of the pocket so much so that the recognition and stabilization residues are not able to maintain binding with D56 of the substrate. The mutations to the recognition residues (R227H and M222K) only marginally contribute to binding and substrate specificity. M222K and R227H in fact cumulatively weaken the ACP recognition by 15% (as seen from individual contribution of thioesterase-ACP residue-residue interaction). This is a necessary evil, as it permits M133L mutation to force tetradecanoyl-ACP out of the pocket with lesser resistance from entrance binding. For dodecanoyl-ACP, the entrance binding is equally weakened, but it is stabilized by hydrophobic packing and in a clash-free manner inside the pocket. Thus, on a relative basis, the binding of C_{12} substrates is stronger than for C_{14} substrates for the triple mutant (i.e., high specificity), and each of these three mutations are tolerated as single point mutants (as seen from experimental screens for mutant purification and expression).

3.5. *In vitro characterization of kinetic parameters*

The WT and the MMD19 variant were both tested *in vitro* to complement the results shown *in vivo*. Dodecanoyl-ACP, tetradecanoyl-ACP, and the BTE WT and MMD19 variant were purified to measure the kinetics of thioester hydrolysis, and the initial rate data was fit to the classic Michaelis-Menten model to obtain the catalytic parameters, k_{cat} and K_M, for each thioesterase-substrate pair. The Michaelis-Menten curves for these four reaction conditions are shown in Fig. S3. While MMD19 exhibited a higher C_{12} to C_{14} product ratio when expressed in *E. coli*, the results of the kinetics study do not indicate that MMD19 is more selective than the WT BTE enzyme: there is no statistically significant difference between the ratios of the catalytic efficiency for dodecanoyl-ACP to tetradecanoyl-ACP for the WT and the MMD19 variant (Table S6 and Table S7). We evaluated if the *in vivo* performance of MMD19 could be supported by the ratio of the turnover number for dodecanoyl-ACP to tetradecanoyl-ACP. This could be plausible if the concentration of dodecanoyl-ACP and tetradecanoyl-ACP were significantly greater than the measured K_M, as the reaction rate can be approximated by the maximum velocity under these conditions. However, the concentration of total ACP in the cytosol has been estimated to be 6×10^4 to 8×10^4 molecules, or 150–200 μM, and the concentration of the individual acyl-ACPs is likely much less (Cronan and Thomas, 2009). Thus, despite the turnover number ratio being significantly higher for the MMD19 variant, the enzymatic performance in the intracellular environment should depend on both the turnover number and the Michaelis-Menten constant. A possible explanation for the discrepancy between the *in vivo* and *in vitro* data is that the *in vitro* acyl-ACP concentrations may not be representative of what the thioesterase



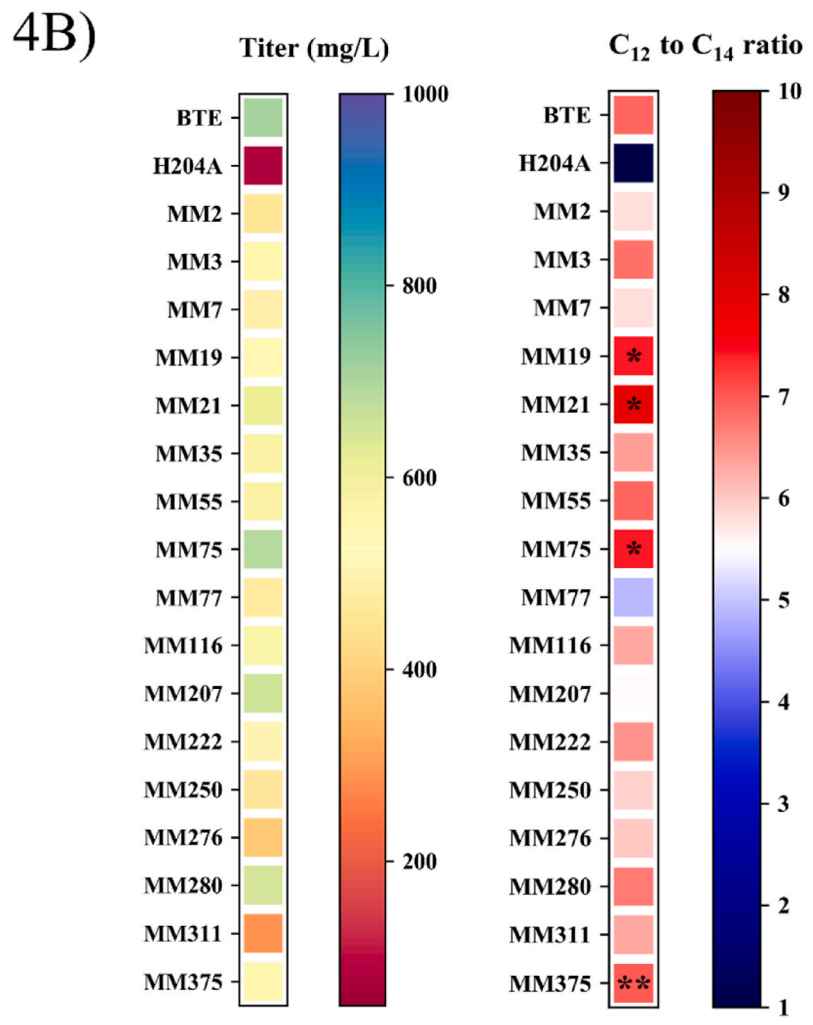


Fig. 4. (A) Pipeline for preparation, screening, and validation of BTE variants with enhanced C_{12} FFA production capability using MALDI-ToF MS screen. (B) Heatmaps showing *in vivo* results of BTE variants selected after MALDI-ToF analysis with total FFA titer on the left and the C_{12} to C_{14} ratio within the product distribution on the right. Titers represent methyl ester derivatives of bulk cell cultures, including FFA and membrane lipids. One asterisk indicates a p-value of less than 0.05 when subjecting C_{12} to C_{14} ratio of the respective variant and the WT to a one-sided *t*-test. Two asterisks indicate a p-value of less than 0.01.

experiences in the cytosol; the thioesterase is exposed to a dynamic pool of acyl-ACPs at various concentrations. Previous authors have shown acyl-ACP thioesterase behavior is not only affected by the preferred substrate but also by other species in the substrate pool; for example, the ChFatB2 thioesterase from *Cuphea hookeriana* produces over 80% octanoic acid when expressed in *E. coli*, yet exhibits a distinct preference for decanoyl-ACP when exposed to an equimolar pool of C_8, C_{10}, C_{12}, C_{14}, and C_{16} ACPs *in vitro* (Feng et al., 2018). An engineered medium-chain

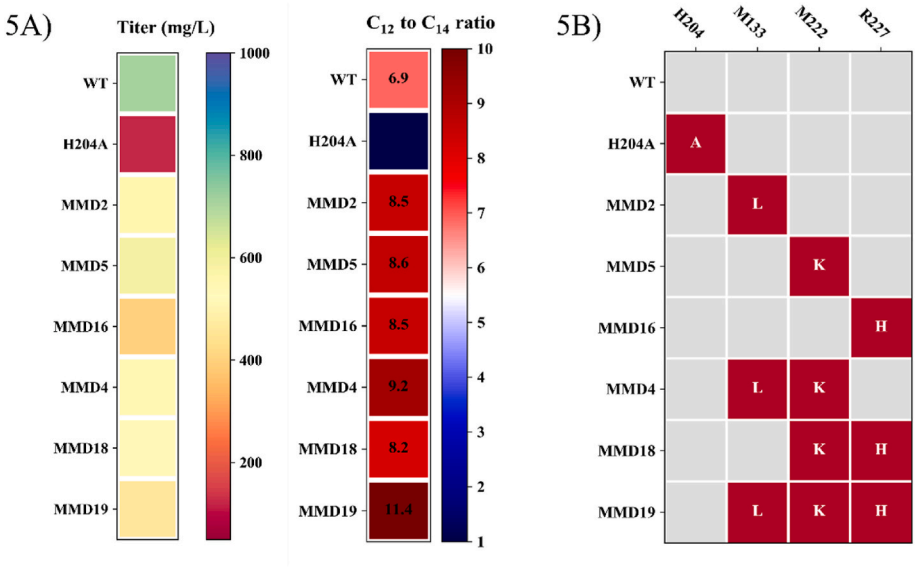


Fig. 5. (A) Heatmaps showing *in vivo* results of most relevant variants from the MMD library which led to the creation of BTE-MMD19, with total FFA titer on the left and C_{12} to C_{14} ratio on the right. Titers represent methyl ester derivatives of bulk cell cultures, including FFA and membrane lipids. The C_{12} to C_{14} ratio is inscribed in each block to illustrate the progression of selectivity in the product distribution as the mutations from the MALDI-ToF were sequentially introduced. (B) Grid of mutations incorporated into the selective members of MMD library. Red squares indicate a substitution along with the new residue incorporated at its respective position. Gray squares indicate no change in the given variant at the designated position.

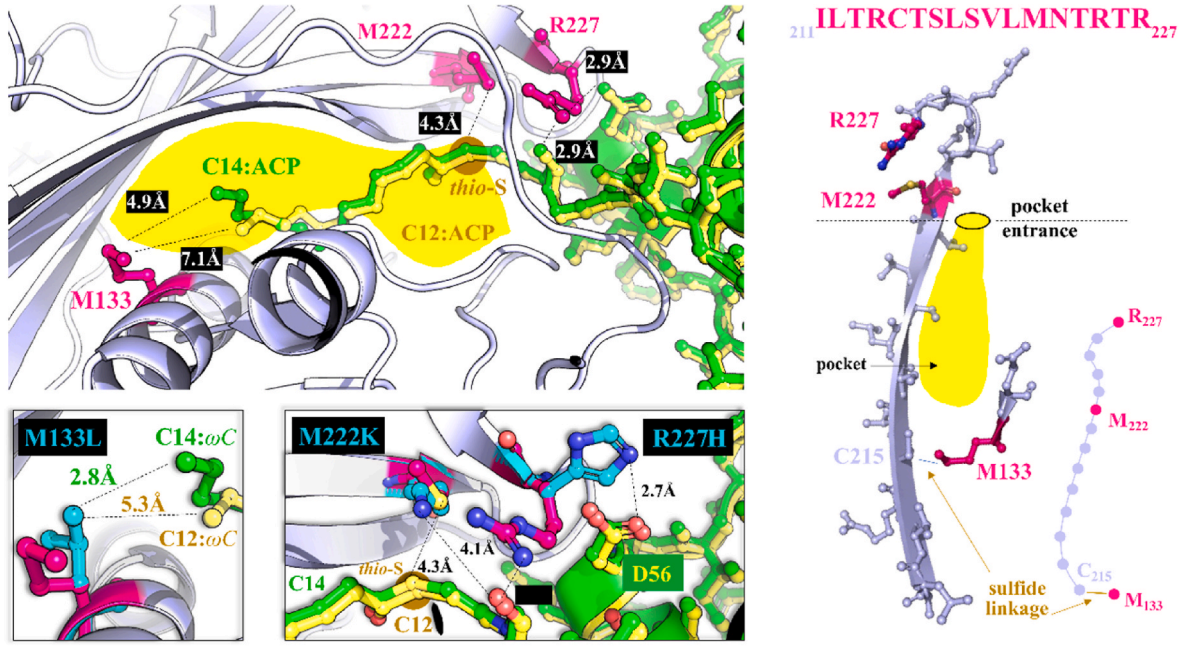


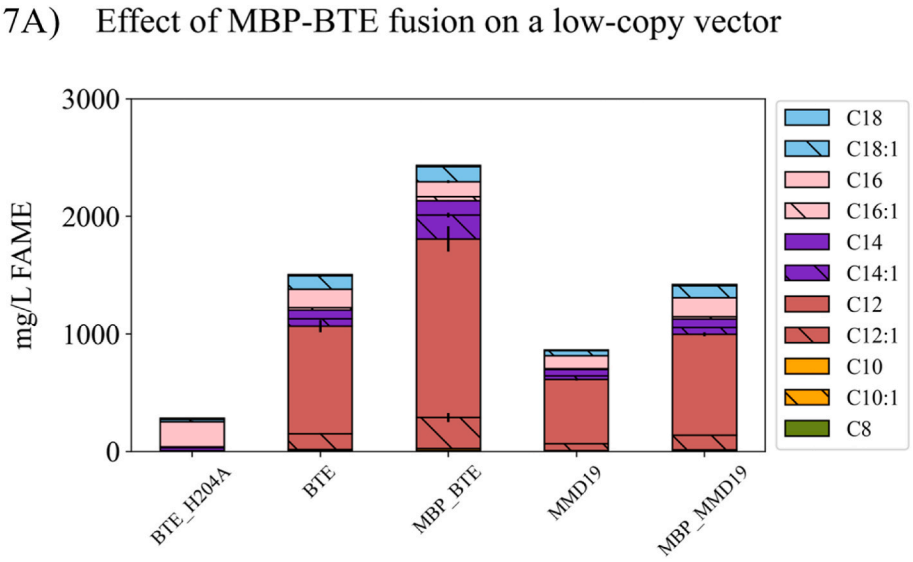
Fig. 6. The role of binding pocket variation between WT BTE and triple mutant in C_{12} and C_{14} specificity. Wild type residues R227, M222, and M133 control the recognition, entry, and stabilization of the acyl-ACP substrate, respectively. Even though two of these residues are far from the third one, they interact via a cascade of 12 covalent bonds (along the β5 sheet) and one sulfide-linkage (between C215 and M133). The pore-entrance mutations R227H and M222K (mutated residues in cyan), in fact weaken recognition and entry of both C_{12} and C_{14} substrates by ~15% (*in silico* binding score scale). However, the M133L mutation protrudes into the binding pocket (marked in yellow) and reduces the overall volume of the pocket (by ~22%) consequently forcing out substrates with more than twelve carbons. The weaker binding strength at the pocket mouth (R227H and M222K) facilitates destabilization of C_{14} substrates by inducing steric clashes near the omega carbon. However, dodecanoyl-ACPs are optimally sized to facilitate clash free hydrophobic packing to stabilize it within the pocket.

specific variant of the archetypal bacterial thioesterase, 'TesA, exhibited nearly no activity when subjected to a competitive assay containing 50 μM hexadecanoyl-CoA and 50 μM octanoyl-CoA despite displaying an initial rate of 0.4 μM octanoic acid per second when catalyzing 100 μM octanoyl-CoA (Grisewood et al., 2017). Finally, expression of BTE in microalgae results in a different product distribution than expression in *E. coli*, as BTE predominantly produces tetradecanoic acid instead of dodecanoic acid in *Chlamydomonas reinhardtii* (Blatti et al., 2012). Taken together, we propose that our MMD19 variant effectively leads to more selective medium-chain fatty acid product distributions in *E. coli*, however its performance is dependent on the acyl-ACP distribution in this organism.

3.6. Addressing solubility to enhance productivity

The kinetic assay indicated that while the MMD19 variant had reduced activity on tetradecanoyl-ACPs as intended, the amino acid substitutions also resulted in the apparent reduction of the turnover number and catalytic efficiency for dodecanoyl-ACPs. This kinetic behavior manifested itself in the production trials, as the MMD19 variant exhibited about a 35% reduction in overall FFA titer in the shake flask trials (Table S3). We reasoned we could remedy the titer loss by

titrating BTE expression through adjusting the strength of the ribosome binding site (RBS), as this has been an effective strategy in previous oleochemical pathway optimization efforts (Hernández Lozada et al., 2020), (Mehrer et al., 2018). We used the RBS strength calculator from *De Novo DNA* for prediction of the protein translation rate of the MMD19 variant and designed a small RBS library to test for an improvement in FFA production (Salis et al., 2010). Surprisingly, the FFA production was unimpacted by the change in RBS strength (Fig. S5). We surmised our strains were already at their expression limit for BTE and were likely generating inclusion bodies from the excess protein burden. Addressing solubility has been demonstrated an effective approach to improve enzymatic activity that is orthogonal to tuning catalytic parameters: Wrenbeck et al. showed that the incorporation of 8 mutations which increased the stability of the Type III polyketide synthase from *A. belladonna* also resulted in a 25*x* fold increase in activity in cell lysates (Wrenbeck et al., 2019). To similarly improve the solubility of the thioesterase, we redesigned the WT and MMD19 coding sequences to fuse the maltose binding protein from *E. coli* to the N-terminus (Kapust and Waugh, 1999). For this demonstration, we lowered the thioesterase expression level and used a buffered, carbon-rich media to generate conditions which accentuated differences in performance between the MBP-tagged and untagged constructs. We cloned tagged and untagged versions of BTE WT and MMD19 into the low-copy, inducible vector, pBTRCK, and subsequently cultured the RL08 transformants in a buffered, yeast extract-tryptone broth supplemented with 20 g per L of glycerol and calcium pantothenate (Clomburg et al., 2012). A C-terminal 6*x* histidine tag was also incorporated into the MBP-tagged constructs to facilitate purification. The FFA production from these constructs are shown in Fig. 7A. In these conditions, the addition of the MBP tag improved the C_{12} fatty acid titer for both the WT and the MMD19 variant by about 60%. However, in these conditions there was no significant difference in the C_{12} to C_{14} ratio. The constructs were cloned into the high-copy vector, pTrc99a, to determine if the strain could accommodate higher expression levels with the MBP tag addition and if the selectivity of the MMD19 variant would be restored. Interestingly, the fusion of MBP lowered the C_{12} fatty acid titer achieved in the culture



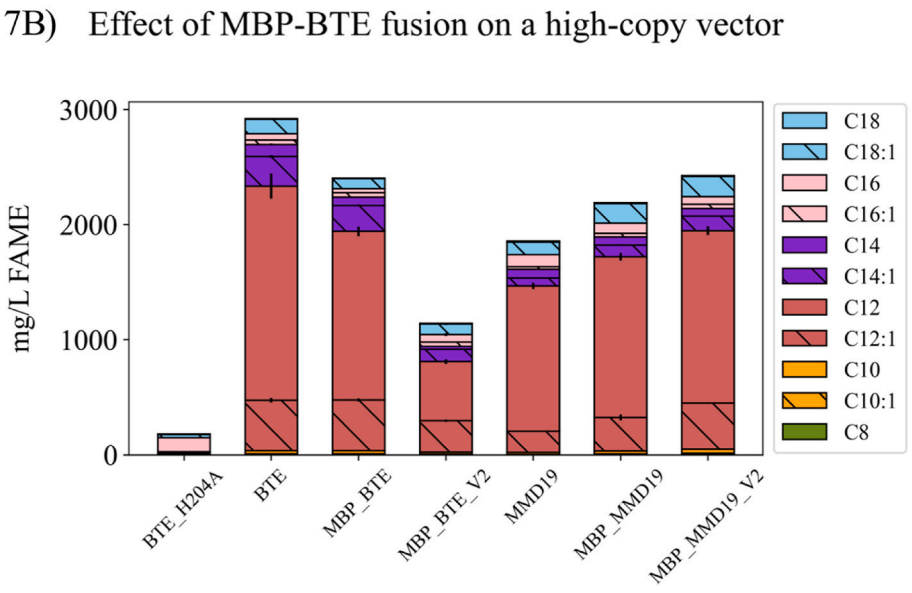


Fig. 7. (A) FFA distribution from RL08 expressing WT BTE and the MMD19 variant on the low-copy pBTRCK vector to demonstrate the effect of improved solubility on production. (B) FFA distribution from RL08 expressing WT BTE and the MMD19 variant on the high-copy pTrc99a vector to maximize expression and titer. Constructs with the MBP fusion were also tested with and without the C-terminal 6*x* histidine tag. Strains were cultured in the buffered, 20 g per L glycerol media, and cultures were induced at an OD of about 0.2 to a concentration of 20 μM IPTG. Concentrations obtained from bulk sample cultures are reported as FAMEs, the methyl ester derivatives of FFA and cell membrane lipids. Error bars represent the standard error of biological triplicates.

expressing the WT enzyme by 17%, yet it increased the titer by 17% in the culture expressing the MMD19 variant (Fig. 7B). We also investigated if removal of the 8*x* his tag would purportedly affect the solubility and thus the fatty acid titer. These samples are denoted as "V2" in Fig. 7B. This modification further reduced the C_{12} product titer of the WT enzyme by an additional 51% of the completely tag-free construct and improved the titer of the MMD19 construct by 14%. Furthermore, the MMD19 constructs exhibited an improvement in selectivity as previously demonstrated in the initial screen. The performance metrics (i.e. titer and C_{12} to C_{14} ratio) for the results shown in Fig. 7 are summarized in Table S8.

4. Conclusion

We engineered a highly selective dodecanoyl-ACP thioesterase which exhibits 90% of C_{12} products in the FFA pool. The mutations conferring this specificity shift occurred in the binding pocket as well as on the positively charged acyl-ACP landing pad, corroborating the general heuristics previously successful thioesterase engineering studies outlined. However, in this study, these design rules were not sufficient to rationally guide our mutagenesis for the purpose of meeting our selectivity goals. Thus, we advocate here for high-throughput screening of randomly generated libraries for the generation of highly selective thioesterases. To address solubility barriers limiting the amount of active enzyme in the cytosol, we fused the maltose binding protein from *E. coli* to the N-terminus of the MMD19 variant. The addition of the solubility tag boosted the titer in rich media by 31%, resulting in 1.9 g per L of C_{12} FFA obtained from rich media shake flask experiments.

We compared the results from the *in vivo* experiments and site saturation libraries to previous IPRO-guided thioesterase engineering with 'TesA to understand why this approach was not fruitful for this campaign (Pantazes et al., 2015), (Steen et al., 2010). While the IPRO algorithm was successful in generating multiple 'TesA variants with improved medium-chain substrate activity, most mutations suggested by the IPRO algorithm inactivated the BTE enzyme. The IPRO approach identifies mutations through changes in interaction energies within a single snapshot of the protein. It is likely that the crystal structure does not capture the confirmation necessary for catalysis nor does it capture the detrimental changes to structure stability caused by mutation incorporated in the IPRO variants. Additionally, the innate connection between thermostability and mutation tolerance could explain why 'TesA from *E. coli* has served as an apt platform for exploring and engineering specificity-altering mutations (Grisewood et al., 2017), (Bloom et al., 2006), (Deng et al., 2020). However, the high-preference for lauroyl-ACPs (also a C_{12} -ACP) in taxonomically distanced, 'TesA homolog, BTE, made it a strategic mutagenesis template. We hypothesized that potential for meeting our design goal of exclusive specificity could be arrived faster if we started with a protein that already recognizes the intended substrate length but is very sensitive to mutations. This work thus presents us with the classic exploration ('TesA) versus exploitation (BTE) dichotomy and prior art on functional protein design using $EdaRose_c$ also corroborates that exploitation is a better strategy while sampling low-energy protein folds with limited compute resources (Simoncini et al., 2017). Thus, we persisted with BTE as the preferred choice for the mutagenesis template. The use of MALDI-ToF mass spectrometry for high-throughput screening of a randomly mutagenized library was the superior approach for our design goals. This enabled the identification of four variants capable of enriching the FFA pool for C_{12} products in *E. coli*. While the general structure-function relationships of medium-chain specific acyl-ACP thioesterases have been documented, this prior knowledge was not sufficient to support our rational engineering campaign with the goal of near-exclusive free fatty acid product profiles. This emphasizes the value of high-throughput screens to complement discovery pipelines while we continue to elucidate the mechanisms for chain-length selectivity.

The intersection of synthetic biology and protein engineering enables platforms for on-purpose production of medium-chain free fatty acids and oleochemical derivatives. A key enabling technology which grants industrial biotechnology its competitiveness in this space is highly specific acyl-ACP thioesterases. Currently, highly exclusive C_8 and C_{12} acyl-ACP thioesterases have been demonstrated in bacterial hosts. The next stage in development will not only include engineering a more specific C_{10} thioesterase, but will also require enhancing the activity of downstream enzymes which convert the FFA pool to value-added derivatives. In addition, translating the benefits of these highly selective thioesterases to oleaginous plants and yeasts will continue to be a focus of synthetic biologists (Park et al., 2021), (Rutter et al., 2015). We present this work on enhancing product selectivity as our contribution to advancing these efforts and realizing a sustainable pipeline for oleochemicals.

Declaration of competing interest

The authors declare no competing interests.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the National Science Foundation (CBET-1703504, link: https://www.nsf.gov/) the 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, link: https://cabbi. bio/), and by U.S. Department of Energy, Advanced Research Projects Agency – Energy, Research under Award Number DE-AR0000ABC, link: https://arpa-e.energy.gov/). The NSF and ARPA-E awards were received by BFP. The CABBI award was received by BFP and JVS. 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 or the National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors are grateful to Qiang Yan and William H. Bothfeld for helpful discussions throughout the course of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymben.2023.02.012.

References

Agnew, D.E., Stevermer, A.K., Youngquist, J.T., Pfleger, B.F., 2012. Engineering Escherichia coli for production of C 12 – C 14 polyhydroxyalkanoate from glucose. Metab. Eng. 14 (6), 705–713.
Amann, E., Ochs, B., Abel, K.J., 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. Gene 69 (2), 301–315.
Amira, P.O., Babalola, O.O., Mary, O.A., 2014. Physicochemical properties of palm kernel oil. Curr. Res. J. Biol. Sci. 6 (5), 205–207.
Baek, M., et al., 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science 876 (August), 871–876.
Banerjee, D., Jindra, M.A., Linot, A.J., Pfleger, B.F., Maranas, C.D., 2022. EnZymClass: substrate specificity prediction tool of plant acyl-ACP thioesterases based on ensemble learning. Curr. Res. Biotechnol. 4, 1–9. December 2021.
Blatti, J.L., Beld, J., Behnke, C.A., Mendez, M., Mayfield, S.P., Burkart, M.D., 2012. Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. PLoS One 7 (9), 1–12.
Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H., 2006. Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA 103 (15), 5869–5874.
Cantu, D.C., Chen, Y., Reilly, P.J., 2010. Thioesterases: a new perspective based on their primary and tertiary structures. Protein Sci. 19 (7), 1281–1295.

Choe, K., Xue, P., Zhao, H., Sweedler, J.V., 2021. macroMS: image-guided analysis of random objects by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J. Am. Soc. Mass Spectrom. 32 (5), 1180–1188.
Chowdhury, R., et al., 2022. Single-sequence protein structure prediction using a language model and deep learning. Nat. Biotechnol. 40, 1617–1623.
Clomburg, J.M., Vick, J.E., Blankschien, M.D., Rodríguez-Moya, M., Gonzalez, R., 2012. A synthetic biology approach to engineer a functional reversal of the β - oxidation cycle. ACS Sustain. Chem. Eng. 1, 541–554.
Cronan, J.E., Thomas, J., 2009. Bacterial fatty acid synthesis and its relationships with polyketide. Methods Enzymol. 459, 395–433.
Deng, X., Chen, L., Hei, M., Liu, T., Feng, Y., Yang, G.Y., 2020. Structure-guided reshaping of the acyl binding pocket of 'TesA thioesterase enhances octanoic acid production in E. coli. Metab. Eng. 61, 24–32. January.
Feng, Y., Wang, Y., Liu, J., Liu, Y., Cao, X., Xue, S., 2017. Structural insight into acyl-ACP thioesterase toward substrate specificity design. ACS Chem. Biol. 12 (11), 2830–2836.
Feng, Y., et al., 2018. Tuning of acyl-ACP thioesterase activity directed for tailored fatty acid synthesis. Appl. Microbiol. Biotechnol. 102 (7), 3173–3182.
Greenspan, P., Mayer, E.P., Fowler, S.D., 1985. Nile red: a selective fluorescent stain for intracellular lipid droplets. J. Cell Biol. 100 (3), 965–973.
Grisewood, M.J., et al., 2017. Computational redesign of acyl-ACP thioesterase with improved selectivity toward medium-chain-length fatty acids. ACS Catal. 7, 3837–3849.
Harris, R., Olson, A.J., Goodsell, D.S., 2008. Automated prediction of ligand-binding sites in proteins. Proteins 70 (4), 1506–1517.
He, X., Man, V.H., Yang, W., 2020. A fast and high-quality charge model for the next generation general AMBER force A fast and high-quality charge model for the next generation general AMBER force field. J. Chem. Phys. 153, 1–11, 114502.
Hernández Lozada, N.J., et al., 2018. Highly active C 8 -Acyl-ACP thioesterase variant isolated by a synthetic selection strategy. ACS Synth. Biol. 7 (9), 2205–2215.
Hernández Lozada, N.J., Simmons, T.R., Xu, K., Jindra, M.A., Pfleger, B.F., 2020. Production of 1-octanol in Escherichia coli by a high flux thioesterase route. Metab. Eng. 61 (April), 352–359.
Jing, F., et al., 2011. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity 1–16.
Jing, F., Zhao, L., Yandeau-Nelson, M.D., Nikolau, B.J., 2018. Two distinct domains contribute to the substrate acyl chain length selectivity of plant acyl-ACP thioesterase. Nat. Commun. 9 (1), 860.
Kapust, R.B., Waugh, D.S., 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8, 1668–1674.
Lennen, R.M., Braden, D.J., West, R.M., Dumesic, J.A., Pfleger, B.F., 2010. A process for microbial hydrocarbon synthesis : overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. Biotechnol. Bioeng. 106 (2), 193–202.
Mehrer, C.R., Incha, M.R., Politz, M.C., Pfleger, B.F., 2018. Anaerobic production of medium-chain fatty alcohols via a β-reduction pathway. Metab. Eng. 48 (May), 63–71.
Michaud, J.M., Madani, A., Fraser, J.S., 2022. A language model beats alphafold2 on orphans. Nature.
Pantazes, R.J., Grisewood, M.J., Li, T., Gifford, N.P., Maranas, C.D., 2015. The Iterative Protein Redesign and Optimization (IPRO) Suite of Programs, pp. 251–263.
Park, K., Sanjaya, S.A., Quach, T., Cahoon, E.B., 2021. Toward sustainable production of value-added bioenergy and industrial oils in oilseed and biomass feedstocks. GCB Bioenergy 13 (May), 1610–1623.
Parsons, S., Chuck, C.J., Mcmanus, M.C., 2018. Microbial lipids : progress in life cycle assessment (LCA) and future outlook of heterotrophic algae and yeast-derived oils. J. Clean. Prod. 172, 661–672.
Parsons, S., Raikova, S., Chuck, C.J., 2020. The viability and desirability of replacing palm oil. Nat. Sustain. 3 (6), 412–418.
Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291 (5509), 1790–1792.
Pfleger, B.F., Gossing, M., Nielsen, J., 2015. Metabolic engineering strategies for microbial synthesis of oleochemicals. Metab. Eng. 29, 1–11.
Ronneberger, O., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596 (August), 583–591.
Rupilius, W., Ahmad, S., 2007. Palm oil and palm kernel oil as raw materials for basic oleochemicals and biodiesel. Eur. J. Lipid Sci. Technol. 109 (4), 433–439.
Rutter, C.D., Zhang, S., V Rao, C., 2015. Engineering Yarrowia lipolytica for production of medium-chain fatty acids. Appl. Microbiol. Biotechnol. 99, 7359–7368.
Salis, H.M., a Mirsky, E., a Voigt, C., 2010. Automated design of synthetic ribosome binding sites to precisely control protein expression. Nat. Biotechnol. 27 (10), 946–950.
Saraf, M.C., Moore, G.L., Goodey, N.M., Cao, V.Y., Benkovic, S.J., Maranas, C.D., 2006. IPRO : an Iterative Computational Protein Library Redesign and Optimization Procedure, vol. 90. June.
Sarria, S., Kruyer, N.S., Peralta-yahya, P., 2017. Microbial synthesis of medium-chain chemicals from renewables. Nat. Biotechnol. 35 (12), 1158–1166.
Sarria, S., Bartholow, T.G., Verga, A., Burkart, M.D., Peralta-Yahya, P., 2018. Matching protein interfaces for improved medium-chain fatty acid production. ACS Synth. Biol. 7 (5), 1179–1187.
Simoncini, D., Schiex, T., Zhang, K.Y.J., 2017. Balancing exploration and exploitation in population-based sampling improves fragment-based de novo protein structure prediction. Proteins 85 (5), 852–858.
Steen, E.J., et al., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463 (7280), 559–562.
Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Simmerling, C., 2020. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J. Chem. Theor. Comput. 16, 528–552.
Trott, O., Olson, A.J., 2010. AutoDock Vina : improving the speed and accuracy of docking with a new scoring function , efficient optimization , and multithreading. J. Comput. Chem. 31, 455–461.
Voelker, T.A., Davies, H.M., 1994. 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 (23), 7320–7327.
Voelker, T.A., et al., 1992. Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257 (7), 72–74.
Wang, C., Wang, M., Han, X., 2015. Applications of mass spectrometry for cellular lipid analysis. Mol. Biosyst. 11, 698–713.
Wilcove, D.S., Pin, L., 2010. Addressing the threats to biodiversity from oil-palm agriculture. Biodivers. Conserv. 19 (4), 999–1007.
Wrenbeck, E.E., Klesmith, J.R., Stapleton, J.A., Adeniran, A., Tyo, K.E.J., Whitehead, T. A., 2016a. Plasmid-based one-pot saturation mutagenesis. Nat. Methods 13 (11), 3–7.
Wrenbeck, E.E., Klesmith, J.R., Stapleton, J.A., Adeniran, A., Tyo, K.E.J., Whitehead, T. A., 2016b. Plasmid-based one-pot saturation mutagenesis. Nat. Methods 13 (11), 928–930.
Wrenbeck, E.E., Bedewitz, M.A., Klesmith, J.R., Noshin, S., Barry, C.S., Whitehead, T.A., 2019. An automated data-driven pipeline for improving heterologous enzyme expression. ACS Synth. Biol. 8 (3), 474–481.
Xue, P., et al., 2020. A mass spectrometry-based high-throughput screening method for engineering fatty acid synthases with improved production of medium-chain fatty acids. Biotechnol. Bioeng. 117 (7), 2131–2138.
Yan, Q., et al., 2021. Metabolic engineering of β-oxidation to leverage thioesterases for production of 2-heptanone, 2-nonanone and 2-undecanone. Metab. Eng. 61 (9), 335–343.
Youngquist, J.T., et al., 2013. Production of medium chain length fatty alcohols from glucose in Escherichia coli. Metab. Eng. 20, 177–186.
Yuan, L., Voelker, T.A., Hawkins, D.J., 1995. Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proc. Natl. Acad. Sci. USA 92, 10639–10643. November.