ACS Catalysis

Research Article

pubs.acs.org/acscatalysis

Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids

Matthew J. Grisewood,[†,▽] Néstor J. Hernández-Lozada,[‡,▽] James B. Thoden,[§] Nathanael P. Gifford,[†] Daniel Mendez-Perez,[‡] Haley A. Schoenberger,[‡] Matthew F. Allan,[†] Martha E. Floy,[‡] Rung-Yi Lai,[‡] Hazel M. Holden,[§] Brian F. Pfleger,[‡] and Costas D. Maranas*[,†]

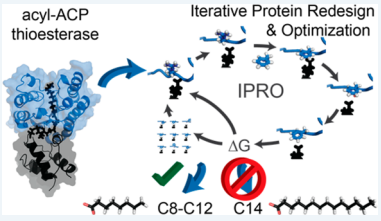
[†]Department of Chemical Engineering, Pennsylvania State University, 158 Fenske Laboratory, University Park, Pennsylvania 16802, United States

[‡]Department of Chemical and Biological Engineering, University of Wisconsin−Madison, 1415 Engineering Drive, Madison, Wisconsin 53706, United States

[§]Department of Biochemistry, University of Wisconsin−Madison, 440 Henry Mall, Madison, Wisconsin 53706, United States

Supporting Information

ABSTRACT: Enzyme and metabolic engineering offer the potential to develop biocatalysts for converting natural resources to a wide range of chemicals. To broaden the scope of potential products beyond natural metabolites, methods of engineering enzymes to accept alternative substrates and/or perform novel chemistries must be developed. DNA synthesis can create large libraries of enzyme-coding sequences, but most biochemistries lack a simple assay to screen for promising enzyme variants. Our solution to this challenge is structure-guided mutagenesis, in which optimization algorithms select the best sequences from libraries based on specified criteria (i.e., binding selectivity). Here, we demonstrate this approach by identifying medium-chain (C_8−C_{12}) acyl-ACP thioesterases through structure-guided mutagenesis. Medium-chain fatty acids, which are products of thioesterase-catalyzed hydrolysis, are limited in natural abundance, compared to long-chain fatty acids; the limited supply leads to high costs of C_6−C_{10} oleochemicals such as fatty alcohols, amines, and esters. Here, we applied computational tools to tune substrate binding of the highly active 'TesA thioesterase in *Escherichia coli*. We used the IPRO algorithm to design thioesterase variants with enhanced C_{12} or C_8 specificity, while maintaining high activity. After four rounds of structure-guided mutagenesis, we identified 3 variants with enhanced production of dodecanoic acid (C_{12}) and 27 variants with enhanced production of octanoic acid (C_8). The top variants reached up to 49% C_{12} and 50% C_8 while exceeding native levels of total free fatty acids. A comparably sized library created by random mutagenesis failed to identify promising mutants. The chain length-preference of 'TesA and the best mutant were confirmed *in vitro* using acyl-CoA substrates. Molecular dynamics simulations, confirmed by resolved crystal structures, of 'TesA variants suggest that hydrophobic forces govern 'TesA substrate specificity. We expect the design rules that we uncovered and the thioesterase variants that we identified will be useful to metabolic engineering projects aimed at sustainable production of medium-chain-length oleochemicals.



KEYWORDS: *molecular dynamics, IPRO, thioesterase, 'TesA, fatty acid, redesign, dodecanoic acid, octanoic acid*

INTRODUCTION

Free fatty acids (FFAs) are energy-rich precursors of membrane lipids, natural oils, liquid transportation fuels (i.e., biodiesel), and high-value oleochemicals (e.g., fatty alcohols, aldehydes, olefins, and waxes).[1−4] Oleochemical properties such as energy content, melting point, and volatility are dictated by the chain length, degree of saturation, and branching pattern of the acyl chain.[5] Fuels and oleochemicals derived from microbially produced FFAs could displace current, unsustainable plant feedstocks and reduce carbon footprints, relative to petrochemical alternatives.[6,7] Unfortunately, natural sources of medium-chain length FFAs and lipids are significantly less abundant than longer-chain compounds. The limited supply and costly petrochemical synthesis alternative leads to higher selling prices for medium-chain oleochemicals (e.g., 1-octanol costs approximately twice as much as 1-hexadecanol per pound).[4] These economic drivers make bioproduction of medium-chain length FFAs and oleochemicals an attractive opportunity if biosynthesis pathways with high yield and selectivity can be assembled and optimized.

While many oleochemical pathways have been demonstrated in model hosts, chain length selectivity remains an unsolved challenge. The product distribution of most metabolic engineering efforts has been restricted to the chain length of the most abundant acyl-thioester in the cell or the distribution created by expression of a thioesterase. Acyl chains that comprise FFAs, lipids, and oleochemicals are made by an iterative series of elongation, keto-reduction, dehydration, and

Received: February 7, 2017
Revised: April 13, 2017
Published: April 20, 2017

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

enoyl-reduction reactions acting on acyl-coenzyme A (acyl-CoA) or acyl−acyl-carrier protein (acyl-ACP) thioesters.[4] An acyl chain is elongated by two carbon atoms per cycle until it is trans-esterified (into phospholipids, waxes, or esters), reduced (to a fatty aldehyde or alcohol), or hydrolyzed (yielding a FFA).[1] The chain-length distribution of these terminal products is controlled by the relative kinetics of elongation, trans-esterification, reduction, and/or hydrolysis. In *Escherichia coli*, the activity of phospholipid synthases (PlsB, PlsC) and fatty acid synthases (FabB, FabF, FabH) constrain the lipid composition to mostly C_{16} or C_{18} acyl-chains with little FFA content.[8] In contrast, *E. coli* can produce high titers of FFA with a wide range of chain-length distributions by the upregulation of native thioesterases (TesA and TesB) or heterologous expression of plant and bacterial thioesterases (see Table S1 in the Supporting Information). In the absence of pathways for catabolizing FFAs and/or acyl-thioesters (i.e., beta-oxidation), the specificity of the acyl-ACP thioesterase controls the chain-length distribution and the chemical properties of downstream oleochemicals (see Figure 1). Collectively, thioesterases exhibit

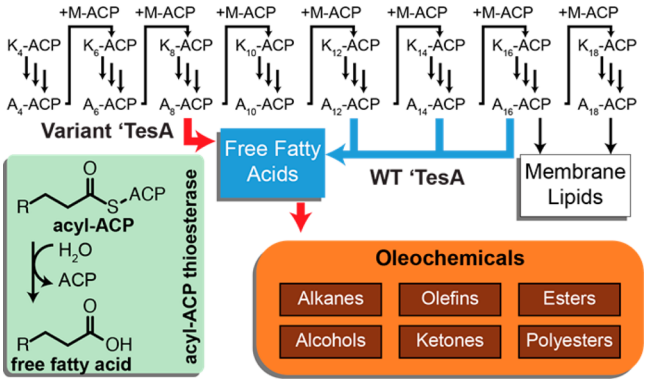


Figure 1. Overview of fatty acid biosynthesis and role of acyl-ACP thioesterases. In each cycle of fatty acid biosynthesis, two carbons are added from a malonyl-ACP (M-ACP), yielding a β-ketoacyl-ACP (K_x-ACP where x is the number of carbons). Three reactions (three vertical arrows) reduce the K_x-ACP to a saturated acyl-ACP (A_x-ACP). In *E. coli*, A_{16}-ACP and A_{18}-ACP are incorporated into membrane lipids. Thioesterases produce FFAs by hydrolyzing the acyl-thioester bond (green inset). Tailoring the specificity of the acyl-ACP thioesterase ('TesA) dictates FFA and downstream oleochemical chain lengths (red arrows).

a wide range of substrate specificities[2,9−14] that has been further diversified through protein engineering and evolution.[15−17] Even with this known diversity, very few thioesterases are specific toward a single aliphatic chain length.[10] Worse, heterologous expression of thioesterases frequently begets unexpected product distributions,[4,10] poor FFA yields,[18,19] or both. Of the many studied thioesterases, 'TesA (a cytosolic *E. coli* TesA lacking the N-terminal periplasmic signal peptide) has been used extensively in metabolic engineering studies and has a crystallographically resolved structure.[20] However, 'TesA has broad substrate specificity with relatively low medium-chain content.[2,9] For these reasons, 'TesA is an attractive system for applying enzyme engineering to improve thioesterase selectivity toward medium-chain lengths.

Enzyme engineering is generally pursued using directed evolution approaches that rely on high-throughput screening of large mutant libraries.[21] These large libraries are constructed using various mechanisms for diversifying the gene pool, including homologous (e.g., DNA shuffling) or nonhomologous (e.g., overlap extension PCR) recombination, random mutagenesis (e.g., error-prone PCR), or combinations thereof (see ref 22 for review). Large library sizes rely on high-throughput screening that takes advantage of optical properties, such as fluorescent or colorimetric assays. Currently, no high-throughput screen that can discriminate between different FFA chain lengths has been developed. When optical screens are unsuitable, more laborious experiments (e.g., mass spectrometry or NMR spectroscopy) can be used in low-throughput screens that mandate small, more-focused libraries. Focused libraries can be generated through site saturation mutagenesis (e.g., degenerate oligonucleotide-primed PCR), but this approach can only include a small number of sites (three saturated sites would yield 8000 variants) with a high percentage of inactive mutants. Smaller library sizes can be formed through site-directed mutagenesis[23] (i.e., rational design), but, so far, it has been very difficult to forecast the effect of multiple mutations. When a protein structure is known, molecular modeling tools can suggest a handful of promising mutations while considering sequence-structure relationships, as well as the approximations.

Structure-based protein redesign procedures, such as the Iterative Protein Redesign and Optimization (IPRO) method[24] used here, offer several advantages over tools that simply suggest "hot spot" residues,[25−30] because they can capture the simultaneous effect of multiple mutations. Another class of computational protein redesign tools uses mostly sequence information to suggest crossover locations for generating combinatorial libraries.[31−37] However, these combinatorial libraries are limited by the parental sequence space and do not take full advantage of available structural information. *De novo* enzyme designs[38−42] are usually less active than native ones (without the aid of directed evolution).[38,39,41] Thus, structure-based protein redesign can harness natural protein performance and suggest directed modifications using structural insight into meeting a single or multiple design objectives. IPRO differs from other structure-based protein redesign procedures in that it employs a mixed-integer linear program to guarantee a global minimum for a given protein backbone structure, it can handle multiple decision criterion simultaneously, and distance restraints can be easily imposed to keep catalytic machinery intact.[24,43,44] More-extensive reviews of existing computational protein engineering procedures have been provided by Pantazes et al.,[45] Samish et al.,[46] and Huang et al.[40]

In this study, we applied the IPRO method[24] to guide 'TesA mutagenesis in search of variants that both improve medium-chain FFA specificity and maintain high thioesterase activity. Specifically, we engineered 'TesA to yield additional dodecanoic acid (C12:0) and octanoic acid (C8:0) at the expense of the natively preferred tetradecanoic acid (C14:0). The employed redesign procedure involved recursively predicting *in silico* 'TesA mutants with enhanced binding capabilities, analyzing the *in vivo* FFA composition, and modifying the computational explorations in a Design-Build-Test-Learn cycle.[47] We identified 3 separate 'TesA mutants that exhibited a statistically significant ($p < 0.05$) improvement in C_{12} composition over wild-type (WT) and 27 mutants with a statistically significant improvement in C_8 composition. In comparison to previously studied thioesterases (see Table S1 in the Supporting Information), our computationally predicted mutants include 4 of the 10 most C_{12}-specific and 1 of the 10 most C_8-specific

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

thioesterases. While computational enzyme redesign that reaches industrially relevant performance metrics has so far remained elusive,[45] the results presented here demonstrate the potential of the adopted Design-Build-Test-Learn paradigm to pinpoint promising enzyme mutants.[45,47,48]

RESULTS AND DISCUSSION

Overview of the Design-Build-Test-Learn Approach. In support of efforts to alter the product profile of 'TesA toward medium-chain-length FFAs, we performed four rounds of site-directed mutagenesis structured around a Design-Build-Test-Learn cycle (Figure 2). In each round, mutagenesis targets and specific amino acid substitutions were selected after analyzing simulations of enzyme−substrate binding performed with IPRO.[24] Genes encoding the designed 'TesA variants were constructed by Quikchange or Gibson Assembly of PCR products (see the "Materials and Methods" section) and cloned into arabinose-inducible expression vectors. The impact of each mutation was assessed by quantifying the FFA content of a 'TesA expressing *E. coli* culture (see Table S1). After each round, the product distributions for each mutant were analyzed and used to improve to the IPRO framework[24] via changes in the scoring function parameters or refocusing the mutagenesis targets. The following sections describe the deployment of our computational enzyme Design phase, the results of the Build-Test phase, and a discussion of what we have Learned about thioesterase selectivity after each round.

IPRO Constraints and Implementation. IPRO enzyme redesign requires a three-dimensional (3D) model of the protein complex structure, a set of predetermined mutable residues known as design positions, and a set of constraints that quantify the desired improvements that protein variants must reach, relative to wild-type. Examples of such constraints include imposing relations that require (i) a stronger interaction energy with the new substrate and (ii) a weaker interaction energy with the native substrate, compared to wild-type. Here, we constructed a structural model of 'TesA bound to an acyl-ACP with acyl chains ranging from C_8 to C_{14}. The model was assembled from published structures of 'TesA (PDB 1U8U)[20] and decanoyl-ACP (PDB 2FAE).[49] Docking between 'TesA and octanoyl-ACP was modeled by systematically rotating the octanoyl-ACP structure about its phosphopantetheine linker until it aligned with bound octanoic acid in the 'TesA crystal structure (PDB 1U8U; see the "Materials and Methods" section). Other 'TesA:acyl-ACP complexes were derived from the 'TesA:octanoyl-ACP complex structure by adding atoms to the ω-1 carbon (i.e., the carbon furthest from ACP) of the octanoyl-ACP structure. The IPRO algorithm was used to search for mutations that led to improved binding between 'TesA and the desired substrate, while simultaneously discouraging binding of undesired substrate(s). Design positions for 'TesA were selected based on proximity to the ω-1 carbon of bound tetradecanoyl-ACP while not considering positions vital for catalytic turnover (see the "Materials and Methods" section). The fitness of each predicted variant was assessed using interaction energy as a proxy for binding energy (i.e., ΔG), thereby reducing force field dependence and requiring fewer calculations. Interaction energy is defined as $G_{Enz:FFA,min} - G_{Enz} - G_{FFA}$, where G is the Gibbs free energy, "Enz" represents the enzyme ('TesA or a variant thereof), and "min" indicates that the molecule(s) have undergone an energy minimization. In contrast, binding energy is defined as $G_{Enz:FFA,min} - G_{Enz,min} - G_{FFA,min}$. The first constraint is aimed

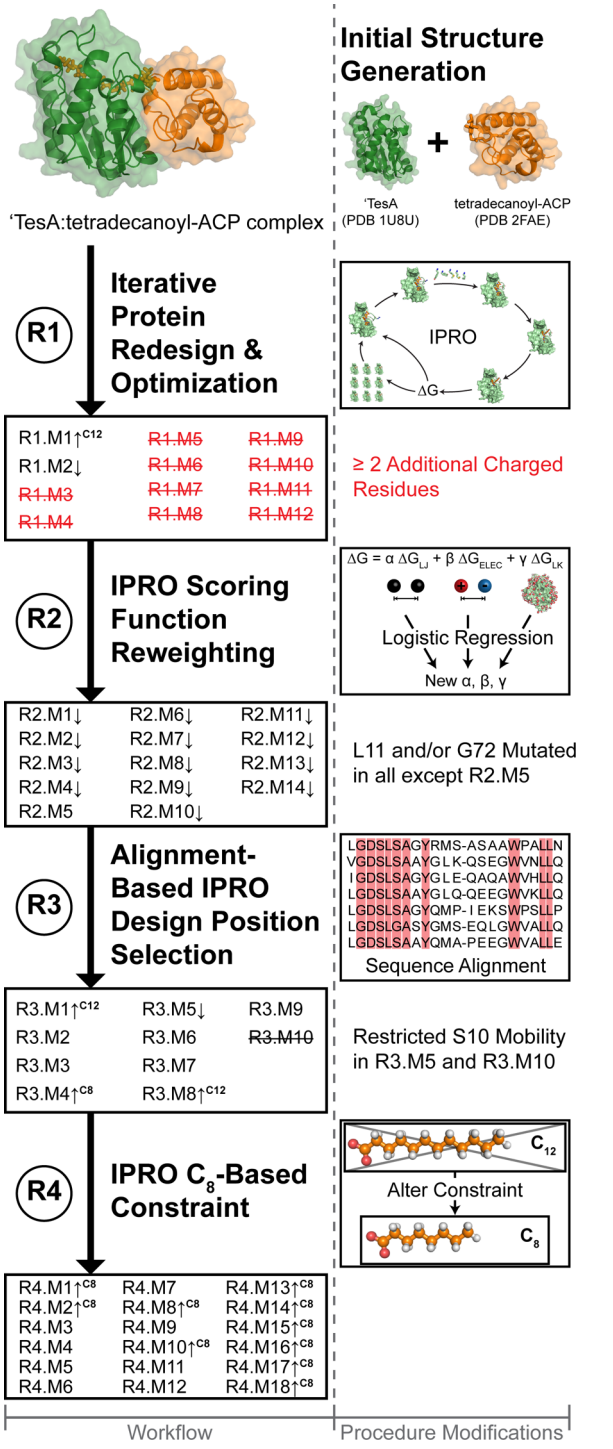


Figure 2. Overview of Predict-Design-Revise-Learn approach used to guide 'TesA redesign. The left-hand column shows a pictorial illustration of the steps traversed and mutants identified with improved specificity toward C_{12}- or C_8-acyl-ACPs. The right-hand column denotes the changes in the computational procedure modifications in response to the experimental results. Variants that produced significant improvements in the C_{12} ($p < 0.05$) or C_8 ($p < 0.005$) fraction while maintaining WT production levels are indicated with an upward arrow followed by the FFA. Major improvements in C_8 are indicated by $p < 0.005$ rather than $p < 0.05$ to highlight top designs. Struck-through variants indicate enzyme inactivity, and names followed by a downward arrow represent variants with reduced total production levels relative to WT.

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

at worsening the interaction energy between 'TesA and the native substrate, tetradecanoyl-ACP (e.g., C_{14}). This first constraint safeguards against binding to even longer acyl-ACPs (e.g., C_{16}, C_{18}), because the repulsive interactions, which disfavor interactions with the long acyl-ACP (e.g., C_{14} and longer), are further exaggerated, because of steric clashes or hydrophobicity. The second constraint requires interaction energy improvements between 'TesA and the medium-chain-length acyl-ACP (e.g., C_{12}). The imposition of the second constraint attempts to enhance binding with C_{12} or even shorter acyl-ACPs (e.g., C_8, C_{10}).

IPRO operates by successively performing redesign iterations for a preset number of cycles until a variant that simultaneously satisfies all constraints and optimizes the objective function is found. IPRO has been applied previously to modify *E. coli* β-glucuronidase substrate specificity,[50] alter the cofactor specificity of *Candida boidinii* xylose reductase,[51] graft a calcium-binding pocket into *Thermoactinomyces vulgaris* thermitase or a copper-binding pocket into *E. coli* thioredoxin,[52] and *de novo* design antibody variable regions that target influenza hemagglutinin, HIV gp120, and Ebola GP1-GP2 viruses.[42] Each IPRO iteration begins with a local backbone perturbation nearby a randomly selected design position from the requisite input set. The second step implements a mixed-integer linear program (MILP) to identify the optimal set of amino acids for the new backbone conformation. The third step of IPRO executes a local, rigid-body docking (i.e., intramolecular movements are prohibited and only relative positioning between the enzyme and ligand is considered) to reorient the ligand (i.e., the acyl-ACP) within the binding site. The fourth step performs an energy minimization of the entire enzyme complex. The fifth and final step evaluates the constraints set forth (i.e., (i) reducing binding to C_{14} and (ii) increasing binding to C_{12}). The results of the iteration are retained or discarded based on the Metropolis criterion, whereby worsening solutions are accepted with a decaying exponential probability as in simulated annealing (see ref 24 for further details). The two constraints are imposed only at the ground state for the thioester hydrolysis with no additional calculations at the transition state. Instead, simple restraints on catalytic distances were imposed during all IPRO iterations to preserve, but not necessarily boost, catalytic activity.[50] Multiple IPRO trajectories (~10 independent trajectories) were simulated to discover alternative routes for improving specificity. For each variant, the difference of interaction energies for the short and long acyl-ACP with 'TesA (i.e., $\Delta IE = IE_{C12} - IE_{C14}$) was calculated. The variants were prioritized based on the extent of the energy differences, and 10−20 variants were selected to build a focused library for experimental testing.

Method Implementation for 'TesA Redesign. The first round (R1) of IPRO-guided mutagenesis helped tune the parameters of the scoring function to improve prediction accuracy. Our initial objective was a modest shift in substrate preference from C_{14} to C_{12}, which comprised up to 20% of the native 'TesA product profile, in order to assess the efficacy of the redesign protocol. Of 12 tested variants, 1 improved C_{12} composition and maintained WT production levels (i.e., $\geq WT_{Total}$, $p < 0.05$; R1.M1; see Table S1). In contrast, 10 were inactive (i.e., $<Control_{Total}$) and included more than 2 charged substitutions (all except R1.M1 and R1.M2). The scoring function implemented within IPRO's MILP rotamer−residue selection algorithm was identified as the source of the charged residue bias. The scoring function energy terms were reweighted using logistic regression on a dataset of high-quality protein structures.[53] The updated scoring function roughly doubled native rotamer recovery relative to the existing scoring function (see the "Materials and Methods" section). The former scoring function was adequate for earlier systems,[24,42,50−52] but the high hydrophobicity of 'TesA (the energy term that was underemphasized in the former scoring function) made this enzyme especially susceptible to unsuccessful designs.

The modified scoring function was used to design a second round (R2) of variants with improved activity on C_{12} acyl chains. While all 14 variants were active, 13 mutants (all except R2.M5) produced less total FFA than WT and none improved the C_{12} fraction ($p < 0.05$). All 13 variants incorporated a mutation at either position L11 or position G72. These positions, in hindsight, were deemed conserved (exhibiting 86.3% and 60.0% sequence conservation, respectively, across the L1-like lysophospholipase subgroup of the SGNH-hydrolase family), since they are immediately adjacent to S10 and N73, which are part of the oxyanion hole.[20] Therefore, both positions L11 and G72 were eliminated from the list of design positions in subsequent rounds.

The third round (R3) of mutagenesis, based on the updated set of design positions and revised scoring function, resulted in a higher fraction of variants that maintained WT activity (80.0%), compared to that of R1 and R2 (8.3% and 7.1%, respectively). Of the 10 tested R3 designs, 2 improved the C_{12} product composition (R3.M1, R3.M8; $p < 0.05$) and 1 led to a major increase in the C_8 mole fraction (R3.M4; $p < 0.005$). Notably, since the design constraints imposed by IPRO in rounds R1−R3 did not preclude binding to acyl-ACPs smaller than C_{12} (i.e., C_6, C_8, C_{10}), it led to the serendipitous isolation of an octanoyl-ACP dominant variant (R3.M4). The discovery of a C_8 mutant R3.M4 and the higher commercial value of octanoic acid[4] prompted a final round of computational predictions focusing on C_8 composition only.

In the fourth round (R4), the IPRO imposed design constraints were similar to that of rounds R1−R3, except that dodecanoyl-ACP (the "desired substrate") was replaced with octanoyl-ACP, while tetradecanoyl-ACP (the "undesired substrate") was replaced with dodecanoyl-ACP. These changes were made to drive more aggressive mutagenesis toward C_8 preference. Of the 18 R4 variants tested, all were active, 16 maintained WT FFA production levels (all except R4.M11 and R4.M12), 13 improved the C_8 mole fraction, but none produced more octanoic acid than R3.M4. The reduced production levels for R4.M11 and R4.M12 could be explained by the rearrangement of aromatic side chains in the binding crevice (i.e., R108F, F139, Y145, and Y145F), relative to the active R4 variants, which may have disrupted the hydrophobicity of the binding crevice.

Through four rounds of the Design-Build-Test-Learn paradigm, 54 variants were tested, 43 were active, 25 maintained WT production levels (23 in R3 and R4), 3 improved the C_{12} mole fraction, and 27 increased the C_8 composition (see Tables S1 and S2 in the Supporting Information). All 3 C_{12}-specific variants and all but 6 of the 27 C_8-specific variants (R1.M2, R2.M2, R2.M3, R2.M4, R3.M3, and R4.M12) maintained or exceeded WT production levels. The success rate of computational enzyme design varies drastically, based on the procedures employed, the system studied, and the ambition of the (re)design (published values can range from as low as 7% to as high as 78%), and the

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

frequency of favorable outcomes from IPRO in this study (24% for R1−R4, 43% for R3 and R4) is consistent with these previously reported values.[47,48,54−57] The variant with the highest C_{12} mole fraction (48% ± 8%, a 1.8-fold improvement over WT), R3.M1, consisted of three mutations: S122K, Y145K, and L146K. Despite attempting to target octanoic acid production in R4, the variant with the highest C_8 mole fraction was R3.M4 (50% ± 3%, a 10-fold improvement over WT), containing mutations M141L, Y145K, and L146K. The FFA profiles for the top C_{12}- and C_8-specific variants are summarized in Figure 3.

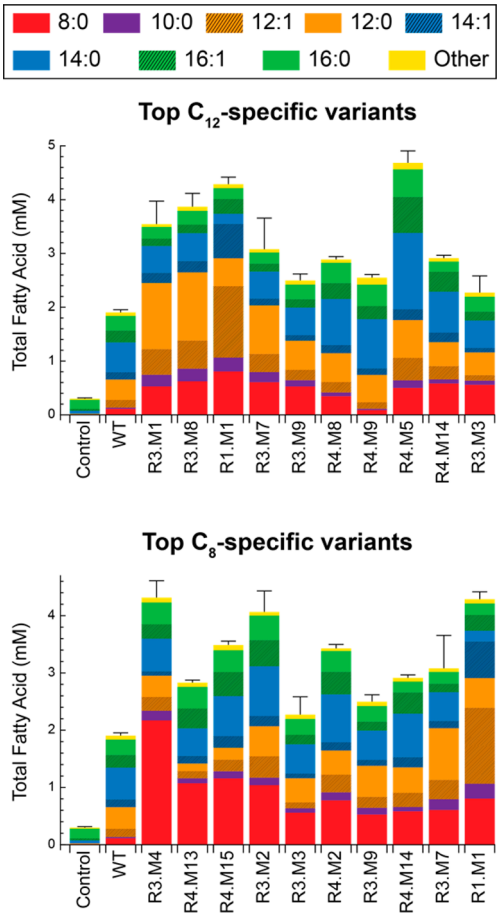


Figure 3. Fatty acid production profiles for the most C_{12}-specific and C_8-specific computationally designed variants. FFA titers are shown as bars, where error bars indicate total FFA standard deviation. Profiles of uninduced cells (Control) and wild-type 'TesA (WT) are provided for reference. All profiles are listed in Table S1.

Computation-Guided Design Outperforms Random Mutagenesis. Several of our best performing variants have a small number of mutations that may have been recovered from classical random mutagenesis approaches. Therefore, we created a small library of randomly mutated 'TesA variants by error-prone PCR mutagenesis and screened for changes in product profile. The purpose of this library was to provide a negative control for testing that a library not directed by IPRO would not achieve the same level of success. The FFA profile of 61 *E. coli* cultures harboring expression vectors for unique 'TesA variants was measured (N = 1). Of the 61 random mutants (RMs) screened, 46 were active (i.e., ≥ 240 μM), and 20 maintained WT FFA production levels (i.e., ≥1750 μM). The best dodecanoic acid producing RM (RM.M39, 44% ± 6%) demonstrated a comparable C_{12} composition to the top computationally predicted variant (R3.M1, 48% ± 8%) but at the expense of a substantial reduction in total FFA titer (p < 0.05, 80% of R3.M1). The best octanoic acid producing RM (RM.M29, 21.9%) produced a lower fraction of C_8 than R3.M4 (50% ± 3%, a 44% reduction) and displayed only 58% of the total activity of R3.M4 (see Table S3 in the Supporting Information). A comparison of these results shows that IPRO-guided mutagenesis generated more hits, more active mutants, and better leads than a library of similar size made through random mutagenesis.

Analysis of Successful 'TesA Redesigns. The best C_{12}-producing variants were dominated by three mutations: S122K, Y145K, and L146K. Mutation S122K (R3.RD3, Table S1) alone was sufficient to shift the C_{12} fraction to 35% of total FFAs, equal to the best C_{12}-producing mutant R3.M1. A nonpolar mutation at the same position (S122L, R3.RD4) had a similar but less pronounced shift toward C_{12} at the expense of C_{14}, indicating that S122 is an important residue in the active site (see Figure S1). The additional mutations in R3.M1 (Y145K and L146K—equivalent to the R3.M7 mutation) also reduced the long-chain composition but produced a higher fraction of C_8 and lowered total activity. Almost all of the top C_8-producing variants contained a mutation at Y145 with lysine or phenylalanine as the dominant substituents. The best C_8-producing mutant (R3.M4) contained Y145K and L146K mutations, as well as a M141L, which, by itself (R3.M2), was able to dramatically increase the C_8 composition. The Y145K mutation drastically increased the fraction of unsaturated products in the C_{12} and C_{14} chain lengths.

Crystal Structures and Simulations Show That Hydrophobic Interactions Govern Specificity. Crystal structures of WT 'TesA and R3.M4 bound to octanoic acid were solved to confirm the structures predicted by IPRO and facilitate analysis of helpful mutations. The WT crystal structure was very similar to the previously published structure[20] with an all-atom root-mean-square deviation (RMSD) of 1.1 Å. The X-ray resolved structures corroborated the structures predicted by our computational methods, quantified by an all-atom RMSD of 1.6 Å for R3.M4:octanoyl-ACP (Figure 4). The structural differences in $loop_{111-120}$ may have affected IPRO's ability to accept or reject mutations. However, the strong structural similarity between the crystallized and modeled structure at positions M141L, Y145K, and L146K indicate that the conformation of these side chains is favorable, despite the movement of $loop_{111-120}$. Comparison of the WT and R3.M4 structures reveal that the largest differences (RMSD ≥ 2.5 Å) occur at (i) the mutated positions (M141L, Y145K, L146K), (ii) the flexible regions as suggested by molecular dynamics (MD)-derived *B*-factors (L11−A19, Q32−S33, G44−D45, N73, I107−R115, and D153−I156), (iii) the solvent-exposed residues adjacent to these flexible loops (i.e., H157 and R160), and (iv) the C-terminus (i.e., L177; see Figure 5). While structural differences in the flexible regions are possibly artifacts due to high residue mobility, the structural differences at the mutated positions provide insight into enzyme specificity determinants. The M141L side chain extends laterally toward the ω-1 carbon of the FFA, yielding an attractive dispersion force (Figure 6A). Y145K partially occludes the binding crevice by forming a barrier between the ω-1 carbon of the FFA and solvent. L146K forms a salt bridge with E143 that may stabilize the conformation of Y145K but does not directly influence

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

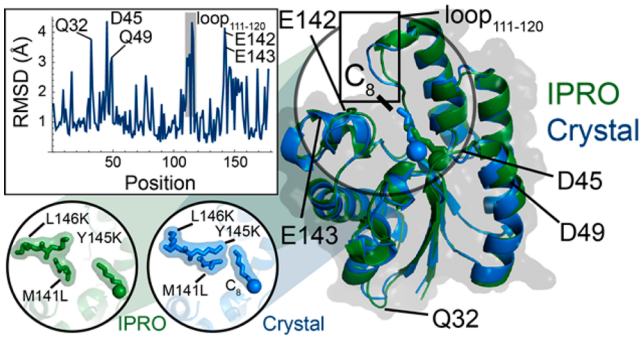


Figure 4. Comparison of computationally predicted and crystallized structure of R3.M4. The IPRO-derived structure (green) and crystallized structure at pH 5.0 (blue) of R3.M4 are shown as cartoons bound to octanoyl-ACP (truncated at the thioester bond) and octanoic acid (C_8), respectively. The carbonyl oxygen of each FFA structure is shown as a sphere. Regions of relatively large structural differences (RMSD ≥ 3.0 Å) are shown in a plot of the root-mean-square deviation (RMSD) versus position and annotated. Design positions (141, 145, and 146) are shown in the bottom left inset, where each residue is shown by sticks and a transparent surface. Hydrogen atoms are excluded from the IPRO-predicted structure.

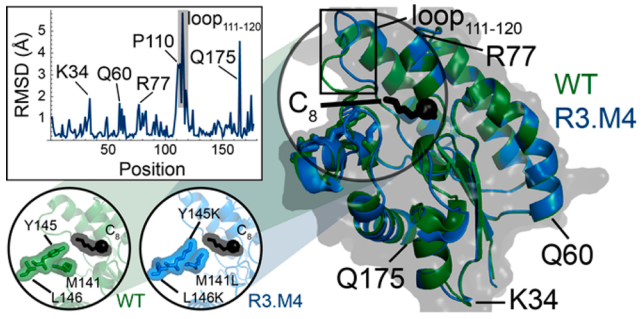


Figure 5. Comparison of WT and R3.M4 crystal structures. Both WT (green) and R3.M4 (blue, pH 5.0) are shown as cartoons to the right of the figure bound to octanoic acid (C_8). The carbonyl oxygen of each FFA structure is shown as a sphere. Regions with relatively large changes (RMSD ≥ 1.5 Å) are annotated. Quantification of the structural differences is shown with the inset plot of RMSD versus sequence position. The structural differences between the residues at positions 141, 145, and 146 (the mutated positions in R3.M4) are depicted in the bottom left inset. Here, each of the three residues, as well as octanoic acid, is represented as sticks with a transparent surface.

binding. Because the catalytic machinery of mutant R3.M4 (and all other active variants) includes an oxyanion hole that necessitates a negative charge for binding, R3.M4 in complex with octanoic acid (pK_a 4.9) was also crystallized at a higher pH to ensure binding site occupancy. The additional experiments included crystallization at pH 7.5 and crystallization at pH 5.0, followed by additional octanoic acid buffered at pH 7.5. These additional experiments revealed substantial differences (RMSD ≥ 2.5 Å) at R16 (2.534 Å), K34 (3.024 Å), A111 (4.362 Å), Y113 (9.311 Å), G114 (3.536 Å), and L177 (3.723 Å; see Figure S2 in the Supporting Information). These highly flexible regions indicate that their movement is highly sensitive to changes in hydrophobicity and their motion may be essential for catalytic turnover. The structure of N112 was not solved for R3.M4 at pH 5.0, because of low-resolution electron density maps.

MD simulations were used to elucidate the mechanisms by which FFA composition is controlled. Seven total trajectories were analyzed using MD. These included WT (bound to C_{14}, C_{12}, or C_8 acyl-ACPs), R3.M1 (bound to C_{14} or C_{12}), and R3.M4 (bound to C_{12} or C_8). The MD results revealed an alternate binding mode nearby T46, S47, N73, and R77 that is only assumed for preferred chain lengths in the WT and R3.M4 trajectories. This binding mode is likely not observed for R3.M1:dodecanoyl-ACP, because of insufficient conformational sampling. In addition to the alternate binding mode, MD analysis revealed that enzymes bound to acyl-ACPs beyond their preferred chain lengths have a deformed loop between G75 and Q80. With the exception of R3.M4:octanoyl-ACP and WT:octanoyl-ACP, all other enzymes bound to an acyl-ACP equal to or just below (≤2 carbon atoms) its preferred chain length maintain a constant conformation of this loop (see Figure S3 in the Supporting Information). Finally, *B*-factors for each of the seven simulated enzymes were estimated and exhibited six regions with increased mobility. These regions were found from L11−A19, Q32−S33, G44−D45, N73, I107−R115, and D153−I156 ($B_i \geq \overline{B_i} + \sigma$).

Combining the MD and crystallography results with previously published investigations of 'TesA acyl-ACP specificity[20] support the theory that hydrophobic interactions formed between $loop_{75-80}$ (i.e., residues forming the loop between G75 and Q80), the acyl chain of acyl-ACP, and $loop_{111-120}$ govern substrate selectivity. Lo et al.[20] previously postulated that $loop_{75-80}$, described therein as the "switch loop" that is in one conformation (i.e., "on") when the acyl-ACP is bound and in another conformation (i.e., "off") when the substrate is not bound. The switch loop is dependent on the acyl chain length and stabilizes the enzyme:substrate complex during hydrolysis. Our MD simulations are consistent with these findings, as the switch loop is in the "on" position for variants bound to an acyl-ACP equal to or just less than (≤2 carbon atoms; i.e., WT bound to C_{12}) its preferred chain length. Prior studies[20,58] revealed that rigidity of the switch loop or the lack of a bound acyl-ACP force the switch loop into the "off" conformation. MD simulations for WT:octanoyl-ACP, R3.M1:tetradecanoyl-ACP, and R3.M4:dodecanoyl-ACP occupy intermediate states between the "on" and "off" conformations. The switch loop is in a completely different conformation for R3.M4:octanoyl-ACP, which is due to the increased hydrophobicity in the binding crevice near the C-1 terminus caused by M141L. For R3.M4:octanoyl-ACP, the movement of the switch loop toward $loop_{111-120}$ allows octanoyl-ACP to slide underneath the switch loop into the alternate binding cavity, which could be important for octanoic acid release. The pH-dependent conformation of $loop_{111-120}$ for R3.M4 from the crystallography experiments suggests that a change in protonation state of a titratable residue (i.e., an acidic or basic amino acid) causes a major conformational change, demonstrating the sensitivity of the hydrophobic region to electrostatics. At pH 7.5, the hydrophobic $loop_{111-120}$ extends away from the typically hydrophobic binding crevice toward the highly hydrophilic bulk solvent (Figure S2 in the Supporting Information), which is unexpected, since nonpolar substances typically aggregate (i.e., the hydrophobic effect). This movement of $loop_{111-120}$ is postulated to be due to a newly charged residue in the binding crevice, weakening the binding crevice's hydrophobic environment. This observation is consistent with octanoic acid serving as the titratable residue. While M141L increases the hydrophobicity near the C-8 atom of R3.M4, Y145K limits the hydrolysis of longer acyl-ACPs by reducing binding crevice hydrophobicity near the ω-1 atom (i.e., C-10, C-12, C-14)

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

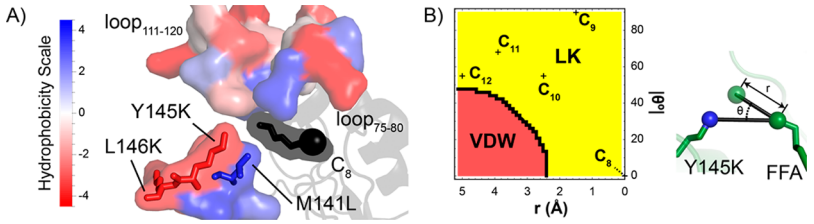


Figure 6. Hydrophobicity drives substrate specificity, as demonstrated by (A) R3.M4 crystal structure and (B) a two-atom model for the Y145K mutation. In panel (A), crystallized R3.M4 (pH 5.0) is bound to octanoic acid (C_8). The two critical hydrophobic loops ($loop_{75-80}$ and $loop_{111-120}$), as well as the residues nearby the design positions, are shown as molecular surfaces and colored according to their hydrophobicity score. The FFA is represented as sticks (with its carbonyl oxygen shown as a sphere) and a black molecular surface, as its hydrophobicity score is unknown. In panel (B), a two-atom model is shown to examine the ability of Y145K to eliminate binding to tetradecanoyl-ACP. The ε-amino nitrogen of Y145K (blue sphere) and the ω-1 carbon of FFA (green sphere) constitute the two-atom model. The model shows the effect of extending the acyl chain, which changes the hypothetical position of the ω-1 atom (transparent green sphere). The parameter r represents the interatomic distance between the C_8 ω-1 atom and the hypothetical ω-1 atom. θ represents the angle between the ε-amino nitrogen, the C_8 ω-1 atom, and the hypothetical ω-1 atom. At (r = 0 Å, θ = 0°), the ω-1 atom occupies the position of the crystallized ω-1 C_8 atom. The energy term with the largest contribution toward the interaction energy between the ε-amino nitrogen and the hypothetical ω-1 atom at a given r and θ is provided in the contour plot (red = van der Waals energy, yellow = Lazaridis−Karplus solvation energy). The positions of the hypothetical ω-1 atoms using ideal FFA geometry (C−C bond length of 1.54 Å, 109.5° angle, 180° dihedral angle) are also labeled in the contour plot. C_{13} (r = 6.4 Å, θ = 62.8°) and C_{14} (r = 7.5 Å, θ = 54.9°) are beyond the boundaries of the contour plot. Molecular structures were generated using PyMOL.

thereby disrupting the conformation of the switch loop. A simple two-atom model involving the FFA ω-1 carbon and the ε-amino nitrogen (i.e., the side chain nitrogen) atom of Y145K had pairwise energies calculated and demonstrated that the Lazaridis−Karplus solvation term, a computationally accessible proxy for hydrophobicity, governs unfavorable interactions with the ω-1 carbon (Figure 6B). The two-atom model illustrates that solvation energy dominates the phase space with the exception of the van der Waals region. The ω-1 carbon would not be expected to occupy the van der Waals region, because the energy is much more unfavorable (i.e., positive) at these close distances ($E \approx r^{-12}$ for van der Waals, $E \approx \exp(-r^2)$ for Lazaridis−Karplus solvation). Longer FFA chain lengths extend closer to the ε-amino nitrogen, exacerbating the repulsive solvation energy. Adding additional carbon atoms to the point where the ω-1 is further from the ε-amino nitrogen will only add to the total repulsive energy, although the marginal cost will decrease with each additional carbon. Finally, the large degree of switch loop and $loop_{111-120}$ mobility could explain the broad substrate specificity that is typically observed for 'TesA.

$\Delta\Delta G$ ($\Delta G_{Variant} - \Delta G_{WT}$) values estimated from the computations exhibited good agreement with experimentally derived values (see the Supporting Information). A Pearson correlation coefficient of 0.6 ± 0.2 was found for R3 and 0.43 ± 0.07 for R4. These correlation coefficients are consistent with earlier computational studies.[24,59,60] Notably, the regions with elevated *B*-factors from the MD trajectories ($loop_{11-19}$, $loop_{32-33}$, $loop_{44-45}$, $loop_{73}$, $loop_{107-115}$, and $loop_{153-156}$) overlap with the regions with elevated *B*-factors from the crystallography experiments ($loop_{30-35}$, $loop_{59-62}$, $loop_{98-100}$, $loop_{111-115}$, and $loop_{153}$) and other published[20,58] crystallographic structures ($loop_{31-35}$, $loop_{75-80}$, and $loop_{111-120}$).

***In Vitro* Assays of WT 'TesA and C_8-Specific R3.M4 Confirm *In Vivo* Results.** 'TesA can catalyze hydrolysis of both acyl-CoA and acyl-ACP substrates. Given the relative availability of these substrates, we compared the kinetic activity of WT 'TesA and R3.M4 on acyl-CoAs ranging from 6 to 16 carbons in length. We monitored reaction progress by tracking the abundance of free CoA released by hydrolysis. The highest *in vitro* WT activity (Figure 7A) was observed for C_{12}-CoA, C_{14}-CoA, and C_{16}-CoA, consistent with the *in vivo* data for the release of FFAs from acyl-ACPs (see Figure 3). In contrast, R3.M4 showed a significant increase in activity on C_8-CoA compared to WT, which is consistent with the observed *in vivo* production of octanoic acid, and a modest decrease in activity on C_{12}-CoA and C_{14}-CoA activity. For both enzymes, we observed an unexpected decrease in activity on hexadecanoyl-CoA (C_{16}-CoA) beyond a threshold concentration. Interestingly, in the R3.M4 mutant, the inhibitory effect of C_{16}-CoA is exacerbated and C_{14}-CoA also shows inhibition (not seen on WT 'TesA). Given the linear reaction progress curves that we observed, we suspected that the enzymes were substrate-inhibited. Therefore, we performed assays with both C_8-CoA and the inhibitory CoA species (see Figures 7B−E). Competitive activity assays were performed at a constant concentration of C_8-CoA (50 μM) and variable concentrations of C_{14}-CoA (Figures 7B for WT and 7C for R3.M4) and C_{16}-CoA (Figures 7D for WT and 7E for R3.M4). In all cases, production of free CoA was inhibited by C_{14}-CoA and C_{16}-CoA in a concentration-dependent manner consistent with the original assay in Figure 7A.

CONCLUSIONS

The potential of the IPRO algorithm to aid in protein engineering efforts was demonstrated using a Design-Build-Test-Learn approach to alter the substrate preference of 'TesA. Our approach leverages computational protein design procedures to achieve successful experimental redesign beyond what has been achievable so far,[15−17] yielding two top variants. One (R3.M1) produced 48% ± 8% C_{12} composition, a 1.8-fold improvement over WT, while maintaining native production levels. Despite a preference for C_{14} production in WT 'TesA, three amino acid substitutions constitute R3.M1, which is the third-most C_{12}-specific thioesterase known to date (see Table 1). Similarly, R3.M4 produced 50% ± 3% C_8 composition, a 10-fold improvement, while maintaining WT production levels. R3.M4 is the tenth-most C_8-specific thioesterase described to date. In total, the Design-Build-Test-Learn paradigm yielded 3 and 21 variants with significant ($p < 0.05$) improvements in C_{12} and C_8 mole fraction, respectively, while maintaining WT production levels.

Despite sampling a similar library size, random mutagenesis yielded fewer active mutants than the computationally guided library. In addition, the random library produced two fewer C_{12}-

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

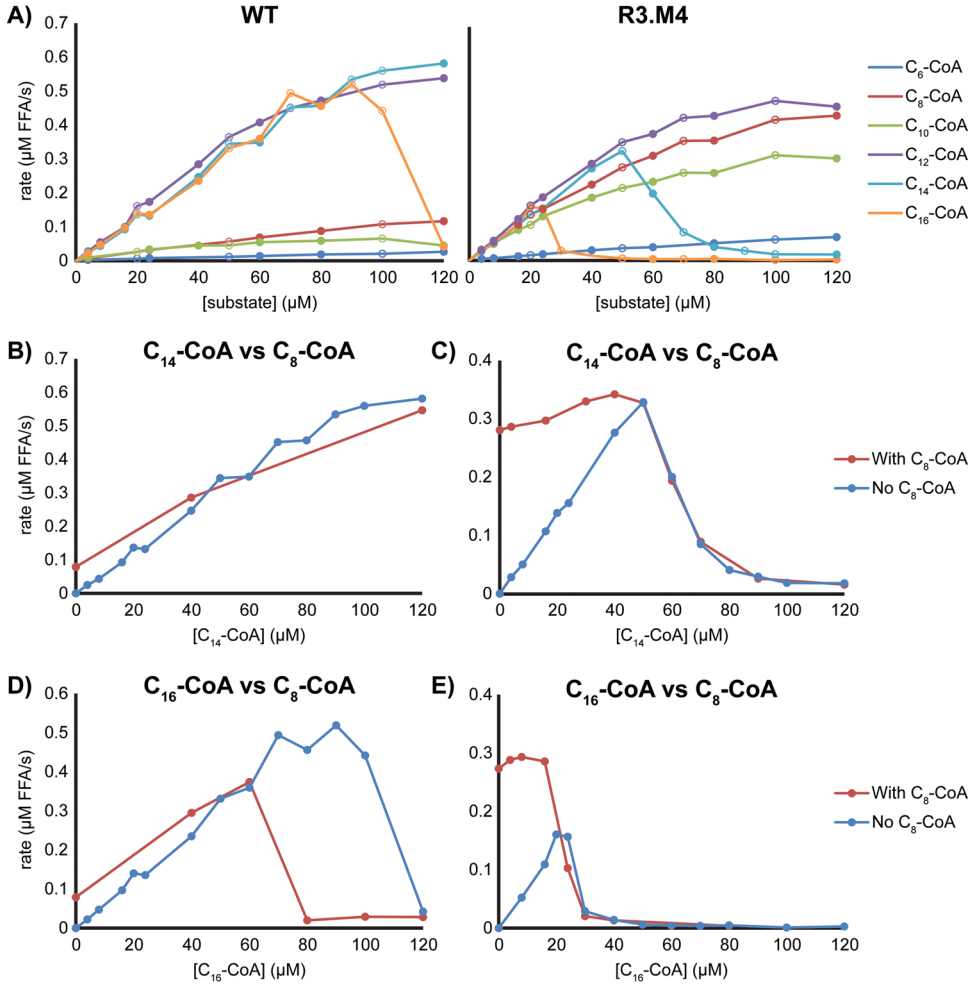


Figure 7. Enzymatic assays of (A) WT 'TesA and R3.M4 confirm the mutant's increase in specificity for C_8 species, and competitive activity assays (B–E) show WT 'TesA and R3.M4 activity on C_8-CoA, as a function of (B and C) C_{14}-CoA concentration and (D and E) C_{16}-CoA concentration. In panel (A), the activity of WT 'TesA (left) and R3.M4 (right), as a function of substrate concentration, is shown for six different acyl-CoA substrates. Open and closed circles indicate measurements taken in separate days. The competitive activity assays (panels (B)–(E)) were performed to see the effect that an increase in C_{14}-CoA and C_{16}-CoA would have on the C_8–CoA activity of the enzymes. Competitive activity assays were performed at a constant C_8-CoA concentration of 50 μM and variable concentrations of C_{14}-CoA ((B) for WT and (C) for R3.M4) and C_{16}-CoA (panel (D) for WT and panel (E) for R3.M4). In all cases, the activity of 'TesA was impacted by the C_{14}-CoA and C_{16}-CoA in a concentration-dependent manner consistent with the original assay in panel (A).

Table 1. Top Ten C_{12}-Specific and C_8-Specific Thioesterases Published to Date[a]

	Dodecanoic Acid (C_{12})			Octanoic Acid (C_8)		
rank	name	composition	source	name	composition	source
	WT	28 ± 1	this study	WT	5.9 ± 0.3	this study
1	Q41635	76.9[b,c]	13	AAC49179	97.5 ± 0.2	10
2	rTE15	49.0[b]	16	AAC72882	93.5[b]	16
3	R3.M1	49 ± 8	this study	EEI82564	87 ± 2	10
4	R3.RD3	48 ± 10	this study	ABG82470	70 ± 4	10
5	R3.M8	46 ± 4	this study	CAD63310	68.0 ± 0.8	10
6	RM.M39	44 ± 6	this study	CvB2MT40	61.4[b]	16
7	R1.M1	43 ± 2	this study	ABJ63754	55.5 ± 0.7	10
8	AEM72521	40 ± 2	10	AEM72522	52 ± 6	10
b9	R3.M7	40 ± 11	this study	TEGm258	50.4[b]	16
10	R4.RD11	37.7 ± 0.6	this study	R3.M4	50 ± 3	this study

[a]Data were collected from Table S1 and sorted by dodecanoic acid (either saturated or unsaturated) and octanoic acid mole fractions in the product pool. WT compositions were provided as a reference. "Composition" refers to mole fraction of the corresponding FFA. Names match those provided in Table S1. The most C_{12}-specific and C_8-specific 'TesA variants found in R1-R4 correspond to the third most C_{12}-specific and tenth most C_8-specific thioesterases identified to date, respectively. [b]Standard deviation not provided [c]Data approximated using image processing software

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837–3849

specific variants with native production levels and 20 fewer C_8-specific variants with at least WT productivity. Unlike random mutagenesis, variants from R3 and R4 achieved high levels of activity by directly enforcing catalytic contacts through imposed restraints and preserving conserved amino acids. Furthermore, R3 and R4 variants systematically tailored the binding crevice environment to generate hydrophobic clusters between $loop_{111-120}$, the acyl-ACP, and the switch loop ($loop_{75-80}$). MD and crystallography results from this work and elsewhere[20,58] suggest that this hydrophobic packing is essential for enzyme functionality. Whereas the *in silico* method directly accounts for hydrophobicity (albeit only approximately) through the Lazaridis−Karplus solvation energy term, random mutagenesis techniques can only improve binding site hydrophobicity by chance alone.

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

MATERIALS AND METHODS

'TesA Model Construction. The structure of 'TesA was derived from PDB 1U8U, where it is in complex with octanoic acid.[20] The acyl-ACP structures were derived from PDB 2FAE, where decanoyl-ACP is held in an internal binding cavity.[49] Other acyl-ACP structures, such as hexanoyl-ACP (from PDB 2FAC)[49] and heptanoyl-ACP (from PDB 2FAD),[49] are similar to decanoyl-ACP, as demonstrated by respective all-atom RMSDs of 1.1 and 1.1 Å. In order to dock octanoyl-ACP with 'TesA, the acyl chain was systematically rotated about the phosphopantetheine linker and superimposed with the bound octanoic acid in 1U8U. The rotation that led to the lowest root-mean-square deviation (RMSD) was energy-minimized within CHARMM34.[61] Acyl-ACPs with different chain lengths were adapted from this initial complex by either deleting atoms or adding atoms using CHARMM's internal coordinate system. Lazaridis−Karplus solvation files and CHARMM input files were constructed using published parameters for lipids and proteins. The constructed topology and parameter files were in close agreement with CGenFF-derived parameters.[62,63]

Structure-Based Redesign and Analysis. All computationally predicted mutants were identified using multiple IPRO trajectories that each ran for 1000 iterations without ensemble structure refinements.[24] The primary constraint of each trajectory was to eliminate binding to larger acyl-ACPs (C_{14} for Rounds R1−R3, C_{12} for Round R4) with a secondary constraint to improve binding to the shorter acyl-ACPs (C_{12} for Rounds R1−R3, C_8 for Round R4). Round R4 C_8/C_{14} and C_8/C_{12} fraction ratios correlate with $r = 0.88 \pm 0.02$, alluding to a similar repulsive force for C_{12} and C_{14} binding (see Figure S4 in the Supporting Information). Design position selection is described in the "Design Position Selection" subsection. Restraints were imposed to ensure that the intermolecular catalytic distances (±0.2 Å) were maintained (i.e., S10, G44, N73, D154, and H157). All other IPRO parameters were set to their standard values, and calculations were run on the Lion-XF system at Penn State University. Error propagation was performed manually and replicated using the Python uncertainties module.[64] Statistical differences were calculated using Welch's *t*-test between the WT and mutant FFA profiles. For a given enzyme-FFA complex, the interaction energy is found using $IE = G_{Enz\text{-}FFA,min} - G_{Enz} - G_{FFA}$. Mutants were sorted by the interaction energy difference between the short-chain FFA and C_{14} ($\Delta IE = IE_{C12,C8} - IE_{C14,C12}$). Therefore, for a given round, Mutant 1 (smallest ΔIE) would be expected to show the biggest change in specificity.

Design Position Selection. For R1, 'TesA residues were sorted by distance to carbon atoms in the acyl group of the 'TesA:tetradecanoyl-ACP complex (i.e., C-1 through C-14). The minimum interatomic distance between the residue's heavy atoms (i.e., not hydrogen atoms) and the acyl carbon atoms was used for sorting. In R1, residues constituting the catalytic triad (i.e., S10, D154, H157) and the oxyanion hole (i.e., S10, G44, N73) were not considered during design position selection. At the time design positions for R1 were designated, IPRO was unable to handle mutations from proline so these residues were not considered during the selection of design positions.[24] L109 was considered to be important for 'TesA functionality and was also removed from consideration as a design position.[20] Residues that were near the undesired end of the acyl group (i.e., near C-1 instead of ω-1) were no longer regarded as potential design positions. Residues considered "near C-1" were those that contained a heavy atom within 4.5 Å of the thioester S atom. A final set of residues that were strongly oriented away from the ω-1 terminus of the acyl moiety were also no longer considered as potential design positions. Residues "strongly oriented away from the ω-1 terminus" were those whose Cα atom was more than 0.75 Å closer to the ω-1 atom than the Cβ atom (glycines not considered). The sorted residues are provided in Table S4 in the Supporting Information with any exceptions annotated. The eight nearest residues that were not filtered out were selected as design positions. The set of design positions used for R1 were L11, G72, L76, I107, R108, A111, F139, and Y145.

For R2, 'TesA residues were once again sorted by distance to carbon atoms in the acyl group of the 'TesA:tetradecanoyl-ACP complex. Residues constituting the catalytic triad and oxyanionic hole, proline residues, residues near the C-1 terminus of the acyl moiety, and L109 were once again removed from consideration, as described for the R1 design position selection procedure. In addition, residues belonging to the three flexible loops (i.e., $loop_{31-35}$, $loop_{75-80}$, $loop_{111-120}$) were no longer recognized as prospective design positions for the same reason that L109 was removed from consideration.[20] Finally, I107 and R108 were removed as possible design positions, because these positions invariably mutated to lysines in R1 and all instances of I107K and R108K abolished catalytic activity (see Figure 2, as well as Table S2). The set of design positions (i.e., the eight nearest residues that were not filtered out) employed for R2 included L11, G72, F139, M141, E142, Y145, G155, and I156.

For R3 and R4, a very different approach was used when compared to R1 and R2. Instead of sorting residues by distance

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

to carbon atoms in the acyl group of 'TesA:tetradecanoyl-ACP, residues were sorted by distance between the residue's Cβ atom and the ω-1 atom of dodecanoyl-ACP from the 'TesA:dodecanoyl-ACP complex. In lieu of the Cβ atom, Cα was used to calculate interatomic distances for glycine residues, and Cγ was used for H180, because Cα and Cβ were not part of the solved crystal structure.[20] Unlike R1 and R2, only residues that were aligned to gaps or had ≥40% sequence conservation were not considered as candidate design positions. Sequence alignment was performed using the conserved domain database, where 81 members (including 'TesA) of the lysophospholipase L1-like subgroup from the SGNH-hydrolase superfamily were found.[65] The sorted residues are provided in Table S4 in the Supporting Information. Residues aligned to gaps or conserved residues are also noted in Table S4. The final set of design positions (i.e., the eight nearest residues that were not filtered out) used for R3 and R4 was I107, R108, L109, S122, M141, E142, Y145, and L146.

Scoring Function Reweighting. A new set of weights for the IPRO scoring function was found that approximately doubles the native rotamer recovery, relative to the unmodified scoring function. The scoring function was modified using a symmetric logistic regression within Weka, a collection of machine-learning algorithms.[66] A dataset of native and non-native rotamers was collected from the Top8000 database, which is a dataset of 8000 high-resolution (<2 Å), quality-filtered (<2.0 MolProbity score[67]), nonhomologous (<70% identity) protein structures.[53] Of these structures, 50 were randomly selected for use with the machine-learning training set. A separate set of 80 structures was randomly selected to validate the results. From these 130 structures, the native rotamer was found by finding the rotamer (of the same amino acid type) with the lowest RMSD to the crystallized side chain. The van der Waals, electrostatic, and Lazaridis−Karplus solvation energies then were calculated for each rotamer (regardless of amino acid type), as well as a binary indicator as to whether this was the native rotamer at the position or not. The data were separated for residues at the protein surface (≤20 C^{β} atoms within 10 Å) and within the core of the protein (>20 C^{β} atoms within 10 Å) using a distance-based metric developed by Kuhlman and Baker.[68] Finally, since the number of non-native rotamers heavily outweighed the number of native rotamers, non-native rotamers were randomly removed until there was an ∼60:40 split of non-native:native rotamers. The rotamer data were used to determine the set of weights that can optimally classify a rotamer as native or non-native. The updated energy equation for the core residues is $E = 0.04E_{VDW} + 0.02E_{Elec} + 0.16E_{LK}$. For the surface residues, the updated energy equation is $E = 0.03E_{VDW} + 0.01E_{Elec} + 0.09E_{LK}$. The inaccuracy of the original scoring function that used equally weighted energy terms (i.e., $E = 0.01E_{VDW} + 0.01E_{Elec} + 0.01E_{LK}$) stemmed mostly from the Lazaridis−Karplus implicit solvation energy term.

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

DNA Synthesis and 'TesA Variant Construction. All mutants were created starting with WT '*tesA* gene cloned into a pBAD18 plasmid[70] to link 'TesA expression to the presence of L-arabinose. R1 mutants were constructed using Agilent Technologies QuickChange II site directed mutagenesis kit following the given protocol. For R2−R4, all mutants were constructed using Gibson assembly strategies with primers containing the desired mutations in the 5′ tails. All cloning was performed in *E. coli* DH5α strain.

'TesA in pBad18 was randomly mutagenized using Gene Morph II random mutagenesis kit following the kit protocol to make a library of 61 mutants with a mutation rate of 1.8 amino acids per gene (Table S2). Primers were designed to include the start and stop codons to ensure keeping those positions of the mutants invariant.

Bacterial Culturing and Fatty Acid Production. FFA production was assayed from small batch cultures (5−50 mL) of *E. coli* strain RL08ara (K-12 MG1655 Δ*fadD* Δ*araBAD* ΔaraFGH $\Phi(\Delta araEp P_{CP18} - araE)^3$) harboring each thioesterase expression vector. Three single colonies of each mutant were grown overnight on LB media containing 100 mg/L of ampicillin. Overnight cultures of each strain were diluted 1:100 into 25 mL of LB media containing 100 mg/L of ampicillin and 0.4% w/v glycerol in a 250 mL baffled shake flask and grown at 37 °C and 250 rpm. When the OD_{600} reached 0.2−0.3, cultures were induced with 0.2% w/v L-arabinose and shaken for 24 h. All mutants were tested in triplicate and error bars represent the standard error of the measurements.

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

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

Protein Expression and Purification of WT 'TesA. The WT *'tesA* gene was cloned into pET28t, a vector previously modified to contain a TEV protease site, rather than a thrombin cleavage site between 'TesA and a N-terminal polyhistidine tag.[71] When translated, the modified enzyme was fused to the following peptide: MGSSHHHHHHSSENLYFQGGGG. The pET28t-*tesA* plasmid was used to transform *E. coli* Rosetta2-(DE3) cells (Novagen). Cultures were grown at 37 °C with shaking in lysogeny broth supplemented with 50 mg/L kanamycin and 50 mg/L chloramphenicol until the OD_{600} reached 0.8. Flasks were cooled in an ice bath, induced with 1 mM isopropyl-β-D-thiogalactopyranoside, and incubated overnight at 21 °C. Cells were harvested by centrifugation and frozen as pellets in liquid nitrogen. The frozen cell pellets were sonicated on ice in a lysis buffer composed of 50 mM sodium phosphate, 20 mM imidazole, 10% glycerol, and 300 mM NaCl (pH 8.0). The lysate was cleared by centrifugation, and 'TesA was purified at 4 °C utilizing Ni-nitrilotriacetic acid resin (Qiagen), according to the manufacturer's instructions. TEV protease was added in a 1:20 molar ratio to the pooled protein solution and subsequently dialyzed against 50 mM sodium phosphate, 300 mM NaCl, and 20 imidazole (pH 8.0) at 4 °C for 36 h. Both the TEV protease and the uncleaved protein were removed by passage over a Ni-nitrilotriacetic acid resin. The cleaved protein was collected and dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl and concentrated to ~35 g/L based on an extinction coefficient of 0.62 L g^{-1} cm^{-1}.

Crystallization and Structural Analysis of WT 'TesA. Crystallization conditions for 'TesA were surveyed by the hanging drop method of vapor diffusion using a laboratory-based sparse matrix screen. The enzyme was initially tested either in the presence or absence of 3 mM C_8 FFA. Crystals were subsequently grown from 22%−26% poly(ethylene glycol) (PEG) 5000 with 100 mM Homo-PIPES buffer (pH 5.0). The protein solution used contained 3 mM C_8 FFA. Crystals belonged to the monoclinic space group $P2_1$ with unit-cell dimensions of a = 40.9 Å, b = 82.1 Å, c = 53.9 Å, and β = 90.4°. The asymmetric unit was comprised of two monomers. Prior to X-ray data collection at 100 K, the crystals were transferred to a cryoprotectant solution composed of 30% PEG 5000, 250 mM NaCl, 3 mM C_8 FFA, and 13% ethylene glycol with Homo-PIPES buffer (pH 5.0). An X-ray dataset was collected with a Bruker AXS Platinum-135 CCD detector using the PROTEUM software suite (Bruker AXS, Inc.) The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. Data were processed with SAINT and scaled with SADABS (Bruker AXS, Inc.). X-ray data collection statistics are listed in Table S5 in the Supporting Information. The structure of WT 'TesA was solved via molecular replacement using the software package PHASER[72] and the PDB 1U8U as the search model.[20] Model refinement with REFMAC[73] and manual model building with COOT[74,75] reduced the overall R-factor to 18.8% at a resolution of 1.65 Å. Refinement statistics are presented in Table S6 in the Supporting Information.

Crystallization and Structural Analysis of R3.M4. The R3.M4 *'tesA* gene was subcloned, expressed, and purified as described for WT. Crystals were obtained at both pH 5.0 and pH 7.5. Those obtained at pH 5.0 were grown from 20%−25% PEG 5000 with 100 mM Homo-PIPES buffer (protein solution contained 3 mM C_8 FFA). The crystals were cryoprotected as described for WT 'TesA. Crystals belonged to the monoclinic space group $P2_1$ with unit-cell dimensions of a = 40.7 Å, b = 55.2 Å, c = 42.3 Å, and β = 105.2° with a single monomer in the asymmetric unit. R3.M4 crystals obtained at pH 7.5 were grown from 24%−28% PEG with 100 mM HEPES buffer (pH 7.5). Again, the protein solution contained 3 mM C_8 FFA. These crystals were isomorphous to those obtained at pH 5.0 and were cryoprotected with a solution composed of 32% PEG, 250 mM NaCl, 3 mM C_8 FFA, and 13% ethylene glycol with 100 mM HEPES buffer (pH 7.5). Given the concern that, at pH 5.0, the C_8 FFA would most likely not bind at full occupancy, these crystals were subsequently moved in a final experiment to solutions buffered at pH 7.5 that contained an additional 3 mM C_8 FFA. These "soaked" crystals were cryoprotected in a similar manner to those grown at pH 7.5. X-ray data from R3.M4 crystals obtained at pH 5.0, pH 7.5, and from the "soaked" crystals were collected as described for WT. The structure R3.M4 at pH 5.0 was solved by molecular replacement, using the WT model as the search probe whereas the structures either grown or soaked at pH 7.5 were solved via Fourier difference analyses. X-ray data collection statistics and model refinement statistics are provided in Tables S5 and S6, respectively.

Purification for Enzymatic Assays of WT 'TesA and R3.M4. The *'tesA* gene was subcloned as described for WT. Expression followed the same procedure, except the construct was transformed to BL21(DE3) cells. The frozen cell pellets were sonicated on ice in a lysis buffer composed of 50 mM sodium phosphate, 10 mM imidazole, and 300 mM NaCl (pH 8.0). The lysate was cleared by centrifugation, and 'TesA was purified at 20 °C utilizing Ni-nitrilotriacetic acid resin (Qiagen), according to the manufacturer's instructions. TEV protease was added in a 1:20 molar ratio to the pooled protein solution and subsequently dialyzed against 10 mM Tris-HCl (pH 7.5) at 4 °C for 18 h. Both the TEV protease and the uncleaved protein were removed by passage over a Ni-nitrilotriacetic acid resin. The cleaved protein was concentrated to 2.5 mL to be solvent-exchanged into 50 mM potassium phosphate (pH 7.0), and 30% glycerol using a PD-10 desalting column (GE), according to the manufacturer's instructions.

Enzymatic Assays of WT 'TesA and R3.M4. Enzymatic thioesterase assay was performed with WT 'TesA and R3.M4 to compare their activities on various chain lengths. The reaction conditions follow as per Shin et al.,[76] except the enzyme concentration used was 40 nM and substrate concentrations ranged from 0 to 120 μM for six saturated acyl-CoA substrates of 6−16 carbons (hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, dodecanoyl-CoA, tetradecanoyl-CoA, and hexadecanoyl-CoA). The assay tracks generation of free-CoA as 'TesA hydrolyses the thiol bond in the acyl-CoA. This hydrolysis is tracked by the increase of absorbance at 412 nm, because of the free-CoA dependent reduction of 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) present in the reaction mixture.[76] Absorbance at 412 nm was followed using a NanoDrop 2000c system (Thermo Scientific), with a path length of 10 mm, and measurements were taken for 2 min in 10 s intervals.

Molecular Dynamics. VMD was used to solvate enzyme−FFA complexes within a 12.0 Å water box with 0.17 M NaCl and contained ~49 000 atoms.[77] Each complex was minimized and slowly heated to 310 K and 1 atm over 7 ns using Langevin dynamics. Force-field parameters were identical to those used for the IPRO trajectories. Periodic boundary conditions were applied, and long-range electrostatic forces were considered using the particle mesh Ewald method. 40 ns production simulations were performed using NAMD over 30 nodes on

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

the Lion-XF cluster at Penn State University, using the NVE ensemble.[78]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.7b00408.

Comparison of computational and experimental $\Delta\Delta G$ values; structure of R3.M1; pH dependence of R3.M4 structure; structures derived from MD; comparison of C_8/C_{12} to C_8/C_{14} for R4; correlation between computational and experimental $\Delta\Delta G$ values; FFA production profiles for thioesterases expressed in *E. coli*; sequences of all thioesterase variants considered in this study; FFA production profiles for C_{12}-specific random variants; 'TesA residues sorted by distance to acyl-ACP used for design position selection; crystallography data collection statistics; and crystallography refinement statistics (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: (814) 863-9958. E-mail: costas@psu.edu.

ORCID

Costas D. Maranas: 0000-0002-1508-1398

Notes

The authors declare no competing financial interest.

[▽]These authors contributed equally to this work.

■ ACKNOWLEDGMENTS

Support for this work was provided by the National Science Foundation (Award Nos. CBET-0967062, CBET-1149678), the Advanced Research Projects Agency-Energy (Award No. DE-AR0000431), and the National Institutes of Health (NIH No. GM115921). N.J.H.-L. is the recipient of a NIH Chemistry-Biology Interface Training Program fellowship (No. T32 GM008505) and a Graduate Engineering Research Scholars fellowship from the UW−Madison College of Engineering.

■ REFERENCES

(1) Lennen, R. M.; Pfleger, B. F. *Curr. Opin. Biotechnol.* **2013**, *24*, 1044−1053.

(2) Steen, E. J.; Kang, Y. S.; Bokinsky, G.; Hu, Z. H.; Schirmer, A.; McClure, A.; del Cardayre, S. B.; Keasling, J. D. *Nature* **2010**, *463*, 559−562.

(3) Lennen, R. M.; Braden, D. J.; West, R. M.; Dumesic, J. A.; Pfleger, B. F. *Biotechnol. Bioeng.* **2010**, *106*, 193−202.

(4) Pfleger, B. F.; Gossing, M.; Nielsen, J. *Metab. Eng.* **2015**, *29*, 1−11.

(5) Knothe, G. *Energy Fuels* **2008**, *22*, 1358−1364.

(6) Howard, T. P.; Middelhaufe, S.; Moore, K.; Edner, C.; Kolak, D. M.; Taylor, G. N.; Parker, D. A.; Lee, R.; Smirnoff, N.; Aves, S. J.; Love, J. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7636−7641.

(7) Liao, J. C.; Mi, L.; Pontrelli, S.; Luo, S. *Nat. Rev. Microbiol.* **2016**, *14*, 288−304.

(8) Magnuson, K.; Jackowski, S.; Rock, C. O.; Cronan, J. E., Jr. *Microbiol. Rev.* **1993**, *57*, 522−542.

(9) Choi, Y. J.; Lee, S. Y. *Nature* **2013**, *502*, 571−574.

(10) Jing, F. Y.; Cantu, D. C.; Tvaruzkova, J.; Chipman, J. P.; Nikolau, B. J.; Yandeau-Nelson, M. D.; Reilly, P. J. *BMC Biochem.* **2011**, *12*, 44.

(11) Zhang, X.; Li, M.; Agrawal, A.; San, K. Y. *Metab. Eng.* **2011**, *13*, 713−722.

(12) Lu, X.; Vora, H.; Khosla, C. *Metab. Eng.* **2008**, *10*, 333−339.

(13) Voelker, T. A.; Davies, H. M. *J. Bacteriol.* **1994**, *176*, 7320−7327.

(14) Dormann, P.; Voelker, T. A.; Ohlrogge, J. B. *Arch. Biochem. Biophys.* **1995**, *316*, 612−618.

(15) Hom, L.; Trinh, N.; Alibhai, M. Methods and compositions related to thioesterase enzymes. U.S. Patent 2,010,154,293 (A1), June 24, 2010.

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

(17) Mayer, K. M.; Shanklin, J. *BMC Plant Biol.* **2007**, *7*, 1.

(18) Zhang, F. Z.; Carothers, J. M.; Keasling, J. D. *Nat. Biotechnol.* **2012**, *30*, 354−359.

(19) Liu, T.; Vora, H.; Khosla, C. *Metab. Eng.* **2010**, *12*, 378−386.

(20) Lo, Y.-C.; Lin, S.-C.; Shaw, J.-F.; Liaw, Y.-C. *Biochemistry* **2005**, *44*, 1971−1979.

(21) Bloom, J. D.; Meyer, M. M.; Meinhold, P.; Otey, C. R.; MacMillan, D.; Arnold, F. H. *Curr. Opin. Struct. Biol.* **2005**, *15*, 447−452.

(22) Packer, M. S.; Liu, D. R. *Nat. Rev. Genet.* **2015**, *16*, 379−394.

(23) Gajewski, J.; Pavlovic, R.; Fischer, M.; Boles, E.; Grininger, M. *Nat. Commun.* **2017**, *8*, 14650.

(24) Pantazes, R. J.; Grisewood, M. J.; Li, T.; Gifford, N. P.; Maranas, C. D. *J. Comput. Chem.* **2015**, *36*, 251−263.

(25) Pavelka, A.; Chovancova, E.; Damborsky, J. *Nucleic Acids Res.* **2009**, *37*, W376−383.

(26) Grosdidier, S.; Fernandez-Recio, J. *BMC Bioinf.* **2008**, *9*, 447.

(27) Zheng, H.; Reetz, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 15744−15751.

(28) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14116−14121.

(29) Kortemme, T.; Kim, D. E.; Baker, D. *Sci. Signaling* **2004**, *2004*, pl2.

(30) Darnell, S. J.; Page, D.; Mitchell, J. C. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 813−823.

(31) Kawarasaki, Y.; Griswold, K. E.; Stevenson, J. D.; Selzer, T.; Benkovic, S. J.; Iverson, B. L.; Georgiou, G. *Nucleic Acids Res.* **2003**, *31*, 126e.

(32) Meyer, M. M.; Hochrein, L.; Arnold, F. H. *Protein Eng., Des. Sel.* **2006**, *19*, 563−570.

(33) Voigt, C. A.; Martinez, C.; Wang, Z. G.; Mayo, S. L.; Arnold, F. H. *Nat. Struct. Biol.* **2002**, *9*, 553−558.

(34) Pantazes, R. J.; Saraf, M. C.; Maranas, C. D. *Protein Eng., Des. Sel.* **2007**, *20*, 361−373.

(35) Moore, G. L.; Maranas, C. D. *AIChE J.* **2004**, *50*, 262−272.

(36) Saraf, M. C.; Maranas, C. D. *Protein Eng., Des. Sel.* **2003**, *16*, 1025−1034.

(37) Moore, G. L.; Maranas, C. D. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5091−5096.

(38) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190−195.

(39) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387−1391.

(40) Huang, P. S.; Boyken, S. E.; Baker, D. *Nature* **2016**, *537*, 320−327.

(41) Garrabou, X.; Wicky, B. I.; Hilvert, D. *J. Am. Chem. Soc.* **2016**, *138*, 6972−6974.

(42) Li, T.; Pantazes, R. J.; Maranas, C. D. *PLoS One* **2014**, *9*, e105954.

(43) Fazelinia, H.; Cirino, P. C.; Maranas, C. D. *Biophys. J.* **2007**, *92*, 2120−2130.

(44) Saraf, M. C.; Moore, G. L.; Goodey, N. M.; Cao, V. Y.; Benkovic, S. J.; Maranas, C. D. *Biophys. J.* **2006**, *90*, 4167−4180.

(45) Pantazes, R. J.; Grisewood, M. J.; Maranas, C. D. *Curr. Opin. Struct. Biol.* **2011**, *21*, 467−472.

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849

(46) Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. *Annu. Rev. Phys. Chem.* **2011**, *62*, 129−149.
(47) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3790−3795.
(48) Baker, D. *Protein Sci.* **2010**, *19*, 1817−1819.
(49) Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R. *J. Mol. Biol.* **2007**, *365*, 135−145.
(50) Grisewood, M. J.; Gifford, N. P.; Pantazes, R. J.; Li, Y.; Cirino, P. C.; Janik, M. J.; Maranas, C. D. *PLoS One* **2013**, *8*, e75358.
(51) Khoury, G. A.; Fazelinia, H.; Chin, J. W.; Pantazes, R. J.; Cirino, P. C.; Maranas, C. D. *Protein Sci.* **2009**, *18*, 2125−2138.
(52) Fazelinia, H.; Cirino, P. C.; Maranas, C. D. *Protein Sci.* **2009**, *18*, 180−195.
(53) Richardson, J. S.; Keedy, D. A.; Richardson, D. C. In *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*; Bansal, M., Srinivasan, N., Eds.; World Scientific Publishing: Hackensack, NJ, 2013; pp 46−61.
(54) Feldmeier, K.; Hocker, B. *Curr. Opin. Chem. Biol.* **2013**, *17*, 929−933.
(55) Khare, S. D.; Kipnis, Y.; Greisen, P., Jr.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J. L.; Silman, I.; Leader, H.; Sussman, J. L.; Stoddard, B. L.; Tawfik, D. S.; Baker, D. *Nat. Chem. Biol.* **2012**, *8*, 294−300.
(56) Gordon, S. R.; Stanley, E. J.; Wolf, S.; Toland, A.; Wu, S. J.; Hadidi, D.; Mills, J. H.; Baker, D.; Pultz, I. S.; Siegel, J. B. *J. Am. Chem. Soc.* **2012**, *134*, 20513−20520.
(57) Chen, C. Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3764−3769.
(58) Lo, Y. C.; Lin, S. C.; Shaw, J. F.; Liaw, Y. C. *J. Mol. Biol.* **2003**, *330*, 539−551.
(59) Morozov, A. V.; Havranek, J. J.; Baker, D.; Siggia, E. D. *Nucleic Acids Res.* **2005**, *33*, 5781−5798.
(60) Meiler, J.; Baker, D. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 538−548.
(61) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(62) Vanommeslaeghe, K.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(63) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.
(64) Lebigot, E. O. *Uncertainties: A Python package for calculations with uncertainties*. http://pythonhosted.org/uncertainties/ (accessed Feb. 23, 2016).
(65) Marchler-Bauer, A.; Derbyshire, M. K.; Gonzales, N. R.; Lu, S. N.; Chitsaz, F.; Geer, L. Y.; Geer, R. C.; He, J.; Gwadz, M.; Hurwitz, D. I.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z. X.; Yamashita, R. A.; Zhang, D. C.; Zheng, C. J.; Bryant, S. H. *Nucleic Acids Res.* **2015**, *43*, D222−D226.
(66) Frank, E.; Hall, M.; Trigg, L.; Holmes, G.; Witten, I. H. *Bioinformatics* **2004**, *20*, 2479−2481.
(67) Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2004**, *32*, W615−W619.
(68) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383−10388.
(69) Huang, Y. M.; Bystroff, C. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2013**, *10*, 1176−1187.
(70) Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. *J. Bacteriol.* **1995**, *177*, 4121−4130.
(71) Thoden, J. B.; Holden, H. M. *J. Biol. Chem.* **2005**, *280*, 32784−32791.
(72) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(73) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.
(74) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(75) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(76) Shin, K. S.; Kim, S.; Lee, S. K. *Biotechnol. Biofuels* **2016**, *9*, 208.
(77) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(78) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

DOI: 10.1021/acscatal.7b00408
ACS Catal. 2017, 7, 3837−3849