

Original Research Article

Enhancing the activity and succinyl-CoA specificity of 3-ketoacyl-CoA thiolase Tfu_0875 through rational binding pocket engineering

Lixia Liu^{a,b}, Shuang Liu^{a,b}, Xiangyang Hu^{a,b}, Shenghu Zhou^{a,b,**}, Yu Deng^{a,b,*}^a National Engineering Research Center of Cereal Fermentation and Food Biomanufacturing, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China^b Jiangsu Provincial Research Center for Bioactive Product Processing Technology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China

ARTICLE INFO

Keywords:

Adipic acid
3-Ketoacyl-CoA thiolase
Claisen condensation
Deep learning
GRAPE strategy

ABSTRACT

The 3-ketoacyl-CoA thiolase is the rate-limiting enzyme for linear dicarboxylic acids production. However, the promiscuous substrate specificity and suboptimal catalytic performance have restricted its application. Here we present both biochemical and structural analyses of a high-efficiency 3-ketoacyl-CoA thiolase Tfu_0875. Notably, Tfu_0875 displayed heightened activity and substrate specificity for succinyl-CoA, a key precursor in adipic acid production. To enhance its performance, a deep learning approach (DLKcat) was employed to identify effective mutants, and a computational strategy, known as the greedy accumulated strategy for protein engineering (GRAPE), was used to accumulate these effective mutants. Among the mutants, Tfu_0875^{N249W/L163H/E217L} exhibited the highest specific activity (320% of wild-type Tfu_0875), the greatest catalytic efficiency ($k_{cat}/K_M = 1.00 \text{ min}^{-1}\text{mM}^{-1}$), the highest succinyl-CoA specificity ($K_M = 0.59 \text{ mM}$, 28.1% of Tfu_0875) and dramatically reduced substrate binding energy ($-30.25 \text{ kcal mol}^{-1}$ v.s. $-15.94 \text{ kcal mol}^{-1}$). A structural comparison between Tfu_0875^{N249W/L163H/E217L} and the wild type Tfu_0875 revealed that the increased interaction between the enzyme and succinyl-CoA was the primary reason for the enhanced enzyme activity. This interaction facilitated rapid substrate anchoring and stabilization. Furthermore, a reduced binding pocket volume improved substrate specificity by enhancing the complementarity between the binding pocket and the substrate in stereo conformation. Finally, our rationally designed mutant, Tfu_0875^{N249W/L163H/E217L}, increased the adipic acid titer by 1.35-fold compared to the wild type Tfu_0875 in shake flask. The demonstrated enzymatic methods provide a promising enzyme variant for the adipic acid production. The above effective substrate binding pocket engineering strategy can be beneficial for the production of other industrially competitive biobased chemicals when be applied to other thiolases.

1. Introduction

Given the limited availability of fossil fuels, utilizing the catalytic activity of enzymes in microorganisms to convert renewable resources into various chemicals has been shown to offer significant advantages [1,2]. The first step in the formation of target chemicals involves the carbon-carbon bond formation and carbon chain elongation, both of which rely on Claisen condensation [3,4]. Typically, Claisen condensation is usually catalyzed by a large family of thiolases [5] to construct diverse carbon skeletons in compounds [6,7]. The thiolase superfamily encompasses an array of enzymes including Archeal thiolase, thiolase I

and II, HMG-CoA synthase (HMGS), β -ketoacyl-ACP synthase I, II and III (KAS I, II and III), β -ketoacyl-CoA synthase (KCS), and polyketide synthase I, II, and III (PKS I, II and III) (Fig. 1) [3,7]. Notably, thiolase I is predisposed to the synthesis of medium-chain (C4–C8) [8] linear aliphatic compounds, which are commonly employed as precursors for fuels [9] and polymers [10,11]. Consequently, advancing our knowledge of the structural and biochemical foundations of thiolase I may enhance the versatility of these enzymes in the enzymatic and microbial production of desired chemicals.

The broad substrate spectrum of thiolase I has impeded its industrial application due to the resultant production of a diverse array of by-

Peer review under responsibility of KeAi Communications Co., Ltd.

* Corresponding author. National Engineering Laboratory for Cereal Fermentation Technology, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China.

** Corresponding author. National Engineering Laboratory for Cereal Fermentation Technology, 1800 Lihu Road, Wuxi, Jiangsu, 214122, China.

E-mail addresses: lixialiu@imut.edu.cn (L. Liu), 6220209029@stu.jiangnan.edu.cn (S. Liu), 1022200129@stu.jiangnan.edu.cn (X. Hu), zhoush@jiangnan.edu.cn (S. Zhou), dengyu@jiangnan.edu.cn (Y. Deng).<https://doi.org/10.1016/j.synbio.2024.04.014>

Received 12 February 2024; Received in revised form 14 April 2024; Accepted 16 April 2024

Available online 20 April 2024

2405-805X/© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

products [2]. Hence, improving the enzyme activity and selectively for specific substrates is the ongoing pursuit of enzyme engineers. With the emergence of rational enzyme design, it is generally believed that the thiolase I can basically achieve the desired transformation [12]. The Claisen condensation reaction occurs within the substrate binding pocket of thiolase superfamily. Therefore, engineering substrate binding pocket would directly affect the substrate specificity and the Claisen condensation efficiency. However, the investigation of the substrate binding pocket of thiolase I has been rarely reported. Blaisse et al. employed methods of sequence and structural alignment to elucidate the distinctions between binding pockets of thiolase Acat5 and Acat2 [5]. Their findings suggest that the incorporation of large benzene rings of phenylalanine residue constricts the binding pocket. This constriction augments the interactions between the amino acid residues of binding pocket and the substrate, thereby enhancing the substrate specificity and enzyme activity towards linear compounds [5]. These results underscore the potential of modifying the dimensions of the substrate binding pocket to refine the catalytic efficiency and substrate selectivity for linear compounds. In addition to enhancing substrate specificity for a particular substrate, tailored mutations can be designed for various substrates to attain targeted substrate specificity. Sofeo et al. modified the binding pocket of acetyl-CoA synthetase (*atACS*) by substituting phylogenetically related residues within the substrate-binding sites. Before modification, the only substrate of *atACS* was acetate, and after modification, it was intended to enhance the substrate specificity towards either linear (hexanoic acid) or branched (methyl valerate) carboxylates [13]. The k_{cat} and K_M for *atACS*^{V399A/W427G} mutant were 8.04 s⁻¹ and 0.09 mM for butyrate. While, the k_{cat} and K_M for *atACS*^{T324G/V399A/W427G} mutant were 5.07 s⁻¹ and 0.14 mM for hexanoate. Currently, enhancements to the catalytic efficiency and substrate specificity of thiolase I predominantly rely on irrational approaches, such as random mutation, which entail a significant workload. Consequently, this study aims to integrate rational design techniques, including deep learning and molecular dynamics simulations, to improve the catalytic performance and substrate specificity of thiolase I.

Recently, we developed a reverse adipate degradation pathway (RADP) for the production of dicarboxylic acids, including glutaric acid [14], adipic acid [15], and pimelic acid [16], which play an important role as biofuels [11] and nylon monomers⁹. The 3-ketoacyl-CoA thiolase, a member of the thiolase I family, serves as the initial and rate-limiting enzyme in the RADP pathway. It catalyzes the Claisen condensation reaction between substrates such as malonyl-CoA, succinyl-CoA, or pimeloyl-CoA, and extension unit acetyl-CoA, resulting in the formation of a 3-oxoacyl-CoA intermediates of dicarboxylic acids [15–17]. However, the wide substrate spectrum of 3-ketoacyl-CoA thiolase often generates mixed products which limited their application in

biosynthesis of specific compounds [2]. In this regard, further improving the substrate specificity of 3-ketoacyl-CoA thiolase was crucial to improve the biosynthesis efficiency based on Claisen condensation. Using Tfu_0875 as a 3-ketoacyl-CoA thiolase in RADP, it was found that the yield of adipic acid was higher than that of other dicarboxylic acids, demonstrating that the substrate specificity and catalytic efficiency of succinyl-CoA were higher [15]. In order to improve the substrate specificity of Tfu_0875 and the efficiency of the Claisen condensation, rationally designing the substrate binding pocket of Tfu_0875 would be a powerful and reliable approach.

Taking inspiration from the engineering of substrate binding pockets in the aforementioned thiolases, it is possible to enhance the substrate specificity and catalytic efficiency of 3-ketoacyl-CoA thiolase through iterative structure-guided mutations aimed at reducing the space within the binding pocket. In this work, we focused on a thermostable 3-ketoacyl-CoA thiolase Tfu_0875 to rationally design its substrate binding pocket by the DLKcat approach [18] and the greedy accumulated strategy for protein engineering (GRAPE) [19] to improve enzyme activity and substrate specificity of succinyl-CoA. After GRAPE, a global optimal mutant (Tfu_0875^{N249W/L163H/E217L}) was obtained, exhibiting that reducing the size of the substrate binding pocket enhances the influence of the surrounding residues on the substrate succinyl-CoA, resulting in improved substrate specificity and catalytic efficiency. When the best mutant was applied in the production of adipic acid, a significant improvement in titer was observed.

2. Results

2.1. Characterization of the 3-ketoacyl-CoA Tfu_0875

The thermophilic actinobacterium *T. fusca* sourced 3-ketoacyl-CoA thiolase Tfu_0875 exhibited a high thermostability and functionally catalyzed the Claisen condensation reaction to lengthen of the carbon chain [20]. Purification via affinity chromatography followed by SDS-PAGE analysis revealed that Tfu_0875 possesses a molecular weight marginally above 40 kDa, aligning with the predicted molecular weight of 40.6 kDa (Fig. S1). Tfu_0875 belongs to the thiolase I which prefers medium-chain length substrates [8]. Therefore, we selected acetyl-CoA, malonyl-CoA, succinyl-CoA, and glutaryl-CoA as substrates to condensation with an extension unit acetyl-CoA by Tfu_0875 for evaluating the substrate specificity [14–16]. As a result, Tfu_0875 demonstrated a relatively broad substrate range, spanning from C2 to C5 carbon chain lengths. When catalyzing substrate succinyl-CoA, the Tfu_0875 exhibited the highest enzyme activity 1.85 μmol·min⁻¹·mg⁻¹ (Fig. 2A).

The pH and temperature had a substantial impact on the activity of Tfu_0875. Herein, the enzyme activity was detected at different reaction

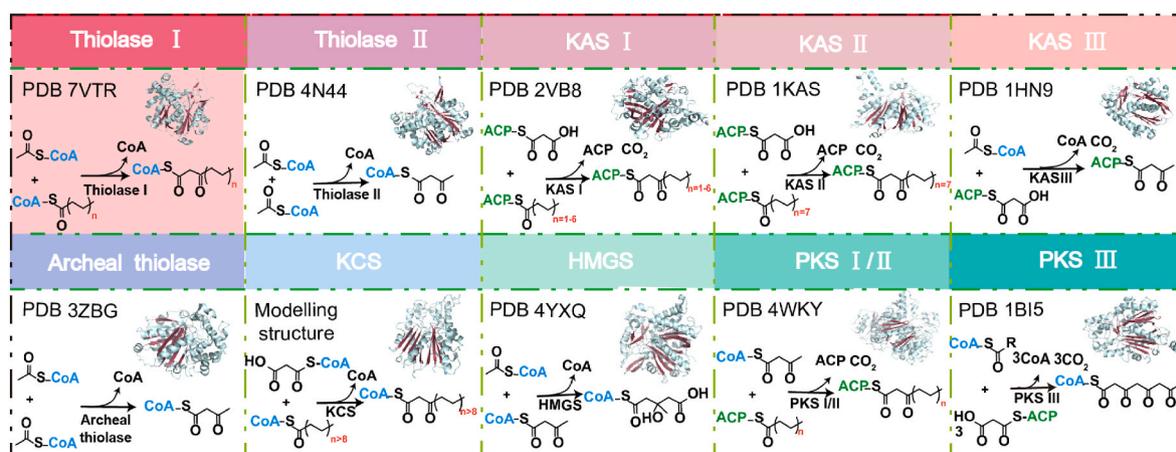


Fig. 1. Classification, stereostructures, and catalytic substrate types of the thiolase superfamily. The corresponding catalytic substrates and monomer structure of each type of thiolases were summarized in the corresponding box. The conserved β -sheets in this superfamily were marked in raspberry.

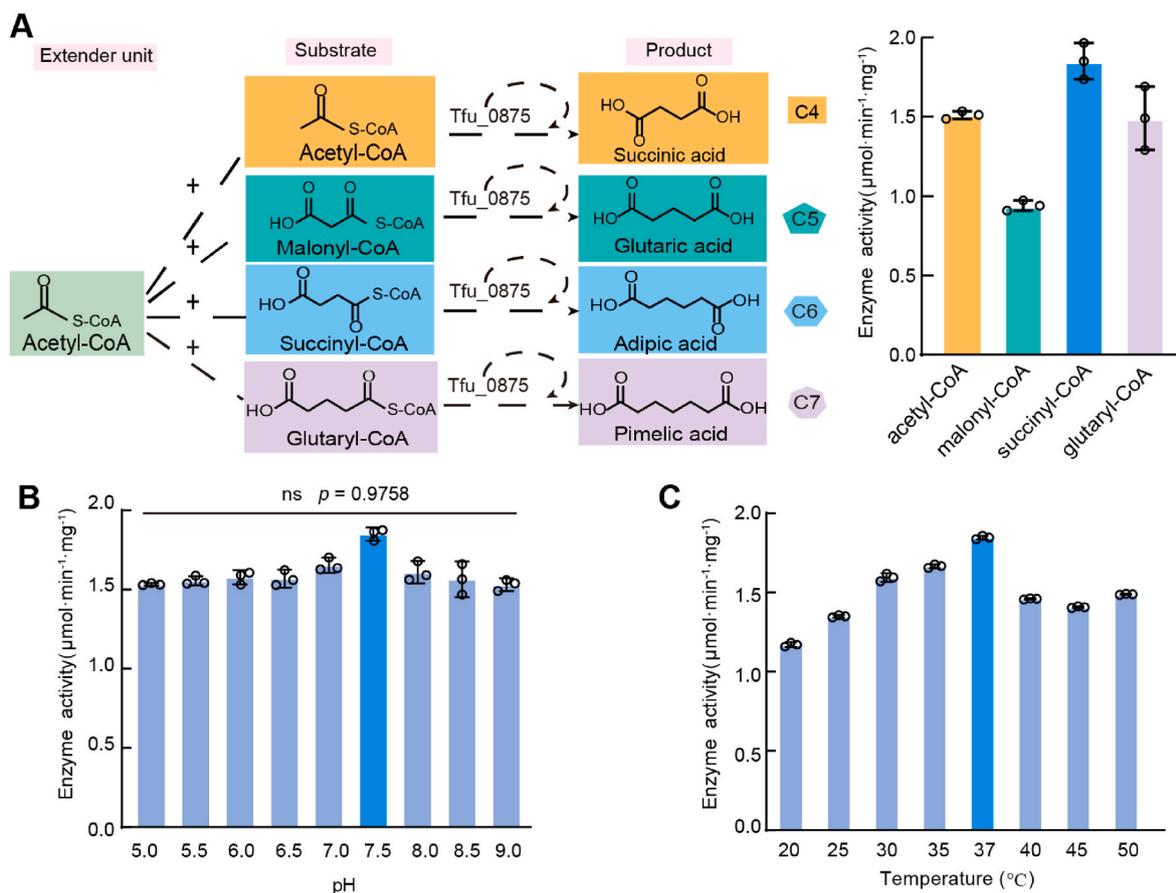


Fig. 2. The detected performance of the Tfu_0875. (A) Claisen condensation reaction (left) and catalytic efficiency (right) of Tfu_0875 on substrates with different carbon chain lengths. (B–C) Effects of different pH values (B) and reaction temperature (C) on the enzyme activity of Tfu_0875. Acetyl-CoA and succinyl-CoA were used as extender unit and substrate, respectively. The bright blue column indicated the highest catalytic activity under the indicated conditions. The data was presented as the mean \pm s.d. ns, no significance.

pH and temperature. Under the pH range from 5.0 to 9.0, the enzyme activity of Tfu_0875 had good adaptability and did not change significantly in acidic, neutral, or alkaline environments (Fig. 2B). This phenomenon was consistent with the other thermophilic enzymes whose enzyme activity is usually not significantly affected by pH changes [21, 22]. On the other hand, the reaction temperature results of Tfu_0875 show that Tfu_0875 demonstrated a wide temperature tolerance, and its optimum reaction temperature at 37 $^{\circ}\text{C}$. The enzyme remained at least 37% of the highest activity when reaction temperature at 20–50 $^{\circ}\text{C}$ (Fig. 2C). Collectively, the high thermostability, broad reaction temperature, and pH range make Tfu_0875 potentially suitable for the production of various chemicals. However, its broad substrate spectrum poses a challenge to this application.

2.2. Screening of effective single mutations of Tfu_0875 via the deep learning

Generally, enzyme specificity for a target molecule is positively related to its activity [23]. Therefore, it is possible to simultaneously enhance the specificity and activity of Tfu_0875 by altering the binding pockets [24]. To identify the amino acid residues that could potentially influence the enzyme activity and substrate specificity, we performed docking simulations of succinyl-CoA into the binding pocket of the crystal structure of Tfu_0875 (PDB ID: 7VTR). Subsequently, we selected mutation candidates from the amino acid residues located within a 5 \AA range of succinyl-CoA. These mutation targets included 16 amino acid residues: M120, G147, L163, E217, S218, R220, E221, T222, K226, F234, I240, S246, P247, L248, N249, and F318 (Fig. 3A). To identify the

key candidates, we performed alanine scans on the above 16 amino acid residues. The results revealed that 9 mutations (M120A, G147A, L163A, E217A, S218A, R220A, E221A, I240A, and N249A) significantly impacted the enzyme activity (Fig. 3B). These amino acid residues were identified as potential key determinants influencing the binding and catalysis of succinyl-CoA. Wherein, the mutant Tfu_0875^{M120A} exhibited a remarkable increase in enzyme activity, with the highest activity being 185% of the wild-type Tfu_0875. So, M120 no longer participated in the subsequent screening, and the remaining 8 key amino acid residues participated in the screening.

Currently, high-throughput screening methods were unavailable to identify positive mutants from the Tfu_0875 saturation mutagenesis library, resulting in a time- and labor-consuming screening process. In this regard, selecting proper mutants or hot spots by computer-assisted or artificial intelligence method is a necessity for developing smart libraries. Deep learning-based k_{cat} prediction (DLKcat) was a tool for predicting enzyme turnover numbers and was able to capture the effects of amino acid substitutions on the k_{cat} values of individual enzymes [18]. Amino acid sequences of Tfu_0875 variants were processed by convolutional neural network (CNN), and the simplified molecular-input line-entry system (SMILES) for succinyl-CoA was processed by a graph neural network (GNN). Subsequently, the protein sequences and SMILES for succinyl-CoA were calculated using a Neural attention mechanism, the k_{cat} values were predicted based on the comprehensive dataset (Fig. 3C). There are 20 amino acids in nature, and only 18 mutations are possible in the remaining 8 amino acid residues after removing their own amino acids and alanine. Among the 144 (8×18) virtual mutants, 17 potential dominant mutants whose predicted k_{cat} values higher than

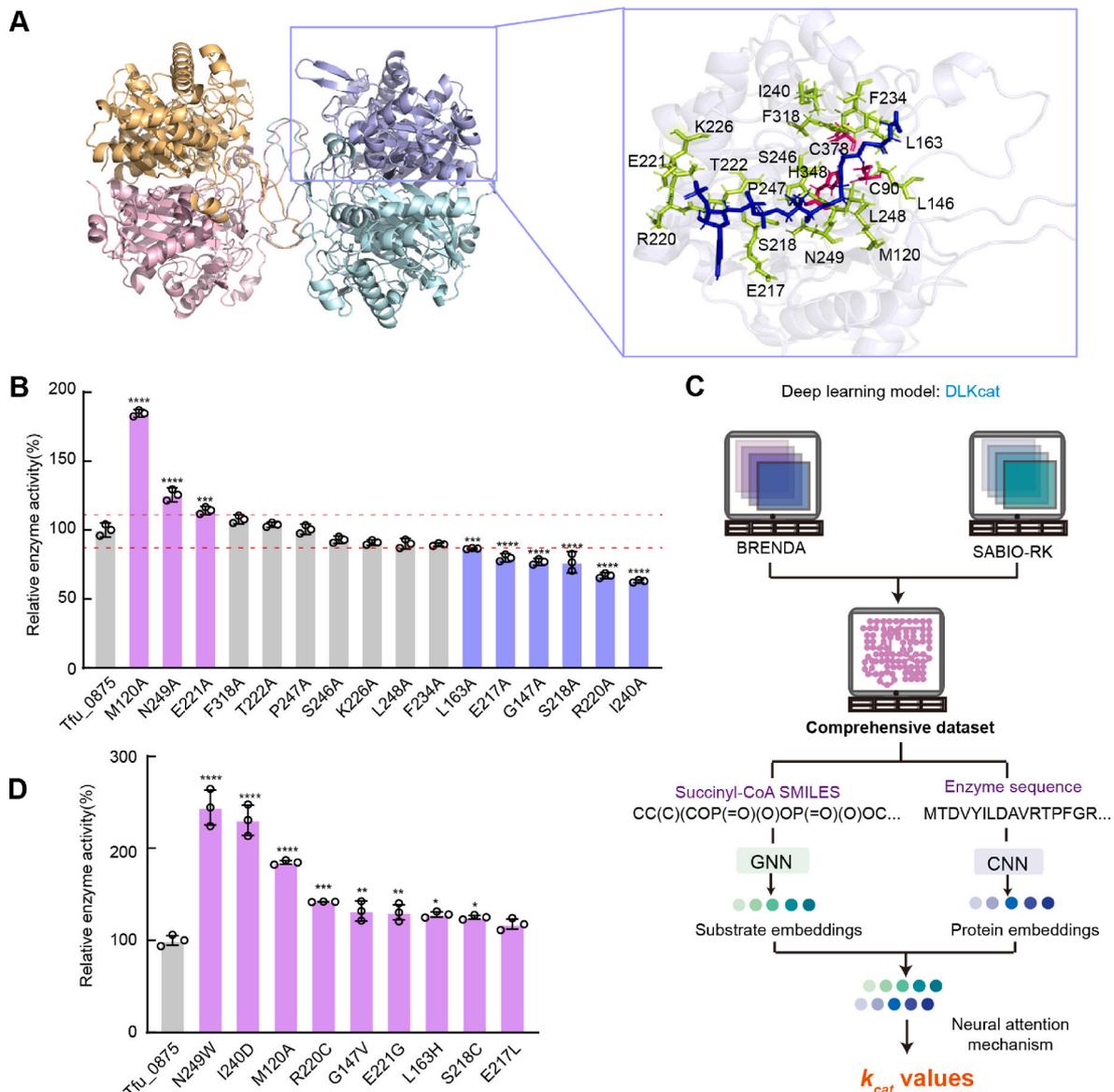


Fig. 3. Screening for effective single mutants of the Tfu_0875. (A). The homologous tetrameric structure of Tfu_0875. Subunits were coloured differently (left). The 16 amino acid residues (limon) and their active residues (C90, C378, and H348; hotpink) within 5 Å range centered on succinyl-CoA (blue) were exhibited (right). (B) Relative enzyme activity of mutants and Tfu_0875 in alanine scanning. Mutants with no significant change, significant increase, and significant decrease in enzyme activity were color as gray, pink, and purple, respectively. (C) The diagram of the deep learning to predict the k_{cat} of mutations by DLKcat. (D) The relative enzyme activities of effective sites as determined by experiment. *, **, *** and **** represent p -values of mutants <0.05, 0.01, 0.001 and 0.0001 in comparison with Tfu_0875, respectively. Error bars are standard deviations ($n = 3$).

that of Tfu_0875 were obtained after DLKcat prediction, and the optimal mutants of the corresponding amino acid residues were verified. Finally, 9 mutants were identified to be effective with enzyme activity increased by 9%–144% of Tfu_0875 (Fig. 3D and Table S4).

2.3. Improving the substrate activity and specificity of Tfu_0875 on succinyl-CoA by the greedy accumulated strategy

The obtained effective single substitutions (Fig. 3D) could be iterated to further improve the enzyme activity of Tfu_0875 to the succinyl-CoA. However, the iteration of these effective mutants may lead to dominant mutation interacting with each other, resulting in unfavorable epistatic interactions and a decrease in enzyme activity during the iteration process [25]. To overcome this challenge, we applied a greedy accumulated strategy for protein engineering (GRAPE) to engineer Tfu_0875 (Fig. 4A) [19]. By incrementally addressing the effective substitutions in

the GRAPE scheme with multilayer upward branches, it becomes possible to determine and quantify negative trade-offs, thus providing effective strategies to avoid apparent dead ends. To do this, the effective single substitutions (Fig. 3D) were classified into two clusters using the K-means approach [26], based on their relative enzyme activity, the distance from the active center (C90, C378, H348) to the $C\alpha$ of succinyl-CoA, as well as the number of hydrogen bonds formed between the substitutions and succinyl-CoA (Fig. 4A). These conditions are regarded as critical factors that influence substrate binding and the catalytic process. Then we performed an iterative site-specific mutagenesis based on the K-means analysis to achieve global optimization [10,19]. Each round of GRAPE engineering would produce the best mutation combination of Tfu_0875 which could serve as a template for further GRAPE engineering. The GRAPE engineering was iterated until the enzyme activity did not significantly increase or even decrease (Fig. 4A).

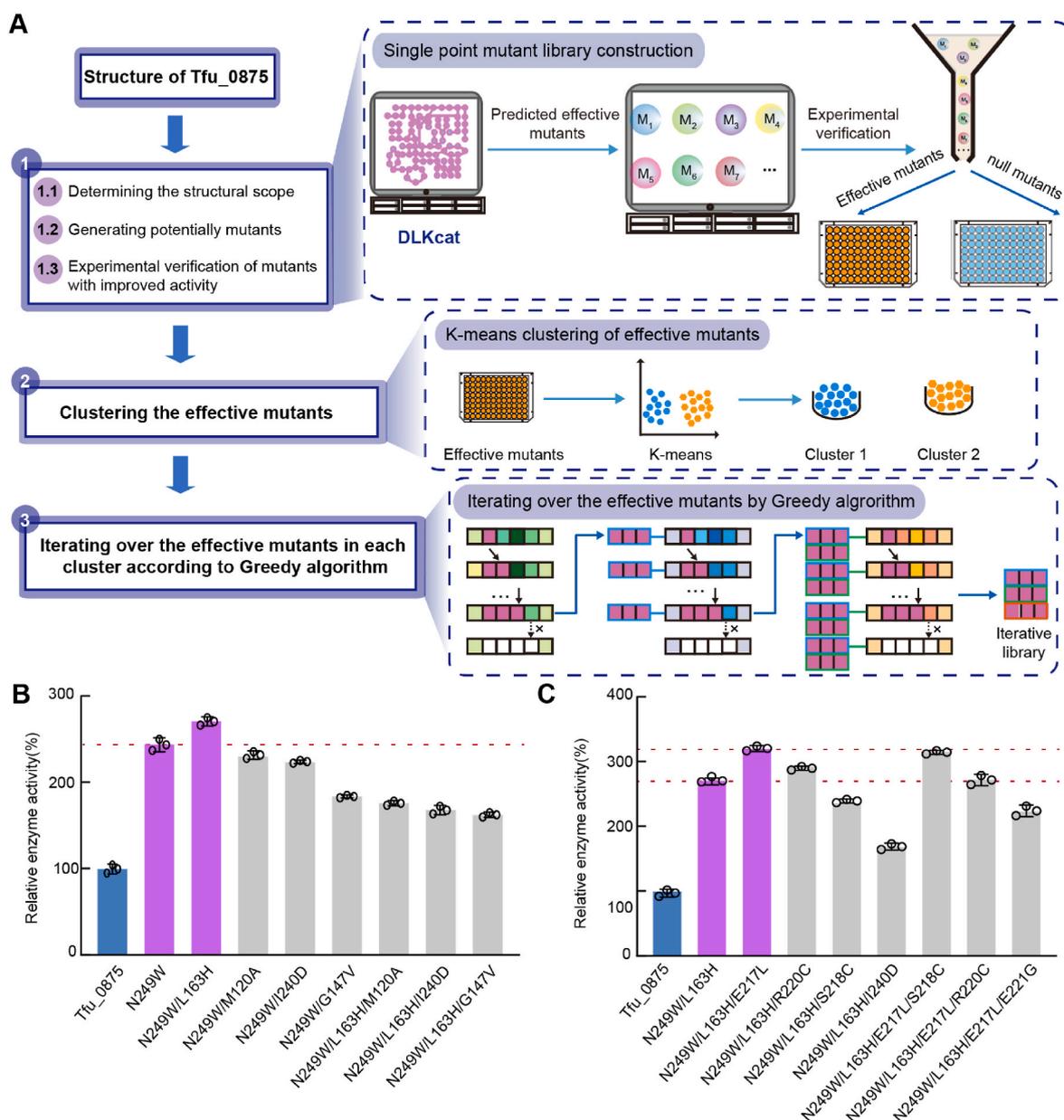


Fig. 4. Applying GRAPE strategy to improve the enzyme activity of Tfu_0875 on succinyl-CoA. (A) Schematic representation of the GRAPE strategy. In step 1, DLKcat were used to predict effective single mutants. Then, the effective single mutants were chosen for experimental validation. Step 2, Classify these single mutants according to their positions, interactions, and activities. In step 3, accumulation of the mutants in each cluster according to the greedy algorithm. (B–C) Enhancing the relative enzyme activity of Tfu_0875 by greedy accumulation in the first cluster (B) and second cluster (C). Relative enzyme activity = (enzyme activity of mutant/enzyme activity of Tfu_0875) × 100%.

After K-means analysis, the M120A, G147V, L163H, I240D, and N249W were characterized into cluster 1, whereas the E217L, S218C, R220C, and E221G were characterized into cluster 2 (Table S5). We selected N249W as the starting template for the first round of GRAPE engineering, as it has been shown to greatly enhance enzyme activity. Other substitutions in the first cluster (M120A, G147V, L163H, and I240D) were combined with the N249W in the first round of GRAPE engineering, resulting in the best mutant Tfu_0875^{N249W/L163H} with 27% enhancement in enzyme activity compared with Tfu_0875^{N249W}. However, using Tfu_0875^{N249W/L163H} as the template for the next round of GRAPE engineering to accumulate any remaining residues in the first cluster (M120A, G147V, and I240D) led to a reduction in enzyme activity (Fig. 4B). The K_M values of Tfu_0875^{N249W} and Tfu_0875^{N249W/L163H} were lower than the Tfu_0875, meanwhile, the k_{cat} values of Tfu_0875^{N249W} and Tfu_0875^{N249W/L163H} were higher than the Tfu_0875

(Table 1). Therefore, Tfu_0875^{N249W/L163H} was identified as the most effective mutant in the first cluster, and was selected as the template for the GRAPE engineering in the second cluster substitutions.

Combining substitutions in the second cluster (E217L, S218C, R220C, and E221G) with N249W/L163H generated a best mutant Tfu_0875^{N249W/L163H/E217L}, and the specific activity of Tfu_0875^{N249W/L163H/E217L} was 320% of that of Tfu_0875. Comparing with Tfu_0875^{N249W/L163H}, the decreased K_M value and increased k_{cat} value of Tfu_0875^{N249W/L163H/E217L} indicated a gradual increase in affinity and turnover number of the mutants to the succinyl-CoA (Table 1). Finally, the K_M and k_{cat} values of the best mutant Tfu_0875^{N249W/L163H/E217L} was 0.59 mM and 0.59 min⁻¹ to succinyl-CoA, 28.1% and 184.4% that of Tfu_0875, respectively. Thus, Tfu_0875^{N249W/L163H/E217L} exhibited a significantly improved catalytic efficiency ($k_{cat}/K_M = 1.00 \text{ min}^{-1}\text{mM}^{-1}$), which was 6.67-fold higher than that of Tfu_0875 (Table 1).

Table 1

Kinetic data of Tfu_0875 and mutants toward substrates with different chain lengths.

| Substrates | Enzymes | K_M (mM) | k_{cat} (min ⁻¹) | k_{cat}/K_M (min ⁻¹ mM ⁻¹) |
|--------------|---------------------------------------|---------------|-----------------------------------|--|
| Succinyl-CoA | Tfu_0875 | 2.10 ± 0.22 | 0.32 ± 0.0022 | 0.15 |
| | Tfu_0875 ^{N249W} | 1.37 ± 0.26 | 0.33 ± 0.0032 | 0.24 |
| | Tfu_0875 ^{N249W/L163H} | 1.14 ± 0.88 | 0.39 ± 0.042 | 0.34 |
| | Tfu_0875 ^{N249W/L163H/E217L} | 0.59 ± 0.83 | 0.59 ± 0.14 | 1.00 |
| | Tfu_0875 | 2.28 ± 0.64 | 0.21 ± 0.0039 | 0.09 |
| Acetyl-CoA | Tfu_0875 | 2.28 ± 0.64 | 0.21 ± 0.0039 | 0.09 |
| | Tfu_0875 ^{N249W} | 1.15 ± 0.81 | 0.31 ± 0.0047 | 0.27 |
| | Tfu_0875 ^{N249W/L163H} | 2.19 ± 0.41 | 0.22 ± 0.036 | 0.10 |
| | Tfu_0875 ^{N249W/L163H/E217L} | 2.18 ± 0.33 | 0.26 ± 0.13 | 0.12 |
| | Tfu_0875 | 3.32 ± 0.86 | 0.19 ± 0.0048 | 0.06 |
| Malonyl-CoA | Tfu_0875 | 3.32 ± 0.86 | 0.19 ± 0.0048 | 0.06 |
| | Tfu_0875 ^{N249W} | 2.89 ± 0.25 | 0.21 ± 0.24 | 0.07 |
| | Tfu_0875 ^{N249W/L163H} | 2.77 ± 0.03 | 0.20 ± 0.02 | 0.07 |
| | Tfu_0875 ^{N249W/L163H/E217L} | 2.57 ± 0.18 | 0.18 ± 0.21 | 0.07 |
| | Tfu_0875 | 2.33 ± 0.55 | 0.21 ± 0.039 | 0.09 |
| Glutaryl-CoA | Tfu_0875 | 2.33 ± 0.55 | 0.21 ± 0.039 | 0.09 |
| | Tfu_0875 ^{N249W} | 2.27 ± 0.88 | 0.20 ± 0.02 | 0.09 |
| | Tfu_0875 ^{N249W/L163H} | 2.28 ± 0.16 | 0.20 ± 0.01 | 0.09 |
| | Tfu_0875 ^{N249W/L163H/E217L} | 2.20 ± 0.82 | 0.30 ± 0.042 | 0.14 |

However, using Tfu_0875^{N249W/L163H/E217L} as the template for the next round of GRAPE engineering to accumulate any remaining residues in the second cluster (S218C, R220C, and E221G) decreased the enzyme activity (Fig. 4C). Finally, the greedy accumulation process resulted in a total of 14 combined mutants. Tfu_0875^{N249W/L163H/E217L} was identified as the most effective mutant in the second cluster, showing a significant improvement in relative enzyme activity.

Furthermore, we investigated the binding specificity and catalytic efficiency of the iteratively mutated enzymes for substrates of varying carbon chain lengths. The results indicated that as the number of iterations in the mutations increased, both the affinity (K_M) and catalytic turnover number (k_{cat}) for succinyl-CoA improved significantly. However, these iterative mutations had only a marginal effect on the binding and turnover numbers for acetyl-CoA, malonyl-CoA, and glutaryl-CoA (Table 1). Finally, compared to succinyl-CoA, the optimum mutant Tfu_0875^{N249W/L163H/E217L} exhibited catalytic efficiencies (k_{cat}/K_M) that were 8.33-fold, 14.28-fold, and 7.14-fold greater for acetyl-CoA, malonyl-CoA, and glutaryl-CoA, respectively. In contrast, the wild-type Tfu_0875 displayed only 1.67-fold, 2.50-fold, and 1.67-fold higher for the same substrates (Table 1). These results suggest that the iterative mutations have significantly improved specificity and activity towards succinyl-CoA, while the effects on other substrates were relatively minor.

2.4. Revealing the structure-function relationship of Tfu_0875^{N249W/L163H/E217L} by molecular dynamics (MD) simulations

In order to reveal the structure-function relationship, succinyl-CoA was docked into the binding pocket of Tfu_0875 and Tfu_0875^{N249W/L163H/E217L} for 100 ns MD simulations [27]. The results showed that the binding energy of Tfu_0875^{N249W/L163H/E217L} with succinyl-CoA was dramatically decreased, compared with that of Tfu_0875 (−30.25 kcal

mol⁻¹ v.s. −15.94 kcal mol⁻¹), indicating an increased affinity after iterative mutation (Table 2). The root mean square deviations (RMSD) were calculated in 100 ns to examine the stability of the Tfu_0875/Tfu_0875^{N249W/L163H/E217L}-succinyl-CoA complexes (Fig. S2). Both Tfu_0875 and Tfu_0875^{N249W/L163H/E217L} reached the equilibrium state after the 100 ns simulation. The root mean square fluctuation (RMSF) of the protein residues was calculated to assess protein flexibility (Fig. S3). The results showed that there was no significant difference in flexibility between Tfu_0875 and Tfu_0875^{N249W/L163H/E217L}.

The trajectories recorded during 100 ns of MD simulation revealed distinct spatial conformation in Tfu_0875 and Tfu_0875^{N249W/L163H/E217L}. By comparison, it was found that the volume of amino acid residues W249, H163, and L217 is larger compared to N249, L163, and E217 (Fig. 5A). Hence, substitutions of L163H, E217L, and N249W resulted in a decrease in the volume of the binding pocket from 1369.3 Å³ to 1146.1 Å³ [28] (Fig. 5B and C). Further analysis of MD results showed that Tfu_0875 could interact with succinyl-CoA to form a total of six hydrogen bonds with R16, S162, R220, and T222, and to form a salt bridge with K226 (Fig. 5D). While, the Tfu_0875^{N249W/L163H/E217L} could interact with succinyl-CoA to form a total of nine hydrogen bonds with T140, L217, I219, R220, K226, A233, A242, S246, and P247, two salt bridges with K226 and K231, and one Pi-cation interaction with R220 (Fig. 5E). The reduction in the size of the binding pocket promotes increased interaction between the pocket and the substrate, leading to the rapid anchoring and stabilization of the substrate. This, in turn, facilitates the occurrence of the reaction.

2.5. Adipic acid production in shake flask fermentation by employing Tfu_0875 and mutants

3-Ketoacyl-CoA thiolase catalysis the carbon chain elongation of succinyl-CoA to generate 3-oxoadipyl-CoA, the precursor of adipic acid [29]. Hence, the production of adipic acid using the modified 3-ketoacyl-CoA thiolase with improved enzyme activity and substrate specificity was expected to increase the yield of adipic acid. To do this, we replaced the 3-ketoacyl-CoA thiolase in the previously constructed strain pAD24 with Tfu_0875^{N249W}, Tfu_0875^{N249W/L163H}, Tfu_0875^{N249W/L163H/E217L} [2,30] (Fig. 6A). In addition, gene *cat1* was overexpressed to provide more precursor succinyl-CoA. An *E. coli* K12 MG1655 Δ atoB Δ sucD Δ pf1B Δ arcA Δ adhE Δ ldhA Δ poxB Δ pta strain was used as the host to redirect carbon flux into adipic acid synthetic pathway and reduce the formation of by-products [31]. In wild-type Tfu_0875 overexpressed strain, the production of adipic acid was growth-coupled [15], and the highest titer of adipic acid (174.7 mg L⁻¹) achieved at 72 h (Fig. 6B). However, after Tfu_0875 was replaced by Tfu_0875^{N249W}, Tfu_0875^{N249W/L163H}, and Tfu_0875^{N249W/L163H/E217L}, the titer of adipic acid increased to 188.0 mg L⁻¹, 209.1 mg L⁻¹, and 235.0 mg L⁻¹ after 72 h of fermentation, 1.08-fold, 1.20-fold, and 1.35-fold of that of Tfu_0875, respectively (Fig. 6C). LC-MS analysis confirmed the presence of the ion peak [M-H]⁻ at *m/z* 145.05, which corresponded to the product ion of adipic acid and was consistent with

Table 2Binding energy of succinyl-CoA for Tfu_0875 and Tfu_0875^{N249W/L163H/E217L}.

| | Binding energy (kcal·mol ⁻¹) of Tfu_0875 | Binding energy (kcal·mol ⁻¹) of Tfu_0875 ^{N249W/L163H/E217L} |
|------------------------|--|---|
| ΔG_{VDW}^a | -21.92 | -36.92 |
| ΔG_{Et}^b | 409.69 | 203.04 |
| ΔG_{polar}^c | -399.44 | -190.10 |
| ΔG_{apolar}^d | -4.27 | -6.27 |
| $\Delta G_{Binding}^e$ | -15.94 | -30.25 |

^a van der Waals energy.

^b Electrostatic energy.

^c Polar-solvation energy.

^d Nonpolar solvation energy.

^e $\Delta G_{Binding} = \Delta G_{VDW} + \Delta G_{Et} + \Delta G_{polar} + \Delta G_{apolar}$

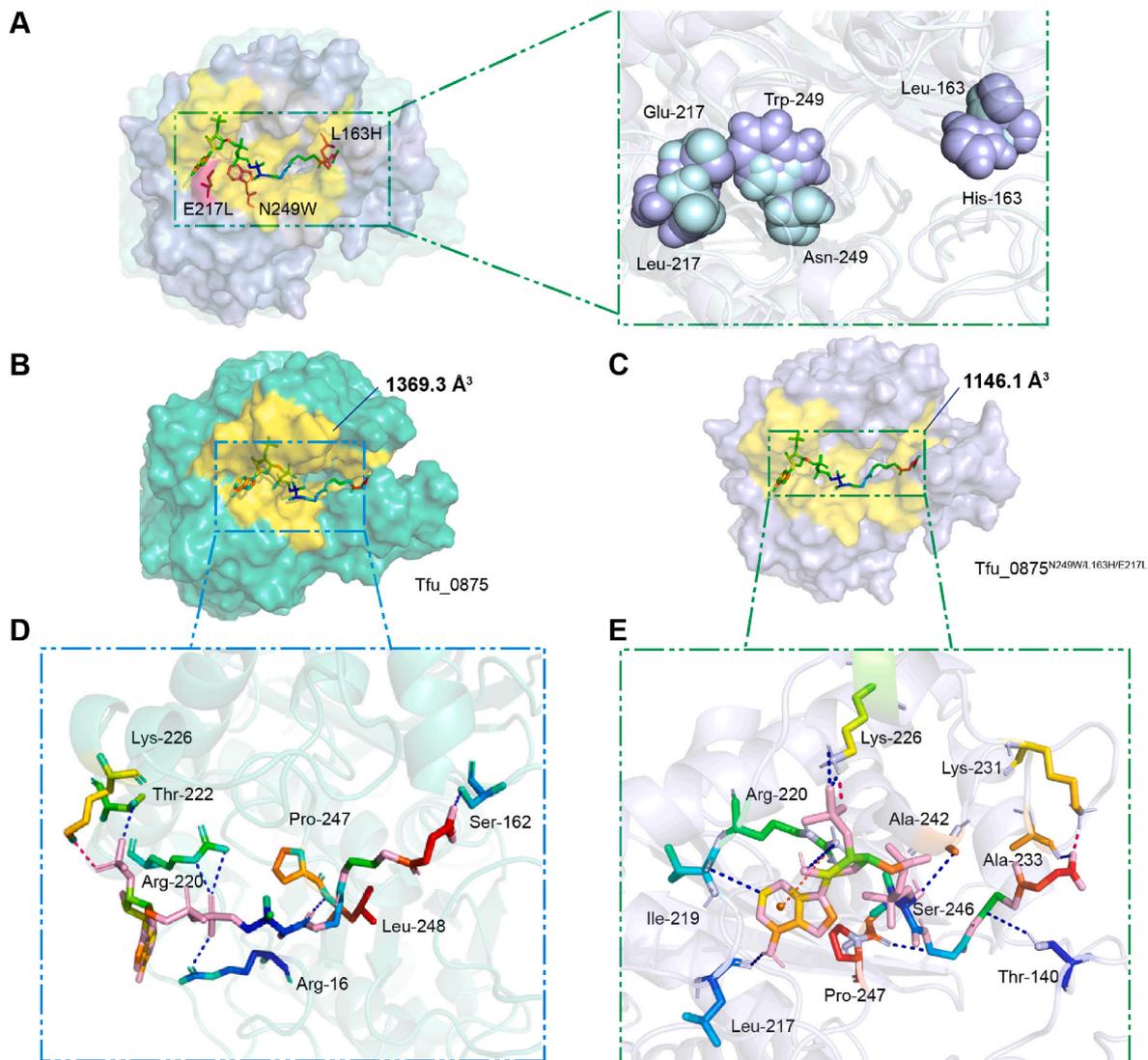


Fig. 5. Structure changes of Tf_u0875 and Tf_u0875^{N249W/L163H/E217L} and their interactions with succinyl-CoA. (A) The steric changes of the substrate binding pocket of Tf_u0875 and Tf_u0875^{N249W/L163H/E217L}. The sphere of N249, L163, and E217 were marked by palecyan, and the sphere of W249, H163, and L217 were marked by light-blue. (B–C) Average surface representation of Tf_u0875 (B, greencyan) and Tf_u0875^{N249W/L163H/E217L} (C, lightblue) in complex with succinyl-CoA at final 100 ns MD trajectory. The substrate binding pockets were marked in yellow, and their volumes were measured using ProteinsPLUS. (D) Detailed view of key interaction between succinyl-CoA and Tf_u0875 after MD simulation. (E) Detailed view of key interaction between succinyl-CoA and Tf_u0875^{N249W/L163H/E217L} after MD simulation. The hydrogen bonds were marked with blue dotted lines, the salt bridges were marked with hotpink dotted lines and the Pi-cation interaction was marked with an orange dotted line.

the standard adipic acid (Fig. 6D). No succinic acid, glutaric acid, and pimelic acid were found in the fermentation solution. It may require the addition of acetyl-CoA synthase, malonic acid synthase or glutaric acid synthase similar to succinyl-CoA synthetase to increase the level of acetyl-CoA, malonyl-CoA or glutaryl-CoA. Overall, the rational design of the substrate binding pocket of Tf_u0875 led to a significant improvement in enzyme activity and specificity and promoted the production of adipic acid. This approach could also be useful in the biosynthesis of other chemicals that require the condensation of acetyl-CoA and succinyl-CoA.

3. Discussion

3-Ketoacyl-CoA thiolase is attractive enzyme for the biosynthesis of products with varying chain lengths and functionalities [8,16,29]. However, the inherent promiscuity of their substrate specificity often results in the production of a mixture of products, which can split the metabolic flux and reduce productivity [32]. In this study, we focused on

the substrate binding pocket of a *T. fusca* sourced 3-ketoacyl-CoA thiolase Tf_u087 to improve enzyme activity and substrate specificity. Using the DLKcat strategy to rationally predict effective amino acid residues, obtained the 11 effective residues for increasing enzyme activity toward succinyl-CoA (Fig. 3B). Then, by using the GRAPE strategy to iteratively optimize these effective residues, the enzyme activity of the optimal mutant Tf_u0875^{N249W/L163H/E217L} was increased to 221 % relative to Tf_u0875 (Fig. 4C). The smaller substrate binding pocket, along with stronger interactions between Tf_u0875^{N249W/L163H/E217L} and succinyl-CoA, are likely the main reasons of the improvement in enzyme activity and specificity. This mechanism provided a reference for future modifications of other 3-ketoacyl-CoA thiolases. Using Tf_u0875^{N249W/L163H/E217L} to produce adipic acid, the highest titer reached 235.0 mg/L, 1.35-fold of that of the Tf_u0875. This highly active Tf_u0875 mutant obtained by rational design also holds potential commercial value for future industrial production of adipic acid.

The substrate binding pocket of the enzymes composed of residues that directly or indirectly involved in the catalysis [33]. Therefore,

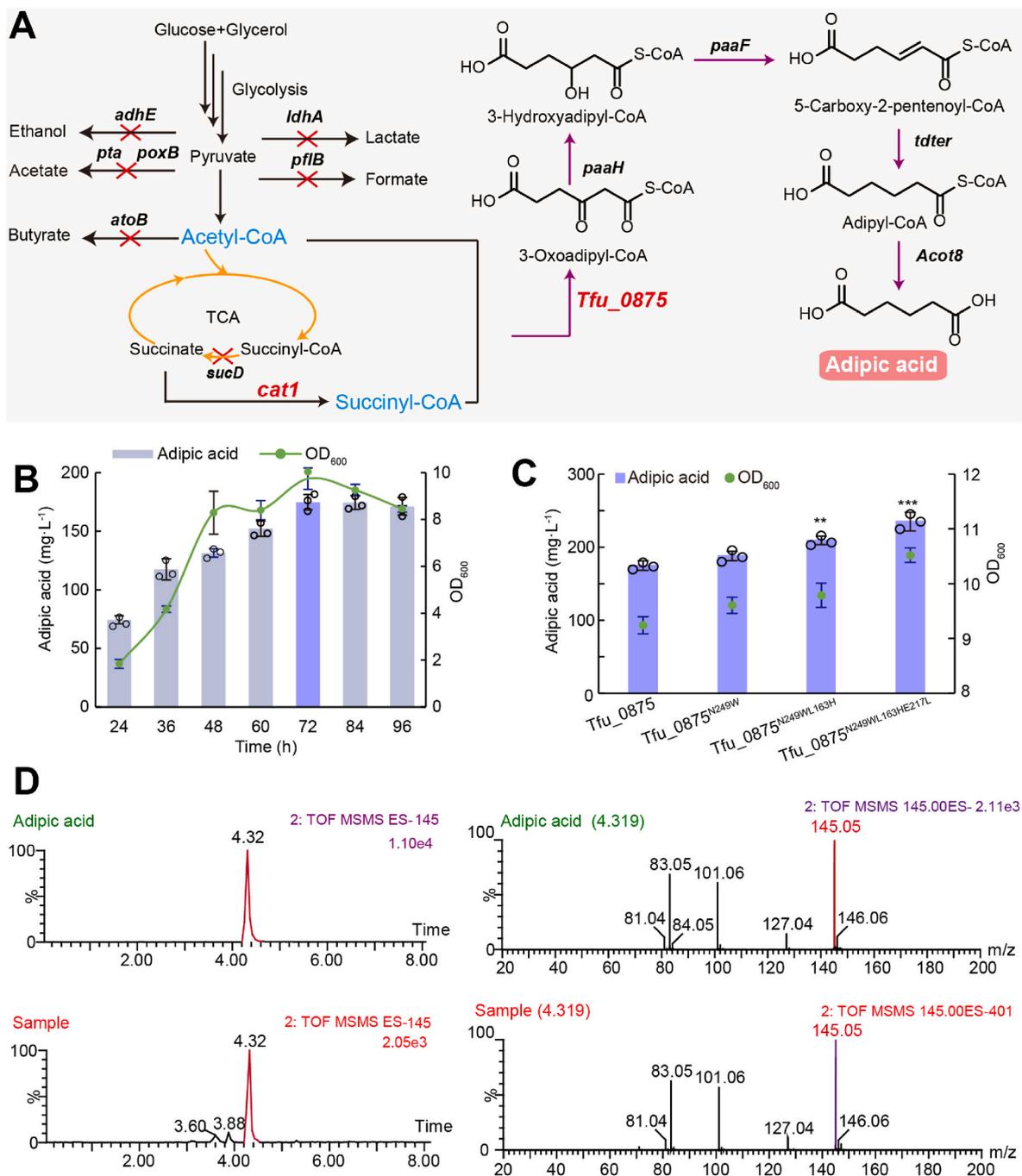


Fig. 6. Adipic acid production by using engineered Tfu_0875 in shake flask fermentation. (A) Metabolic pathway for the production of adipic acid. The orange arrows indicated the TCA pathway. The purple arrows indicated the adipic acid synthesis pathway. The red arrows indicate knocked-out genes. PaaH, 3-hydroxyacyl-CoA dehydrogenase; PaaF, 3-hydroxyadipyl-CoA dehydrogenase; TER, *trans*-enoyl reductase; Acot8, adipyl-CoA transferase; Cat1, coenzyme A transferase. Pta, phosphate acetyltransferase; PoxB, pyruvate oxidase; AdhE, acetaldehyde dehydrogenase; LdhA, lactate dehydrogenase; PflB, pyruvate formate-lyase; AtoB, acetyl-CoA acetyltransferase; SucD, succinyl-CoA synthetase alpha subunit; ArcA, DNA-binding transcriptional dual regulator. (B) The titer of adipic acid produced by Tfu_0875 under different fermentation time. (C) Adipic acid production by Tfu_0875 and its mutants at 72h. ** and *** represent *p*-values of mutants <0.01 and 0.001 in comparison with Tfu_0875, respectively. Error bars represent the s.d. from three independent assays. (D) LC-MS analysis of adipic acid produced in the culture medium (below) and adipic acid standard (above).

adjusting the substrate binding pocket has the potential to increase substrate specificity and the enzyme activity [34], directing more metabolic flow towards the target product rather than synthetic by-products, ultimately increasing the overall yield [9]. However, the identification and rational mutation of effective amino acid residues from the substrate binding pocket remains a challenge. In resolving this challenge, Robinson et al. used phylogenetics and bioinformatics to identify key residues in the substrate binding pocket of thiolase. They

then used a random forest model of machine learning to screen key residues from thiolase OleA, leading to the identification of residues 172, 173, 284, 287, and 316 as significant determinants within the substrate binding pocket [35]. However, this machine learning method requires the acquisition of a large number of dataset (>1000) for random training, which determines the chemical characteristics of each sequence [35]. In contrast, the DLKcat is not limited by the dataset and can directly capture changes in substrate turnover, a direct indicator of

enzyme activity [18]. Moreover, combining the identified effective substitutions into a single mutant is a time- and labor-consuming task [36]. Rational iterative mutation methodologies can significantly reduce the workload of protein engineering [37]. The GRAPE strategy used in this study is a systematic clustering and greedy combination approach [19]. It groups mutant residues based on similar principles at the beginning of the iteration, which helps reduce workload and effectively mitigate initial iteration-induced epistasis.

Generally, the size of the substrate binding pocket and the strength of interactions with substrates are crucial factors that determine specificity and catalytic efficiency [24]. In this study, we observed that the enhanced interactions between succinyl-CoA and the substrate binding pocket were accompanied by improvements in activity (Fig. 5B and C). This finding is consistent with a previous report where the substrate binding pocket of thiolase CaTHL was engineered, resulting in the mutant CaTHL^{V77Q/N153Y/A286K} [9]. This mutant exhibited a remarkable over 3-fold increase in activity. The improved activity can be attributed to the increased interaction of the surrounding amino acids with the substrate, thereby enhancing the catalytic efficiency [9]. Furthermore, the study by Blaisse et al. also supports our findings. They found that the substrate binding pocket of thiolase Acat5 contained a Phe residue that reduced the space in the binding pocket, leading to a strong substrate preference for linear substrates [5]. These findings are consistent with our study, where reducing the space in the binding pocket improves linear substrate specificity, while increasing the interaction with surrounding residues enhances enzyme activity. Thus, our results align with previous studies, providing further evidence for the importance of modulating the binding pocket for substrate specificity and enzyme activity.

Given the pivotal role of 3-ketoacyl-CoA thiolase, this enzyme offers the potential for synthesizing a wide range of chemicals by utilizing different acyl-CoAs as both extender units and substrates [2,8]. This enzyme not only enables the synthesis of medium-chain dicarboxylic acids but also facilitates the production of ω -phenylalkanoic acids, ω -hydroxy acids, ω -1-keto acids, ω -1-methyl alcohols, and 2-methyl acids [2]. The diverse range of products synthesized by this enzyme serves as a promising starting point for metabolizing and generating a wide array of chemical compounds. Consequently, future efforts can focus on modifying the enzyme to catalyze specific substrates, thereby enhancing the efficient synthesis of various compounds and increasing the yield of desired products. A relevant study by Bonk et al. exemplifies this concept by improving the yield of polyhydroxyalkanoates through rational modification of the substrate specificity of 3-ketoacyl-CoA thiolase BktB [32]. Compared with other 3-ketoacyl-CoA thiolase, Tfu_0875 exhibits superior thermal stability and pH adaptability, making it an ideal platform enzyme for biosynthesis.

Overall, this study demonstrates the effectiveness of using deep learning approaches to identify subtle variations that can enhance the substrate specificity of succinyl-CoA and Claisen condensation in Tfu_0875. The combination of these identified variations, Tfu_0875^{N249W/L163H/E217L}, designed through the GRAPE strategy, proves to be a valuable catalyst for efficient Claisen condensation in traditional fermentation processes. This work represents an advancement in enzyme design methodology, complementing conventional computational design approaches, and paves the way for highly efficient adipic acid production.

4. Materials and methods

4.1. Genes, strains, and medium

E. coli JM109 was used as the host for plasmid cloning. *E. coli* BL21 (DE3) was used as the host to express the proteins. *E. coli* K12 MG1655 Δ 8 was used as the host to produce the adipic acid. The *E. coli* strains were grown on Lysogeny Broth (LB) at 37 °C. LB medium containing 50 μ g mL⁻¹ kanamycin was used for plasmid cloning, collecting bacteria,

and seed cultivation. Modified MOPS medium [2] which contained 4 g L⁻¹ glucose, 20 g L⁻¹ glycerin, 20 mM succinic acid, 50 μ g mL⁻¹ kanamycin, and 50 μ g mL⁻¹ chloramphenicol, pH 7.4 was used for adipic acid production in shake flasks fermentation. The genes used in this study were listed in Table S1. Plasmids utilized in this study were donated by other colleagues in our laboratory (Table S2 and Fig. S4) Tfu_0875 was subcloned to pET28a vector.

4.2. Construction of mutant library

The mutagenesis libraries for specific amino acid residues were generated using the pET-28a-Tfu_0875 as a template. 2 × Phanta Flash Master Mix (520, Vazyme Biotech co. Ltd., Nanjing, China) was used to perform site-specific mutagenesis. In the process of mutation, we used F-primers/R-primers containing mutational bases as primer pairs for PCR to linearize plasmids pET28a-Tfu_0875. The full-length pET28a-Tfu_0875 PCR product containing the mutations were purified by the PCR Product Purification Kit (Sangon Biotech, Shanghai, China) and digested by restriction enzyme *DpnI* (TaKaRa, Dalian, China). Finally, the digested products were transformed into *E. coli* JM109. Primers used in this study were listed in Table S3.

4.3. Protein expression and purification

The *E. coli* BL21(DE3) harboring Tfu_0875 and its mutants were grown on LB medium containing 50 μ g mL⁻¹ kanamycin at 37 °C and 250 rpm overnight as seeds. These seeds were exponentially growing to OD₆₀₀ = 1.0 at 37 °C, and then were treated with 0.8 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) to induce. These cells were cultured for 12–14 h at 25 °C. All cells were collected by centrifugation at 4500g for 5 min and washed with buffer A (50 mM Tris-HCl, pH 7.4, and 200 mM NaCl) twice, and then disrupted by ultrasonication on ice. Then the lysate was centrifuged at 6000 g for 30 min, and the supernatant was filtered by a 0.45 μ m membrane filtration. The collected supernatant was loaded onto Ni-NTA affinity chromatography (5 mL, GE health, Shanghai, China). The target protein was eluted with buffer A (50 mM Tris-HCl, pH 7.4, and 200 mM NaCl) with 20 mM imidazole and buffer B (50 mM Tris-HCl, pH 7.4, 200 mM NaCl, and 200 mM imidazole) [9]. All plasmid schematic diagrams were illustrated in Fig. S1. The protein concentration was determined by Bradford's method (Sangon Biotech, Shanghai, China).

4.4. Enzyme assay

The spectrophotometric method was used to assay the activity of 3-ketoacyl-CoA thiolase by monitoring the increase in absorbance at 412 nm [38]. Briefly, the solution was performed in a 100 μ L total reaction volume containing 50 mM Tris buffer (pH 7.4), 40 mM KCl, 1 mg/mL acetyl-CoA, 1 mg/mL dicarbonyl-CoA, and 5 mg/mL Tfu_0875 or its mutants. The reaction solution was performed at 37 °C for 30 min and then added with 100 μ L of 10 mM DTNB solution. The 3-ketoacyl-CoA thiolase could condense 2 molecules of acyl-CoA to release CoA, whose -SH reacted with DTNB to generate 2-nitro-5-thiobenzoic acid (NTB⁻). The increase in absorbance due to 2-nitro-5-thiobenzoic acid (NTB⁻) formation was registered for 1–3 min at 412 nm.

4.5. Clustering

K-means algorithm [26] is an unsupervised partition-based clustering algorithm, meanwhile, it is one of the most widely used algorithms. It characterizes data using k prototype vectors (or, the centers or centroids of k clusters) and discovers these prototypes using a coordinate descent optimization approach that minimizes a sum-of-squares cost function. The distance between two data elements is used by the k-means algorithm to describe their similarity. When the criterion function approaches the optimum or the maximum number of iterations,

the k-means algorithm is ended. The mutants were characterized by a range of parameters, including enzyme activity improvements, distance to C90, distance to H348, distance to C378, difference in distance to C α of succinyl-CoA and the H bonds with succinyl-CoA. For discriminating individuals, we used the WEKA program [39].

4.6. Molecular dynamics (MD) simulations

The structures of succinyl-CoA and Tfu_0875 were prepared, and subsequently, succinyl-CoA was docked into the binding site. MD simulations of ligand-receptor docked complex were performed using Gromacs MD package [27]. Here, the AMBER99SB-ILDN force field was used to create the protein Tfu_0875 topology file, while the AMBER14SB force field was used by the ACPYPE script to create the ligand succinyl-CoA topology file. The molecular force field for the MD simulation was chosen as OPLS3e (all-atom type force field) [40], and the system was solvated using the TIP3 water model. The system charge is neutralized by adding ions. Energy minimization was achieved using the steepest descent algorithm, with cutoff of 1.0 nm for Coulomb interactions and van der Waals interactions. The coupling of temperature-pressure parameters adopts Berendsen coupling algorithm. Finally, simulations were performed for 100 ns for each system under periodic boundary conditions at 310 K temperature and 1.0 bar pressure. The RMSD (Root Mean Square Deviation) of the backbone atoms was calculated and graphically analyzed to understand the nature of protein-ligand interactions.

4.7. Analytical methods

A spectrophotometer (UV-1800, AOE instruments, Shanghai, China) was used to measure optical density at 600 nm to assess cell proliferation. The fermentation broth was centrifuged at 12000 g for 10 min to obtain the supernatant, then diluted 1:1 by 10 mM H₂SO₄ and filtered with 0.22 μ m membrane filtration (Biosharp, Shanghai, China). Adipic acid was qualitative analysis by liquid chromatography-mass spectrometry, which was run on a Waters MALDI SYNAPT Q-TOF MS (Waters, Milford, Massachusetts) in the negative mode electrospray ionization (ESI⁻) [41]. When using MS/MS, *m/z* 145 was selected as the mass transition ion of adipic acid in multiple-reaction monitor (MRM). Adipic acid in supernatant was quantitative analysis by high performance liquid chromatography (HPLC, Agilent 1260 Infinity II, Santa Clara, CA) equipped with an HPX-87H ion-exclusion column (Bio-Rad, Hercules, CA). 5 mM H₂SO₄ was used as mobile phase with a flow rate of 0.6 mL/min at 35 °C. Adipic acid and other organic acids were determined by Agilent 1260 Variable Wavelength Detector set to 210 nm Fig.6D.

Funding sources

This work was supported by the National Key R&D Program of China (2022YFC2104600), National Natural Science Foundation of China (22378170), the Distinguished Young Scholars of Jiangsu Province (BK20220089), and the Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-KJGG-015).

Notes

The authors declare no competing financial interest.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

Lixia Liu: designed the study and wrote the manuscript, performed the experiments and analyzed the results. **Shuang Liu:** performed the experiments and analyzed the results. **Xiangyang Hu:** performed the experiments and analyzed the results, All authors reviewed, approved, and contributed to the final version of the manuscript. **Shenghu Zhou:** supervised the project, designed the study and wrote the manuscript, critically revised the manuscript. **Yu Deng:** supervised the project, critically revised the manuscript.

Acknowledgment

We thank the staff at BL17B1, BL18U1 and BL02U1 beamlines at SSRF of the National Facility for Protein Science in Shanghai (NFPS), Shanghai Advanced Research Institute, Chinese Academy of Sciences, for providing technical support in X-ray diffraction data collection and analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.synbio.2024.04.014>.

References

- [1] Wackett LP, Wilmot CM. Hydrocarbon biosynthesis in microorganisms. In: Himmel ME, editor. Direct microbial conversion of biomass to advanced biofuels. Amsterdam: Elsevier; 2015. p. 13–31.
- [2] Cheong S, Clomburg JM, Gonzalez R. Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. Nat Biotechnol 2016;34:556–61. <https://doi.org/10.1038/nbt.3505>.
- [3] Jiang C, Kim SY, Suh D-Y. Divergent evolution of the thiolase superfamily and chalcone synthase family. Mol Phylogenet Evol 2008;49:691–701. <https://doi.org/10.1016/j.ympev.2008.09.002>.
- [4] Zhou S, Hao T, Xu S, Deng Y. Coenzyme A thioester-mediated carbon chain elongation as a paintbrush to draw colorful chemical compounds. Biotechnol Adv 2020;43:107575. <https://doi.org/10.1016/j.biotechadv.2020.107575>.
- [5] Blaisse MR, Fu B, Chang MCY. Structural and biochemical studies of substrate selectivity in *Ascaris suum* thiolases. Biochemistry 2018;57:3155–66. <https://doi.org/10.1021/acs.biochem.7b01123>.
- [6] Pereto J, Lopez-Garcia P, Moreira D. Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. J Mol Evol 2005;61:65–74. <https://doi.org/10.1007/s00239-004-0280-8>.
- [7] Tan Z, Clomburg JM, Cheong S, Qian S, Gonzalez R. A polyketoacyl-CoA thiolase-dependent pathway for the synthesis of polyketide backbones. Nat Catal 2020;3:593–603. <https://doi.org/10.1038/s41929-020-0471-8>.
- [8] Liu L, Zhou S, Deng Y. The 3-ketoacyl-CoA thiolase: an engineered enzyme for carbon chain elongation of chemical compounds. Appl Microbiol Biotechnol 2020;104:8117–29. <https://doi.org/10.1007/s00253-020-10848-w>.
- [9] Kim S, Jang Y-S, Ha S-C, Ahn J-W, Kim E-J, Lim JH, Cho C, Ryu YS, Lee SK, Lee SY, Kim K-J. Redox-switch regulatory mechanism of thiolase from *Clostridium acetobutylicum*. Nat Commun 2015;6:8410. <https://doi.org/10.1038/ncomms9410>.
- [10] Xu J, Cen Y, Singh W, Fan J, Wu L, Lin X, Zhou X, Huang M, Reetz MT, Wu Q. Stereodivergent protein engineering of a lipase to access all possible stereoisomers of chiral esters with two stereocenters. J Am Chem Soc 2019;141:7934–45. <https://doi.org/10.1021/jacs.9b02709>.
- [11] Li G, Huang D, Sui X, Li S, Huang B, Zhang X, Wu H, Deng Y. Advances in microbial production of medium-chain dicarboxylic acids for nylon materials. React Chem Eng 2020;5:221–38. <https://doi.org/10.1039/c9re00338j>.
- [12] Reetz MT, Qu G, Sun Z. Engineered enzymes for the synthesis of pharmaceuticals and other high-value products. Nat Synth 2024;3:19–32. <https://doi.org/10.1038/s44160-023-00417-0>.
- [13] Sofeo N, Hart JH, Butler B, Oliver DJ, Yandean-Nelson MD, Nikolau BJ. Altering the substrate specificity of acetyl-CoA synthetase by rational mutagenesis of the carboxylate binding pocket. ACS Synth Biol 2019;8:1325–36. <https://doi.org/10.1021/acssynbio.9b00008>.
- [14] Zhao M, Li G, Deng Y. Engineering *Escherichia coli* for glutarate production as the C5 platform backbone. Appl Environ Microbiol 2018;84:e00814–8. <https://doi.org/10.1128/AEM.00814-18>.
- [15] Zhao M, Huang D, Zhang X, Koffas MA, Zhou J, Deng Y. Metabolic engineering of *Escherichia coli* for producing adipic acid through the reverse adipate-degradation pathway. Metab Eng 2018;47:254–62. <https://doi.org/10.1016/j.ymben.2018.04.002>.
- [16] Bao Q, Zhi R, Zhou S, Zhao Y, Mao Y, Li G, Deng Y. Claisen condensation reaction mediated pimelate biosynthesis via the reverse adipate-degradation pathway and

- its isoenzymes. *Chembiochem* 2022;23:e202200098. <https://doi.org/10.1002/cbic.202200098> (202200091 of 202200098).
- [17] Sui X, Zhao M, Liu Y, Wang J, Li G, Zhang X, Deng Y. Enhancing glutaric acid production in *Escherichia coli* by uptake of malonic acid. *J Ind Microbiol Biotechnol* 2020;47:1–8. <https://doi.org/10.1007/s10295-020-02268-6>.
- [18] Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MKM, Kerkhoven EJ, Nielsen J. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat Catal* 2022;5:662–72. <https://doi.org/10.1038/s41929-022-00798-z>.
- [19] Cui Y, Chen Y, Liu X, Dong S, Tian Ye, Qiao Y, Mitra R, Han J, Li C, Han X, Liu W, Chen Q, Wei W, Wang X, Du W, Tang S, Xiang H, Liu H, Liang Y, Houk KN, Wu B. Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. *ACS Catal* 2021;11:1340–50. <https://doi.org/10.1021/acscatal.0c05126>.
- [20] Liu L, Zhou S, Deng Y. Rational design of the substrate tunnel of β -ketothiolase reveals a local cationic domain modulated rule that improves the efficiency of Claisen Condensation. *ACS Catal* 2023;13:8183–94. <https://doi.org/10.1021/acscatal.3c01426>.
- [21] Lischnig T, P H, S W. Thermostability of endo- β -Xylanase from the thermophilic fungus *Thermomyces lanuginosus*. *Biotechnol Lett* 1993;15:411–4. <https://doi.org/10.1007/BF00128286>.
- [22] Bruins ME, Janssen AEM, Boom RM. Thermozyms and their applications. *Appl Biochem Biotechnol* 2001;90:155–86. <https://doi.org/10.1385/ABAB:90:2:155>.
- [23] Chu W-T, Wang J. Energy landscape topography reveals the underlying link between binding specificity and activity of enzymes. *Sci Rep* 2016;6. <https://doi.org/10.1038/srep27808>.
- [24] Yu H, Ma S, Li Y, Dalby PA. Hot spots-making directed evolution easier. *Biotechnol Adv* 2022;56:107926. <https://doi.org/10.1016/j.biotechadv.2022.107926>.
- [25] Bu Y, Cui Y, Peng Y, Hu M, Tian Ye, Tao Y, Wu B. Engineering improved thermostability of the GH11 xylanase from *Neocallimastix patriciarum* via computational library design. *Appl Microbiol Biotechnol* 2018;102:3675–85. <https://doi.org/10.1007/s00253-018-8872-1>.
- [26] Jain AK. Data clustering: 50 years beyond k-means. *Pattern Recogn Lett* 2009;31:651–66. <https://doi.org/10.1016/j.patrec.2009.09.011>.
- [27] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, David VDS, Hess B, Lindahl E. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 2013;29:845–54. <https://doi.org/10.1093/bioinformatics/btt055>.
- [28] Schoning-Stierand K, Diedrich K, Fahrrolfes R, Flachsenberg F, Meyder A, Nittinger E, Steinegger R, Rarey M. ProteinsPlus: interactive analysis of protein-ligand binding interfaces. *Nucleic Acids Res* 2020;48:W48–53. <https://doi.org/10.1093/nar/gkaa235>.
- [29] Hao T, Li G, Zhou S, Deng Y. Engineering the reductive TCA pathway to dynamically regulate the biosynthesis of adipic acid in *Escherichia coli*. *ACS Synth Biol* 2021;10:632–9. <https://doi.org/10.1021/acssynbio.0c00648>.
- [30] Wu M, Di J, Gong L, He Y-C, Ma C, Deng Y. Enhanced adipic acid production from sugarcane bagasse by a rapid room temperature pretreatment. *Chem Eng J* 2023;452. <https://doi.org/10.1016/j.cej.2022.139320>.
- [31] Waegeman H, Beauprez J, Moens H, Maertens J, De Mey M, Foulquie-Moreno MR, Heijnen JJ, Charlier D, Soetaert W. Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). *BMC Microbiol* 2011;11. <https://doi.org/10.1186/1471-2180-11-70>.
- [32] Bonk BM, Tarasova Y, Hicks MA, Tidor B, Prather KLJ. Rational design of thiolase substrate specificity for metabolic engineering applications. *Biotechnol Bioeng* 2018;115:2167–82. <https://doi.org/10.1002/bit.26737>.
- [33] Falivene L, Cao Z, Petta A, Serra L, Poater A, Oliva R, Scarano V, Cavallo L. Towards the online computer-aided design of catalytic pockets. *Nat Chem* 2019;11:872–9. <https://doi.org/10.1038/s41557-019-0319-5>.
- [34] Wang F, Zhu M, Song Z, Li C, Wang Y, Zhu Z, Sun D, Lu F, Qin H-M. Reshaping the binding pocket of lysine hydroxylase for enhanced activity. *ACS Catal* 2020;10:13946–56. <https://doi.org/10.1021/acscatal.0c03841>.
- [35] Robinson SL, Smith MD, Richman JE, Aukema KG, Wackett LP. Machine learning-based prediction of activity and substrate specificity for OleA enzymes in the thiolase superfamily. *Syn Biol* 2020;5. <https://doi.org/10.1093/synbio/ysaa004>.
- [36] Qu G, Li A, Acevedo-Rocha CG, Sun Z, Reetz MT. The crucial role of methodology development in directed evolution of selective enzymes. *Angew Chem Int Ed* 2020;59:13204–31. <https://doi.org/10.1002/anie.201901491>.
- [37] Reetz MT, Carballeira JD. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* 2007;2:891–903. <https://doi.org/10.1038/nprot.2007.72>.
- [38] Wang ZQ, Song H, Koleski EJ, Hara N, Park DS, Kumar G, Min Y, Dauenhauer PJ, Chang MCY. A dual cellular-heterogeneous catalyst strategy for the production of olefins from glucose. *Nat Chem* 2021;13:1178–85. <https://doi.org/10.1038/s41557-021-00820-0>.
- [39] Frank E, Hall MA, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg LE. Weka—a machine learning workbench for data mining. In: Maimon O, Rokach L, editors. *Data mining and knowledge discovery handbook*. US, Boston, MA: Springer; 2010. p. 1269–77.
- [40] Jorgensen WL, Maxwell DS, Tirado-Rives Julian. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 1996;118:11225–6. <https://doi.org/10.1021/ja9621760>.
- [41] Li L, Chang M, Tao G, Wang X, Liu Y, Liu R, Jin Q, Wang X. Analysis of phospholipids in *Schizochytrium* sp. S31 by using UPLC-Q-TOF-MS. *Anal Methods* 2016;8:763–70. <https://doi.org/10.1039/c5ay02795k>.