Supporting Information

Substrate access path-guided engineering of L-threonine aldolase for improving diastereoselectivity

Wenlong Zheng[a], Zhongji Pu[a], Lanxin Xiao[b], Gang Xu[b], Lirong Yang[a b], Haoran Yu*[a b] and Jianping Wu *[a b]

[a] ZJU-Hangzhou Global Scientific and Technological Innovation Centre, Hangzhou, 311200, Zhejiang, China

[b] Institute of Bioengineering, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, 310027, Zhejiang, China.

Contents

1. EXPERIMENTAL SECTION

1.1 Chemicals, materials and strains

PLP and 2, 4-dinitrophenylhydrazine (DNPH) were obtained from Aladdin. Glycine, sodium hydroxide, 4-methylsulfonyl benzaldehyde (MTB) and *N*, *N*-dimethylformamide (DMF) were purchased from Sinopharm Chemical Reagent Co., Ltd. Restriction enzymes (*Dpn* I)and DNA polymerase (Primer STAR) were obtained from TaKaRa. ClonExpress DNA recombinase (MultiS One Step Cloning Kit) was provided by Vazyme Biotech Co.,

Ltd. L-*syn* (*anti*)-MTPS standards were synthesized in our laboratory. Other reagents and solvents were purchased from J&K Scientific. The gene of LTA from *Leishmania major* was artificially synthesized by Tsingke Biotechnology Co., Ltd. The expression vector and host strain used in the work were pET-28a and *Escherichia coli* BL21 (DE3), respectively.

1.2 Molecular dynamic simulation, mutant building and docking

The crystal structure of *Lm*LTA (PDB ID: 1SVV) was downloaded from RCSB protein data bank (https://www.rcsb.org). Before the docking and molecular dynamic (MD) simulations, the structure was pretreated using the "Prepare Protein" module of Discovery Studio 4.0 by standardizing atom names, nonstandard amino acid substitution, inserting missing atoms or residues, removing alternate conformations, removing water and ligand molecules. Then, the prepared structure was uploaded to the PlayMolecule server (www.playmolecule.com/proteinPrepare) for estimating pKa values of the titratable residues at pH 8.5. Notably, His100 was protonated at the Nε position since the histidine transfer proton with the products, and other titratable residues using default value of the PlayMolecule server. Next, the protonated protein was used as the initial structure of the standardized molecular dynamic (MD) simulations by Amber 20. The process includes the following steps: solvation, adding ions, minimization, heating, equilibration and production MD. The detailed procedure included adding hydrogen atoms and a water box to solvate the protein (forcefield: ff19SB; solvatebox type: OPCBOX), adding counter ions to neutralize the system charge. Steepest descent and conjugate gradient algorithm were used for energy minimization with the maxcycle 5000 (Cycle of switch from steepest descent to conjugate gradient at cycle 1000). In the heating protocol, the temperature ranged from 0 K to 300 K with time step of 2 fs. In the equilibration protocol, NVT and NPT were performed 250 ps in sequence with the time step of 2 fs. In the production protocols, the simulations time was 100 ns with the time step of 2 fs and the output structure conformations were saved with the interval of 2 ps. System is equilibrated if the global

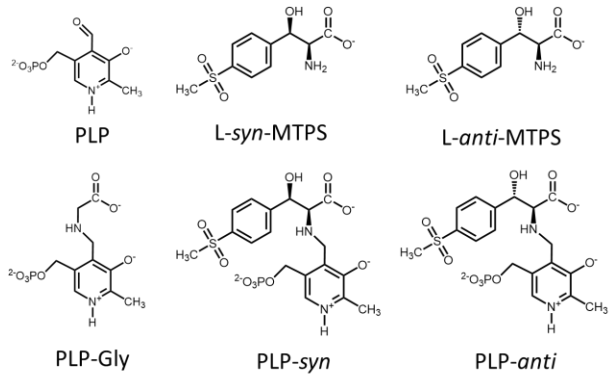
RMSD values are convergent. Down-sampled and RMSD-based cluster analysis were applied in the analysis protocol. The equilibrated structure would be selected for following mutant structure building and molecular docking.

The mutant structure was built by the "Design Protein"module of Discovery Studio with the equilibrated structure of wild type LmLTA as template. Only the specified residues were substituted by the target amino acids. After energy minimization, five mutated candidate conformations were given. Then the conformation with the lowest energy was selected to performed MD for improvement by 100 ns.

Autodock Vina was employed for molecular docking, and the operation methods referred to the tutorial on the official website (Tutorial – AutoDock Vina (scripps.edu)). Based on the catalytic mechanism of LTA, glycine or products would form an aldimine complex with PLP in active center. Aldimine complex namely "intermediate state" was taken as ligand to perform molecular docking[1-2]. In this work, the complex of glycine, L-*syn*-MTPS, L-*anti*-MTPS with PLP were called PLP-Gly, PLP-*syn*,PLP-*anti*, (Scheme S1). PLP-Gly, PLP-*syn*,PLP- *anti* were docked to *Lm*LTA or its mutants as ligands, respectively. Before molecular docking, the ligands were also optimized by adding hydrogen at pH 8.5, and energy minimized by the "Prepare Ligand" module of Discovery Studio. Notably, the hydroxyl group at C_β of PLP-syn or PLP-anti needs to be deprotonated.

A reasonable docked pose with the lowest energy and right conformation was selected to performed MD simulations for improvement. Actually, the protein-ligand complexes were performed for 50 ns-constrained MD simulations (restrained distance of N_ε of His100 and hydroxyl oxygen atom at C_β between 2.7 Å and 2.9 Å) and 50 ns-unconstrained MD simulations. Notably, repeated 50 ns MD simulations with random seed were necessary. Finally, the structure at the last 10ns with converged RMSD value would be selected for further analysis by downsampling. RMSF and Non-bond interactions analyzed by

"Analyze Trajectory" module of Discovery Studio. The RMSD was computed along the trajectories on the protein backbone by "RMSD trajectory tool" module of VMD.



Scheme S1. Chemical structures and intermediate states involved in the paper[1,2]

1.3 PCR-based methods for mutant library construction

In the process of site saturation mutation, the codon of the mutation site was replaced by the NNK degeneracy codon. The primers used in the work are listed in Table S1. The 20 μL PCR reaction mixture was comprised of 0.3 μL template, 1 μL of upstream and downstream primers respectively, 10 μL of Primer STAR, and 7.7 μL deionized water. The PCR was carried out under the following steps: 98 °C 5 min for pre-denaturation; then 30 cycles of 98 °C 30 s, 60 °C 30 s, 72 °C 90 s; finally, 72 °C 10 min for thorough amplification. PCR products were digested by restriction enzyme *Dpn* I at 37°C for 30 min to digest templates. Then, the digested products were transformed into competent cells of *E. coli* BL21 (DE3).

Table S1 primers used in this study

Name	Primers
V19F	CAGCTTC**NNK**AACGACTACAGCGTGG
V19R	GTCGTT**MNN**GAAGCTGTATGGTTTCGG
N20F	GCTTCGTG**NNK**GACTACAGCGTGGGC
N20R	TGTAGTC**MNN**CACGAAGCTGTATGGTTTC
D21F	CGTGAAC**NNK**TACAGCGTGGGCATGC
D21R	GCTGTA**MNN**GTTCACGAAGCTGTATGG

A43F	CCAGCAC**NNK**GGTTATGGCCAGGACA
A43R	ATAACC**MNN**GTGCTGGGTCATGTTG
G44F	GCACGCA**NNK**TATGGCCAGGACAGCC
G44R	GCCATA**MNN**TGCGTGCTGGGTCATG
Y45F	AGGT**NNK**GGCCAGGACAGCCACTGCGCA
Y45R	TGTCCTGGCC**MNN**ACCTGCGTGCTGGGT
R139F	CATGAAAAT**NNK**AGCGAACATATGGTGATTCC
R139R	TGTTCGCT**MNN**ATTTTCATGCAGTGCGCTC
S140F	TGAAAATCGTTTTGAACATATGGTGATTCCGAA
S140R	CATATGTTC**MNN**ACGATTTTCATGCAGTGCG
E141F	ATCGTAGC**NNK**CATATGGTGATTCCGAAACT
E141R	CCATATG**MNN**GCTACGATTTTCATGCAGTG
Y319F	TGATATG**NNK**ACCGTTGAGCCGCTGAAAGATG
Y319R	CAACGGT**MNN**CATATCAAAGTCGTTATTCA
T320F	GATATGTAT**NNK**GTTGAGCCGCTGAAAG
T320R	GGCTCAAC**MNN**ATACATATCAAAGTCGT
V321F	GTATACC**NNK**GAGCCGCTGAAAGATGGTACCT
V321R	GGCTC**MNN**GGTATACATATCAAAGTCGTTAT
E322F	ACCGTT**NNK**CCGCTGAAAGATGGTAC
E322R	AGCGG**MNN**AACGGTATACATATCAAAGTCG

1.4 Cultivation and enzyme solution preparation

primary screening: Clones were picked and transferred into 96-deep-well plates (2.2 mL) containing 500 μL LB liquid medium with 50 μg/mL kanamycin sulfate, and then cultured at 37 °C by 8 h as seeds. Next, 50 μL of the seeds were transferred into a new 96-deep-well plate containing 450 μL LB medium with 50 μg/mL kanamycin sulfate, then cultured at 37 °C. The cells were induced with 0.05 mM IPTG when OD_{600} reached about 0.6. The cells were harvested after induction at 18 °C by18 h. The freeze-thaw lysis method was used in this work for the preparation of the crude enzyme solution. The harvested cells were washed with 500 μL potassium phosphate buffer (50 mM, pH 8.5) two times, and then, the cells were suspended in 250 μL of distilled water by 1 h. Next, the 96-deep-well plate was put into -80 ℃ overnight. Finally, The froze 96-deep-well plate was put into an incubator at 37 ℃ buy 1h for melting, and the crude enzyme solution was prepared.

second screening: The recombinant strains were stored in 15% glycerol at -80 °C. 50 μL of the freezer stocks were inoculated into 5mL LB liquid medium containing 50 μM kanamycin and cultured at 37 °C in an orbital incubator shaker at 200 rpm for 12 h. 1 mL of the cultures were added into 250 mL erlenmeyer flasks containing 50 mL of the LB medium and grown aerobically at 37 °C and 200 rpm in an orbital incubator shaker. The cells were induced with 0.05 mM IPTG when the OD 600 reached about 0.6. The cells were harvested after 18 h of induction at 18 °C. Cells were centrifuged at 1200 rpm for 2 mins. Cell pellets were re-suspended in potassium phosphate buffer (50 mM, pH 8.5), and then disrupted by ultrasonic wave. The cells extracts were centrifuged at 12000 rpm for 10 min, and the crude enzyme solution was prepared. Ni-NTA column affinity chromatography was used for protein purification. LTA were eluted by 250 mM imidazole buffer (pH 8.0). The salt and imidazole were eliminated by ultrafiltration (4000 rpm, 30 min, 3 times). The enzyme solution was stored in 15% glycerol at -80 °C. BCA protein assay kit was used to determine protein concentration.

1.5 Activity assay and diastereoselectivity

High-performance liquid chromatography (HPLC) equipped with a UV detector was applied to perform analysis of substrate and product. The concentration and de *value* of products were measured using a chiral column (CHIRALPAK ZWIX (-); DAICEL CO., LTD, Japan). The chromatographic grade methanol containing 1.9 mL/L formic acid and 2.6 mL/L ethylenediamine was used as mobile phase at a flow rate of 0.5 mL/min. The column was maintained at 40 °C and the UV absorbance at 225 nm was detected. The reaction mixture contains 0.4 M glycine, 20 mM MTB, 50 μM PLP, 5% N, N-dimethylformamide (v/v) and 10% (v/v) crude enzyme solution or 1-2% (v/v) pure enzyme; glycine-NaOH buffer at pH 8.5. Samples were taken every 5 min and 2 % (v/v) 6 M HCl was added to terminate the reaction. The sample was diluted 20-40 folds with methanol. One unit (U) of activity was defined as the amount of enzyme able to synthesize 1 μmol of

L-*syn*-MTPS in one minute. Investigating the corresponding relationship of *de* value and conversion by plotting the graph including the factors of time, *de* value and conversion rate, and tracking the change of *de* value and conversion rate with reaction time. All experiments were conducted in triplicate.

2. RESULT

2.1 Structure evaluation and path hypothesis

The three-dimensional structure of *Lm*LTA is a homologous tetramer, same with the reported LTAs, and each subunit has a small domain and a big domain. Every *Lm*LTA possesses four active centers that are located at the interface of the two domains. Each active pocket consists of several amino acid residues from three adjacent subunits. For example, the active pocket located at subunit D consists of amino acid residues from subunits D, C and A (Fig. S1A). Three different intermediate states including PLP-Gly, PLP-*syn* and PLP-*anti* formed by external aldimine of PLP complex with glycine, L-*syn*-MTPS and L-*anti*-MTPS (Fig. S1) were docked into the *Lm*LTA, resulting in models LTA-Gly, LTA-*syn* and LTA-*anti*, respectively. The models were carried out for 50-ns molecular dynamic (MD) simulations and it was found that with the alignment of ligands PLP-*syn* and PLP-*anti*, the PLP-glycine sections of the two ligands were almost in the same position of the active center, while the orientation of methylsulfonyl phenyl (MTP) groups was distinct (Fig. S1B). The phenomenon consistent with the path hypothesis proposed in our previous work. The hypothesis assumed that the active pocket of LTA has two substrate access paths named *syn* path and *anti* path. L-*syn* configuration products were formed by the substrate aldehyde entering the active center from *syn* path undergo the condensation reaction with PLP-glycine complex. On the contrary, L-*anti* configuration products were formed by the substrate aldehyde entering the active center from *anti* path[1,2]. (Fig. 1, S1).

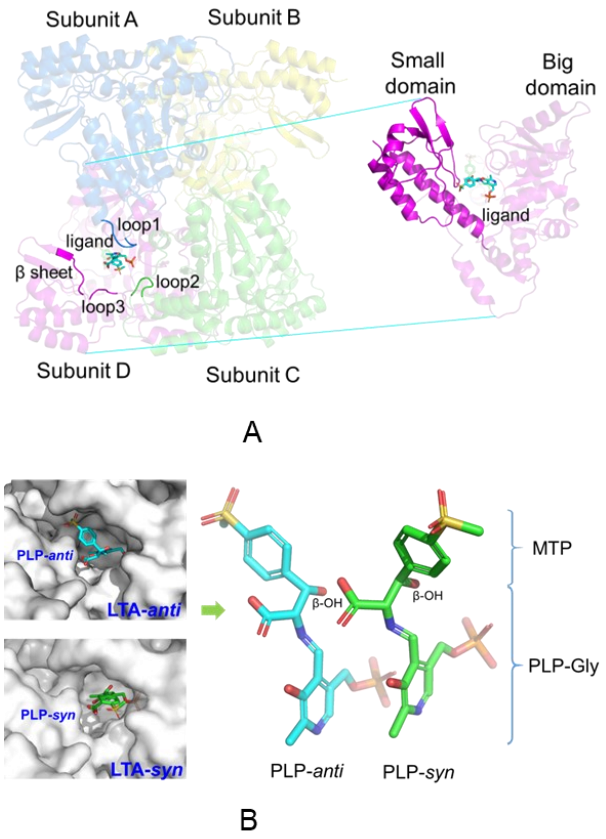


Fig. S1. Models of *Lm*LTA with the docked ligands. **A:** Ribbon representation of *Lm*LTA model with the docked ligand PLP-*syn* colored in cyan. **B:** The alignment of ligands PLP-*syn* and PLP-*anti* from docking models LTA-*syn* and LTA-*anti*, respectively.

2.2 Structure evaluation of *Lm*LTA

The crystal structure of *Lm*LTA (PDB ID:1SVV) was deposited in 2004, but it is still not published. Here, some works have been done to verify reliability of the crystal structure. According to RCSB protein data bank, LTA from 6 different microorganisms have been reported, including LTA (PDB ID: 3WLX) from *Escherichia coli* with resolution of 2.51 Å; LTA (PDB ID: 5VYE) from *Pseudomonas putida* with resolution of 2.28 Å; L-threonine (phenylserine) aldolase (PDB ID: 1V72) from *Pseudomonas putida* with resolution of 2.05 Å; LTA (PDB ID: 1JG8) from *Thermotoga maritima* with resolution of 1.8 Å; LTA (PDB ID: 3WGB) from *Aeromonas jandaei* with resolution of 2.6 Å; LTA (PDB ID: 1SVV, analyzed in this work) from *Leishmania major* with resolution of 2.1 Å. Among them, 3WLX, 1JG8 and 3WGB have been successfully applied in many works including

structure-guided directed evolution, and MD simulations or QM/MM-based catalytic mechanism analysis [3-4]. The crystal structure of 1SVV was compared with 3WLX, 1JG8 and, 3WGB respectively. As a result, the RMSD of structure alignment were 1.67 Å, 1.61 Å and1.74 Å, respectively (Fig S2A, B, C). All of the crystal structures have very similar homotetramer structure. And the family specific positions were highly conserved, especially in active centers and pockets. Besides, Ramachandran plot was constructed to evaluate stereochemical quality of the structure of 1SVV. And the total of residues in most favored regions and additional allowed regions took the share of 99.7% (Fig S2D), which indicated reliability of this crystal structure

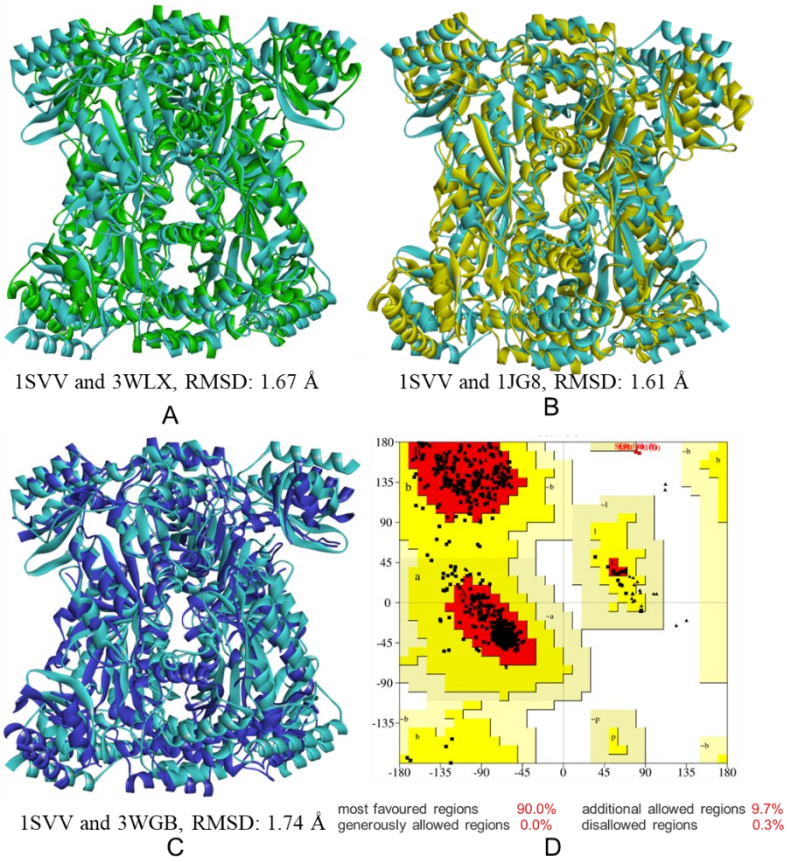


Fig. S2. Structure evaluation of *Lm*LTA (PDB ID: 1SVV). **A**: Structure alignment of 1SVV with 3WLX, 1JG8 and 3WGB, respectively. **B:** The Ramachandran plot of *Lm*LTA

2.3 Improving the diastereoselectivity of LmLTA by sequence alignment

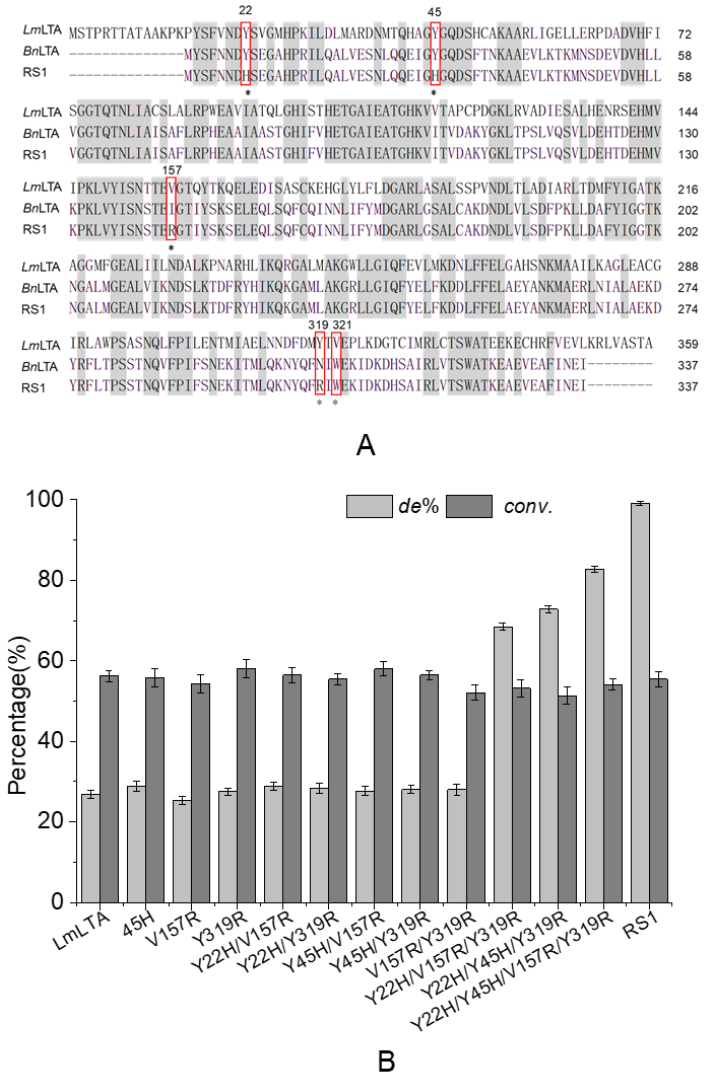


Fig. S3. Improving the diastereoselectivity of *Lm*LTA by sequence alignment. A, Sequence alignment between LmLTA, BnLTA and a mutant RS1 with improved diastereoselectivity. B, Diastereoselectivity of LmLTA mutations constructed at four amino acid residues Y22, Y45, V157 and Y319.

2.4 The property comparison of LmLTA and mutants

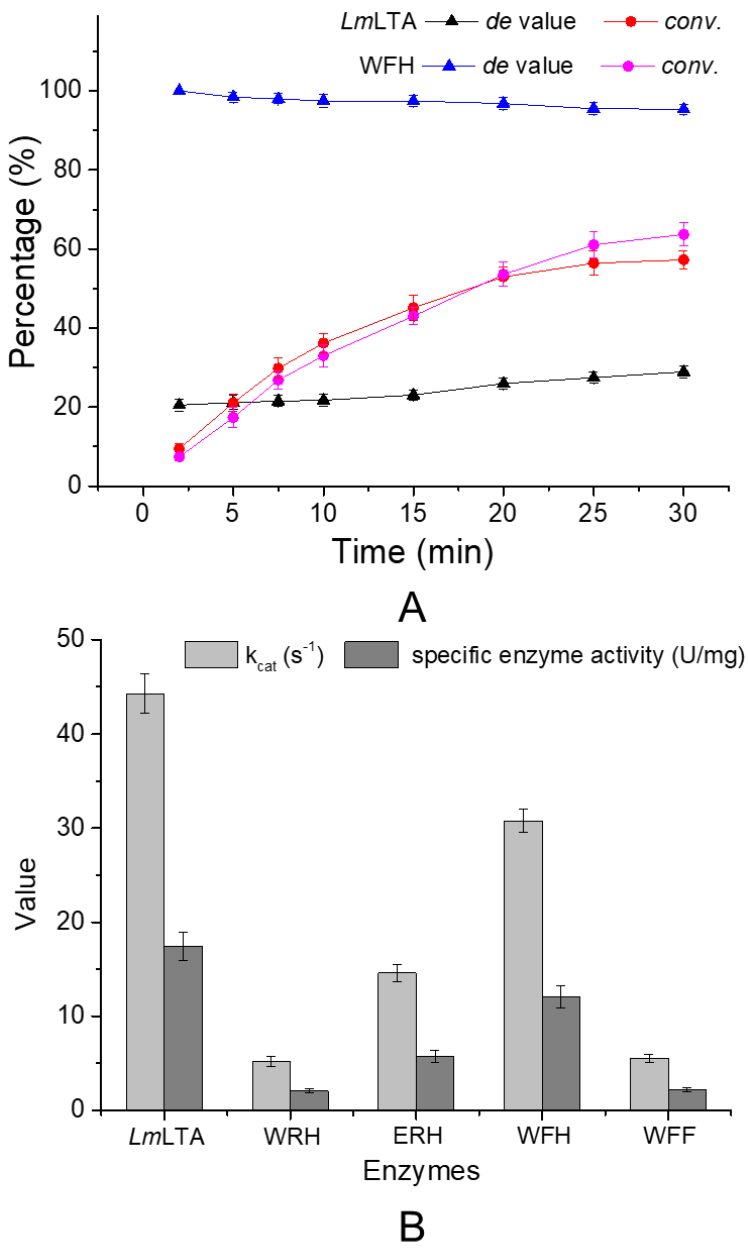


Figure S4. The property comparison of LmLTA and mutants. A: conversion rate and de value comparison of LmLTA and WFH. **B**: Specific enzyme activity and kcat comparison of LmLTA and mutants WRH, ERH, WEH, and WFF. Reaction conditions: 1 mL of Gly-NaOH buffer (100 mM, pH 8.5) containing Gly (0.4 M), aldehyde (20 mM), PLP (50 μM), 5% N, N-dimethylformamide (v/v), and pure LTA 1-2% (v/v).

Table S2 Diastereoselectivity and Conversion Rate Characterization of *Lm*LTA and Mutants

Mutants	*de*$_{(syn)}$	*conv.*	Mutants	*de*$_{(syn)}$	*conv.*	Mutants	*de*$_{(syn)}$	*conv.*
*Lm*LTA	26.8%	53.2%	V321E/Y319A	31.9%	46.5%[a]	V321W/Y319A	30.5%	44.9%[a]
V321E	45.2%	51.5%	V321E/Y319E	39.8%	43.8%[b]	V321W/Y319E	50.3%	56.8%
V321F	30.2%	50.4%	V321E/Y319Q	46.8%	50.2%	V321W/Y319Q	48.4%	54.5%
V321Q	41.4%	53.4%	V321E/Y319F	43.4%	54.3%	V321W/Y319F	76.6%	53.1%
V321R	33.2%	52.8%	V321E/Y319W[a]	50.2%	52.2%	V321W/Y319W	50.6%	46.9%[b]
V321W	56.2%	56.4%	V321E/Y319R	58.8%	50.6%	V321W/Y319R	62.3%	54.3%
V321Y	34.2%	55.9%	V321E/Y319S	40.5%	48.4%[a]	V321W/Y319S	38.2%	43.9%[a]

Reaction conditions: 1 mL of Gly-NaOH buffer (100 mM, pH 8.5) containing Gly (0.4 M), aldehyde (20 mM), PLP (50 μM), 5% N, N-dimethylformamide (v/v), and LTA 10% (v/v). a: reaction time 1h. b: reaction time 2 h. other mutants reaction time 10-30 min

2.5 Molecular dynamic simulation

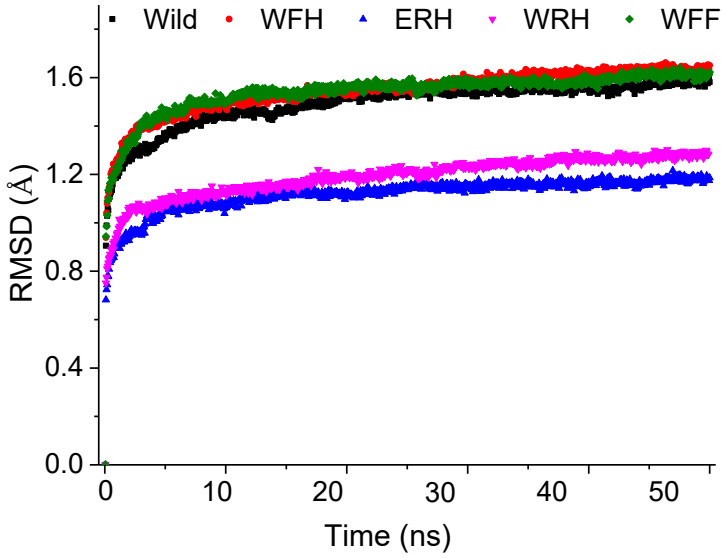


Figure S5. RMSD values of wild type and mutants including WFH, WFF, ERH and WFF with PLP-*syn* as ligands. The RMSD of backbone atoms were calculated using the starting structure of each simulations as a reference.

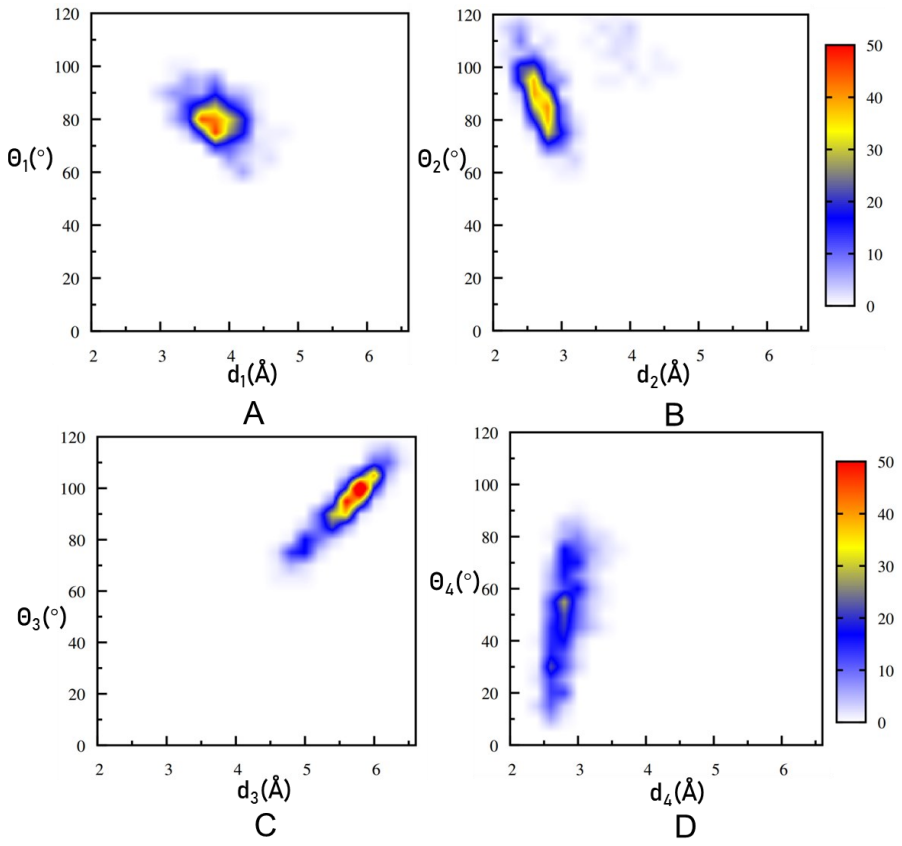


Fig. S6. Statistics of electrostatic interactions and hydrogen bonds in Fig. 3A and 3C, respectively. The axes is distance, and the y-axes is angle. Hydrogen bond would be formed by distance <3.8 Å and the angle between 90°and 180°. The force will be stronger with the angle close to 180° at the same distance. The formation of salt bridge requires distance<5.6 Å and angle> 0°. The magnitude of the electrostatic force depends on the charge distribution on the atoms.

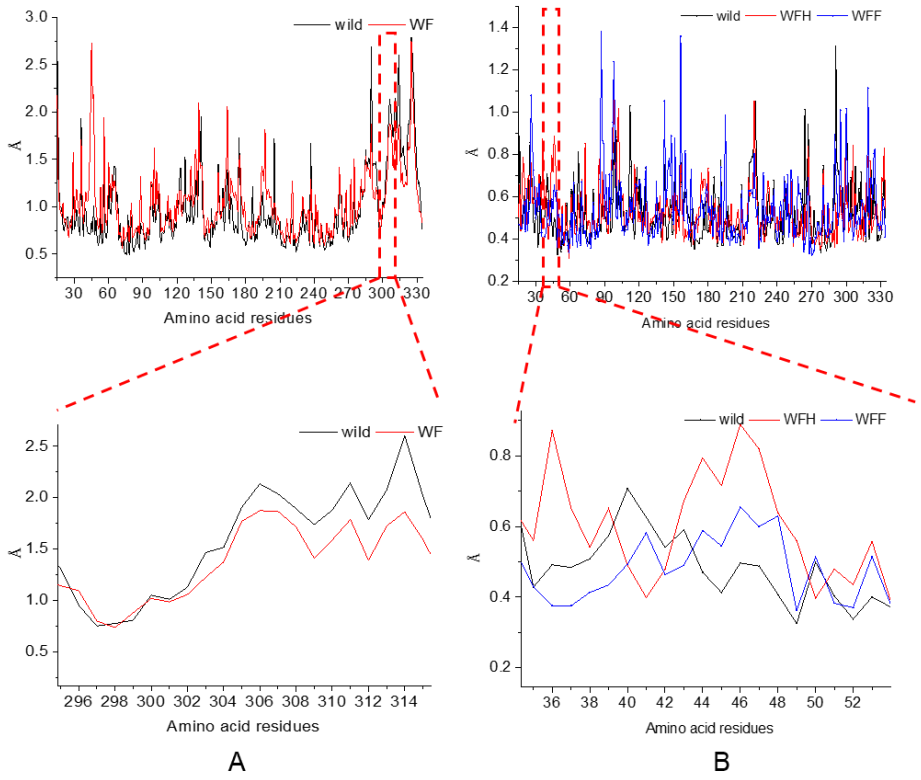


Figure S7. RMSF of model wild-type, WF, WFH and WFF.

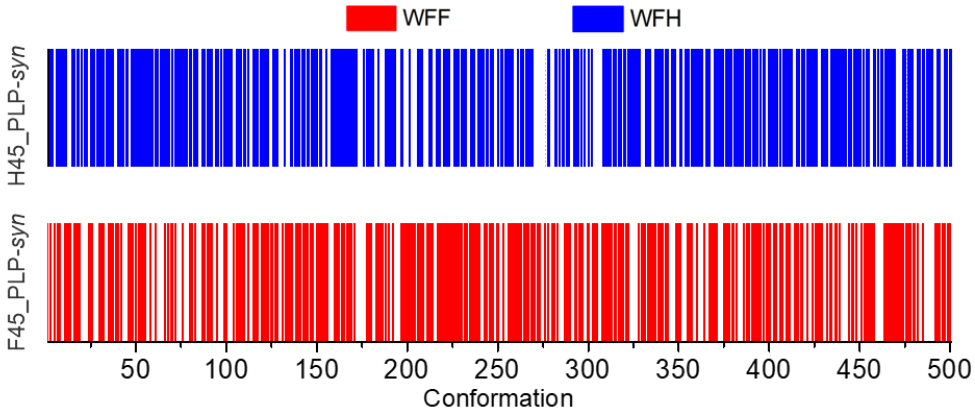


Fig. S8. The occurrence frequency of π−π interaction between ligand PLP-*syn* and residues F45 of WFF, and H45 of WFH, respectively. The x-axes is the conformations extracted from MD simulation, 500 frames in total. The barcodes (red for mutant WFF: F45_PLP-*syn*, blue for mutant WFH: H45_PLP-*syn*) represent the formation of π-π interaction in the corresponding conformation of MD simulation. While the blanks represent that no π-π interaction was formed in the corresponding conformation of MD simulation.

2.6 Result of HPLC and LCMS

Liquid phase mass spectrometry analysis results of L-*syn*-MTPS, L-*anti*-MTPS, referring to our previous work[1].

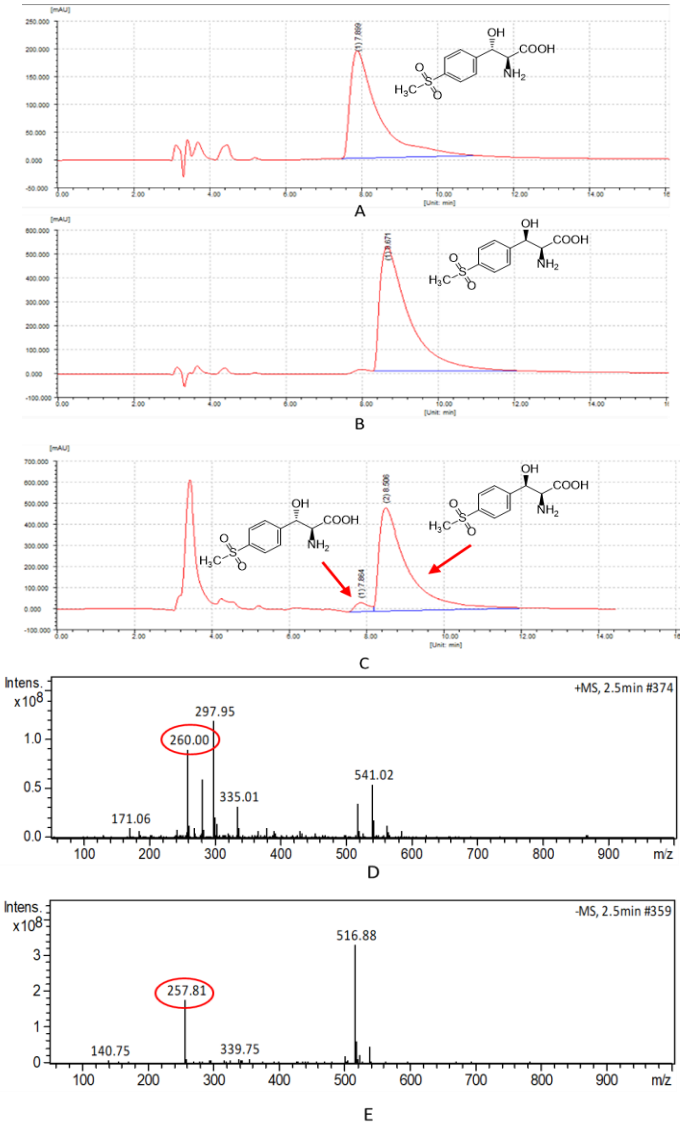


Figure S9. HPLC and MS analysis of L- MTPS (1a). (A) HPLC analysis of standard L-*anti*-MTPS, the retention time is 7.899 min. **(B)** HPLC analysis of standard L-*syn*-MTPS, the retention time is 8.671 min. **(C)** HPLC analysis of reaction solution catalyzed by LTA, the retention time of L-*anti*-MTPS and L-*syn*-MTPS are 7.864 min and 8.506 min, respectively. **(D)** MS analysis of positive ion source (260.00, M+H). **(E)** MS analysis of negative ion source (257.81, M-H). Relative molecular mass of L-MTPS is 259.28 (M).

Reference

1. W. Zheng, H. Yu, S. Fang, K. Chen, Z. Wang, X. Cheng, G. Xu, L. Yang and J. Wu, *ACS Catal.*, 2021, 3198-3205.

2. W. Zheng, K. Chen, Z. Wang, X. Cheng, G. Xu, L. Yang and J. Wu, *Org. Lett.*, 2020, **22**, 5763-5767.

3. J. F. Rocha, S. F. Sousa and N. M. F. S. A. Cerqueira, *ACS Catal*. 2022, **12**, 4990−4999

4. L. Wieteska, M. Ionov, J. Szemraj, C. Feller, A. Kolinski and D. Gront, *Biotechnol. J.*, 2015, **199**, 69-76.

5. M. L. di Salvo, S. G. Remesh, M. Vivoli, M. S. Ghatge, A. Paiardini, S. D'Aguanno, M. K. Safo and R. Contestabile, FEBS journal, 2014, 281, 129-145.