Electronic Supplementary Material (ESI) for Catalysis Science & Technology.

Computational Protein Design to Improve Thermostability of Lipase from *Pseudomonas alcaligenes* by the CREATE Strategy

Zhonglang Yu[#a], Haoran Yu[#ab], Jinling Xu[a], Zhe Wang[ac], Ziyuan Wang[a], Tingting Kang[a], Kaitong Chen[a], Zhongji Pu [ab], Jianping Wu[ab], Lirong Yang[ab], Gang Xu*[a]

[#] Equal contribution

[a] College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China

[b] Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311200, China

[c] Huadong Medicine Co., Ltd, Hangzhou 310011, China

Contents

Experimental Methods

Plasmid, site-directed mutagenesis and purification of enzyme. The plasmid pET-30a (+) was used to over-express *Pa*L in *E. coli* BL21 (DE3). The primers for site-directed mutagenesis were designed using software CE Design V1.04, and the site-directed mutagenesis was completed by PCR. After culturing 100 mL of *E. coli* in a shaker at 37 ℃ 200 rpm to OD 1.2, then added isopropyl-β-D-thiogalactopyranoside (IPTG) to induce protein expression for 20 h at 18 ℃. The *Pa*L was purified by using Ni-NTA resin.

Measurement of relative enzyme activity, residual enzyme activity and half-life at 50 ℃. Adding 0.1 mg of pure enzyme to 100 μL 50 mM *d,l*-menthol propionate, dilute to 2 mL with pH 9.0 Gly-NaOH, and react at 35°C for 20 minutes. Mixing 100 μL of the reaction solution with 100 μL of n-hexane, and stand for extraction for 30 min. Afterwards, the supernatant was detected by a gas chromatograph (GC). The GC detector is FID, the chiral column used is CP-CYCLODEXTRIN β-2, 3, 6-M-19, FS50X.25, and the mobile phases are 30 mL/min N_2, 30 mL/min H_2 and 300 mL/min Air. The reaction amount of *L*-menthol propionate reacted at 35 °C for 20 min was calculated as 100%, and the calculation formula was as follows:

$$Rn = \left(\frac{S_{L0}}{S_D} - \frac{S_{Lt}}{S_D}\right) \times \frac{n}{2}$$

Here, S_{L0} is the initial peak area of *L*-menthol propionate; S_{Lt} is the peak area of *L*-menthol propionate after 20 min of reaction; S_D is the area of *D*-menthol propionate, and *n* is *D, L*-menthol added amount (mmol). All measurements were performed in triplicate, and the data are displayed as mean ± s.d.

The conversion rate of *L*-menthol propionate catalyzed by wild type was measured as 100%, and the relative enzyme activity of the variant was measured. After the pure enzyme was heated at 50°C

for 1 h, the ratio of the conversion rate of l-menthol propionate to the unheated conversion rate of the variant was measured to obtain the residual enzyme activity. After heating the pure enzyme at 50 °C for 0 min, 30 min, 60 min, 90 min and 120 min, respectively, the reaction amount of *L*-menthol propionate was measured. The deactivation kinetics of PaL was fitted to equation $\ln(E/E0)=-k_d*t$ and half-life at 50 ℃ was calculated using the equation $t_{1/2}=\ln(2)/k_d$。

The measurement of T_m. Using a capillary tube to take 1 mg/mL into the tray and use nano-DSF (Prometheus NT.48) for detection. The original data results are shown in Figure S4.

The change in free energy of inactivation relative to WT ($\Delta\Delta G^{\neq}$). The free energy of inactivation was calculated using the following equation, where RA_V is the residual enzyme activity of the variant, and RA_{WT} is the wild-type residual enzyme activity:

$$\Delta\Delta G^{\neq} = RTln(\frac{\ln(RA_{WT})}{\ln(RA_V)})$$

MD Simulations. The *Pa*L 3D structure obtained previously was used as starting structure of MD simulations (*ChemCatChem* 2021, 13, 2691.). The system was solvated using TIP3P water model and then neutralized by adding Na+. The ff14SB force field was used to describe all protein parameters. Sander program in AmberTools21 was used for classical molecular dynamics simulations. The system was minimized by using the steepest method to optimize 5000 steps, and then using conjugate gradient to optimize 5000 steps. Then the system was heated to the target temperature of 300K for a period of 20 ps in constant pressure periodic boundary conditions (NPT). The SHAKE algorithm was used in which all bonds involving hydrogen are constrained. And a

standard cut-off of 10 Å was used for non-bonded interactions. Forty nanosecond production simulations were performed in triplicate under the NPT ensemble at 300 K with the time step of 2 fs. Based on the analysis of root-mean square deviation (RMSD) by AmberTools21, the last 10-nas trajectories were used for analysis of root-mean square fluctuation (RMSF). The RMSD of backbone atoms was calculated with respect to the starting structure, and the RMSF was calculated for each residue of wild type and mutant *Pa*Ls.

Dynamics Cross-Correlation Matrix. DCCM calculation is based on the following formula. The covariance *c*(i,j) is calculated by:

$$c_{i,j} = < \Delta r_i \cdot \Delta r_j >$$

The cross-correlation coefficient is calculated by:

$$C_{i,j} = \frac{c(i,j)}{[c(i,j)c(i,j)]^{1/2}} = \frac{< \Delta r_i \cdot \Delta r_j >}{< \Delta r_i^2 >^{1/2} < \Delta r_j^2 >^{1/2}}$$

Bio 3D was used to calculate DCCM by analyzing 10-ns MD simulations trajectory after stabilization. The output the correlation matrix of Cα was plot by OriginPro 9.0.

Salt bridges and hydrogen bonds analysis

Analysis of salt bridges was carried out with the visual molecular dynamics (VMD) program. The MD simulation trajectories were used as the input and a salt bridge is considered to be formed if the distance between any of the oxygen atoms of acidic residues and the nitrogen atoms of basic residues are within the cut-off distance 4 Å in at least one frame. The protein interactions calculator (PIC) server (http://pic.mbu.iisc.ernet.in/index.html) was used to calculate hydrogen bonds for *Pa*L wild type and variants. VMD version 1.9.3 was used to calculate the distance of hydrogen bonds in the

MD simulations trajectories according to the calculated results of PIC.

DNA Sequence and Amino Acid Sequence

DNA Sequence of *Pa*L WT

ATGCACCATCATCATCATCATTCTTCTGGTCTGGTGCCACGCGGTTCTGGTATGAAAG
AAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCCCAGATCTGGGTACCG
ACGACGACGACAAGGCCATGGCTGATATCGGATCCGAATTCATGGCGACGGTCAAGA
CCACCCATCGCACCATTGCCGGCTGGGATGGAACGCCTCTCGGGGCCTTCGTCATCGA
ACCGCAGGATGCGGGCGGCGGGCGCTATCCGCTGCTGGTGATGCCCAGCAGCTGGGC
GGTGCCCAGCGTGGAGTACGTGGGGGTGGCGCAGTCGCTGGCACAGCGCGGCTATGT
GGTGATCAGCTACAGCTCGCGTGGCTTCTGGGAATCCGGCGGCAGCATCGACATCGC
CGGGCCCTCCACGGTGGAGGATGTCAGCGCGCTGATCGACTGGGCGCTGGACAACAC
CCGCGCTGACCCGGACCGGATCGGCGTTTCGGGCATCTCCTACGGTGCGGGCACGAG
CCTGCTGGCGGCGGCACGCGATCCGCGCATCAAGGCCGTGGCCGCGCTCAGTGGCTG
GGCCGATCTGCAGGCGTCGCTGTACAGCAACGACACCCCGAGCGCGCAGGGCATCGC
ACTGCTGGTTGCGGCCGGCCTGGTGACCGGACGCCCTGGCGCGGAGCTGGCCACGAT
CAACCGCAATGTTCTGGCGGGCAACTACCAGGGCGCGGTGGATTCGCTGTTGCCGGT
GGCGGCGCAGCGCAGTCCCGCGGCCAGCATCGACGAGATCAACGCCAACCAGCCGG
CGGTGTTCCTGGCCAACGCCTTCAACGACAGCCTGTTCCCGCCGGGCCAGCTGGTTGA
TTTCTTCAACCGGCTGAAAGGCCCGAAGCAGCTGCAGCTGCGCCACGGTGACCACGC
GCTCAACGAAGCGCTCGGTGCACTGGGCATTCCCAACGAAGTCTATGACCAGGTGGG
TGACTGGTTCGATCACTATCTGAAGGCAGTGGCCAACGGCATCGACCGCCAGCCGGC
GGTGCAGCTGAAGTCACAGAAGGGCAGCTGGAGCAGTTACCCGGACTGGCAGGCGA
CCAGCAAGGGTGCGGTCAGCTACGGCCTGACCGCACCCTCTGGTTTGCTGCTGCCGAC
CGGCGGCCTGGCCGAGCACGGTGGCGGCACCGGCTGGAACTACCGCATCGGCAGCGG
TTTGCTGACTGCGGCCAATTCCGGCGTGGCGATGGCCTCTGGCGCGCTGCAGATGATC
AACCTGCCGCCGGGTGCCTACGTACCGTTTGTTGGCCGCAGTGCCGCCGGTGTCTGGC
AGGGGCCGATCCAGTGGAGCGCCAAACGCCTGGACGGTGCGCCGGAAGTGCGCCTG
ACGGTCACGCCCAGCCGTGCCAACACCACGCTGTACGCCTACCTGTACGCCGAGGAT
GTGCTGGGCAATGGTCAGTTGATCAGCCACAAGCCGTACACCTTGCGCGGTGCCACG
CCGGGCCAGGCGAAAACGCTCGACCTGCGGCTGGAAGCGAGCAGCTGGAACCTTCCA
GCCGGCAGCCGCCTGACCTTGGTGGTCGATACCGTGGACCTGCGCTATGCGGGCATC
AGCCAGCTGGGTGGGGCGGTGACCTTCACCTCGCCAGCCAATGCGCCGTCGGTGCTG
AAAGTGCCGCTGCATTGA

Amino Acid Sequence of wild type

MATVKTTHRTIAGWDGTPLGAFVIEPQDAGGGRYPLLVMPSSWAVPSVEYVGVAQSLA

QRGYVVISYSSRGFWESGGSIDIAGPSTVEDVSALIDWALDNTRADPDRIGVSGISYGAGT

SLLAAARDPRIKAVAALSGWADLQASLYSNDTPSAQGIALLVAAGLVTGRPGAELATINR

NVLAGNYQGAVDSLLPVAAQRSPAASIDEINANQPAVFLANAFNDSLFPPGQLVDFFNRL

KGPKQLQLRHGDHALNEALGALGIPNEVYDQVGDWFDHYLKAVANGIDRQPAVQLKSQ

KGSWSSYPDWQATSKGAVSYGLTAPSGLLLPTGGLAEHGGGTGWNYRIGSGLLTAANS

GVAMASGALQMINLPPGAYVPFVGRSAAGVWQGPIQWSAKRLDGAPEVRLTVTPSRAN

TTLYAYLYAEDVLGNGQLISHKPYTLRGATPGQAKTLDLRLEASSWNLPAGSRLTLVVD

TVDLRYAGISQLGGAVTFTSPANAPSVLKVPL

Amino Acid Sequence of 4M variant

MATVKTTHRTIAGWDGTPLGAFVIEPQDAGGGRYPLLVMPSSWAVPSVEYVGVAQSLA

QRGYVVISYSSRGFWESGGHIDIAGPSTVEDVSALIDWALDNTRADPDRIGVSGISYGAGT

SLLAAARDPRIKAVAALSGWADLQASLYSNDTPSAQGIALLVAAGLVTGRPGAELATINR

NVLAGNYQGAVDSLLPVAAQRSPAAYIDEINANQPAVFLANAFNDSLFPPGQLVDFFNRL

KGPKQLQLRHGDHALNEALGALGIPNEVYDQVGDWFDHYLKAVANGIDRQPAVQLKSQ

KGSWSSYPDWQATSKGAVSYGLTAPSGLLLPTGGLAEHGGGTGWNYRIGSGLLTAANS

GVAMASGALQMINLPPGAYVPFVDRSAAGVWQGPIQWSAKRLDGAPEVRLTVTPSRAN

TTLYAYLYAEDVLGNGQLISHKPYTLRGATPGQAKTLDLRLEASSWNLPAGSRLTLVVD

TVDLRYAGISQLGGAVTFTSPANAPSVLKVPL

CREATE strategy

Note 1. Identification of non-target regions

Considering the mutations in the functional regions might influence catalytic efficiency of *Pa*L, we applied MD simulations to identify the functional regions as nontarget residues. It has been reported that regions maintaining dynamics correlation with active sites also played important roles in catalysis. The dynamics cross-correlation map (DCCM) of the wild-type *Pa*L was hence computed using MD simulations (Figure 1), which shows the correlation coefficients (Cij) indicating the extent to which the fluctuation of an atom is correlated (or anticorrelated) with one other atom. The catalytic triad of *Pa*L is Ser114-Asp224-His252 and the DCCM showed that regions including G112-I130 and F222-N255 were strongly correlated with the catalytic triad (Figure S2). These regions could be considered to be highly related to the active center and were hence selected as non-target residues.

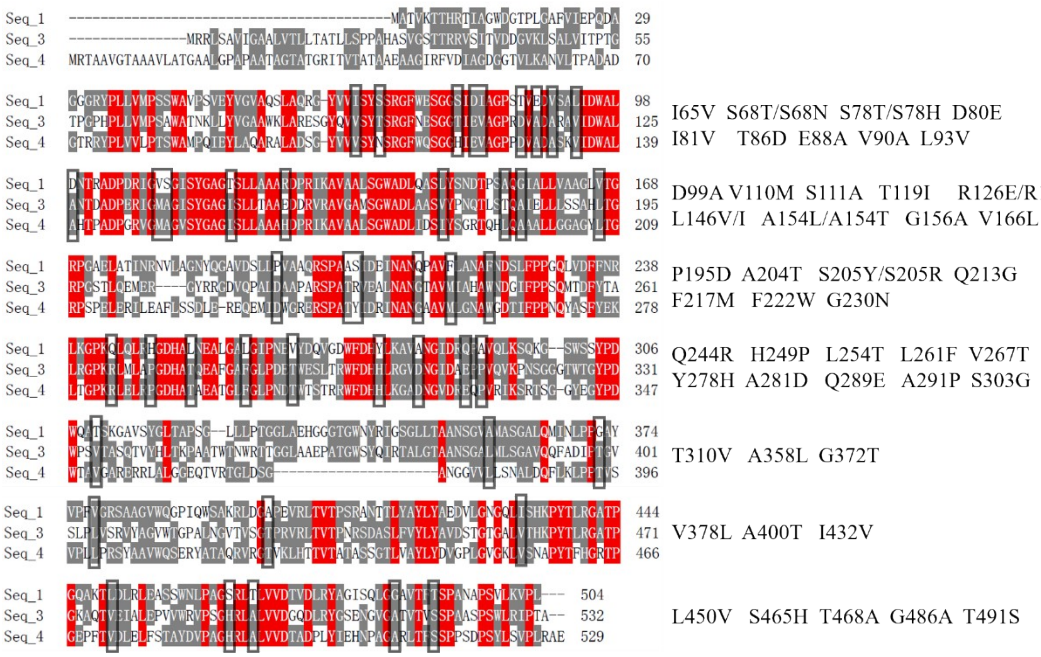


Figure S1. Sequence alignment between *Pa*L and its thermophilic homologous. The box indicates the mutations suggested.

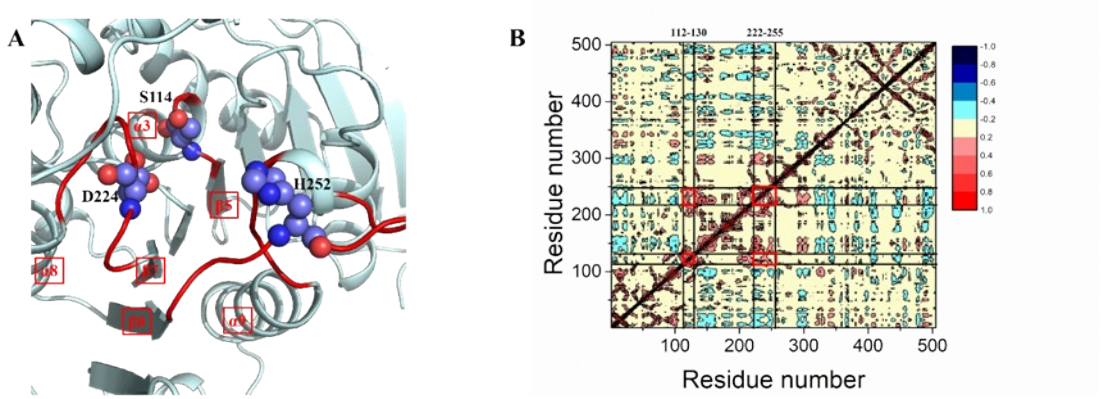


Figure S2. The catalytic triad of *Pa*L and its dynamics correlated regions. (A) The three-dimensional structure of the *Pa*L catalytic triad Ser114-Asp224-His252. The loop area is indicated by a red line. (B) Dynamics cross-correlation map for the C_α atom pairs of WT. Correlation coefficient (C_{ij}) was shown as different colors. C_{ij} with values from 0 to 1 represents positive correlations, whereas C_{ij} with values from −1 to 0 represents negative correlations.

Table S1. Sequence alignment library

Entry	Enzyme source	Homology (%)	NCBI Accession
1	CocE/NonD family hydrolase [*Thermobifida cellulosilytica*]	51.01	WP_068753090.1
2	Acyl esterase [*Streptomyces thermolilacinus* SPC6]	47.59	OEJ93670.1

Table S2. Stable mutations predicted by three different methods

Methods	**Mutations predicted**
Sequence alignment with thermophilic	I65V S68T S68N[a] S78T S78H D80E I81V T86D E88A V90A L93V D99A V110M S111A T119I R126E R126H L146V L146I A154Q A154T G156A V166L P195D A204T S205Y S205R Q213G F217M F222W G230N Q244R H249P L254T L261F V267T Y278H A281D Q289E A291P S303G T310V A358L G372T V378L A400T I432V L450V S465H T468A G486A T491S **52 in total**
PROSS	A12P Y34F L37I S47N G52A V53Q S56K Q59K S68N S69P E74D S78E I81L S85D E88A L93V D99A N100H D106S V110M S111A R126H Q143V S148P D150E S153H G156A A162L L165N V166L A172P A175Q T176Q N178Q N180D A183D N185D S192E V196W A198P Q199E A204T S205Y E208R G230S N237Q K240T Q244 L258T I263L Y268W D273R A281G A283D Q289E K298N Q308A T310L S311Q K312A A314R V315R S316T N342S R344T L350D A360V A363L L369V G372T V375I F377L V378I G379D G384A I390V Q391M A394T K395Q V403L N413D Y417F A422D E423V V425A Q430R A447P K448F A456P S457T N460D S465H A479Q I481H L484P N495D V499T **97 in total**
FireProt	W14Y E25T Q27A D28G G32P M39L S41T V45W S47Q S56K S68N E74Q S78E D80E S85P E88A R102P S111A Q143I S148L G156A V166L N178I V181F N185Y Q199K S205Y Q213G N223Y S225A G230N N237E A260L I263L E266D V267L D269T V271I D273R A281G A283D Q289E A291P S300G S302E S303G Q308K A309S T310V S311M K312A G318A G332P S347T T351S A352G S355M A358W A360L Q365D I367F A373V F377L V378L G379Q G384A Q387L Q391M K395Q A400Q V403L E423V V425A N428Y Q430K A447P K448F A456P S458A W459Y N460W S465H I481H Q483P A487W V488L T491S A496D V499Y **88 in total**

[a]The highlighted mutations were predicted by two or three methods.

Characterization of variants

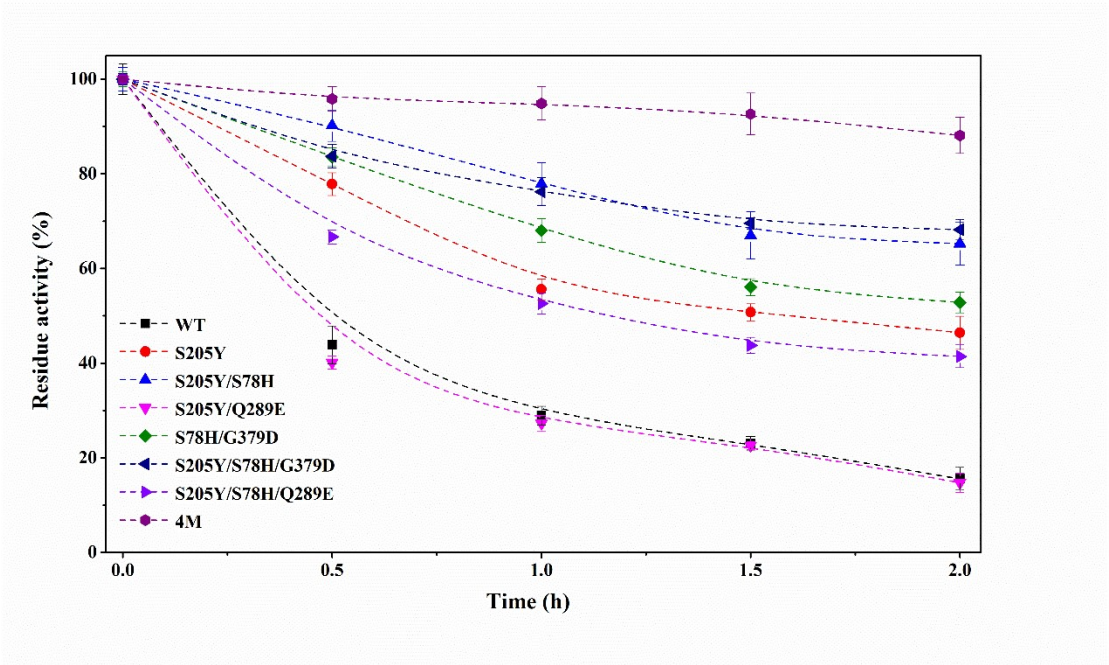
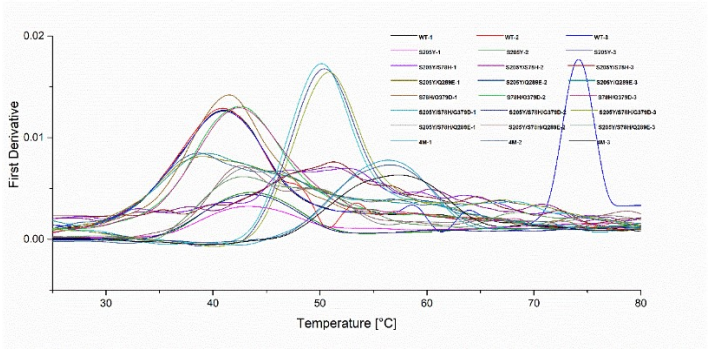


Figure S3. Half-life measurement of wild type and variants at 50 ℃.



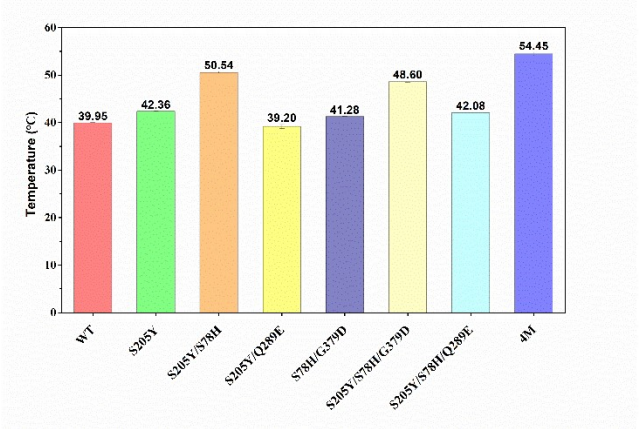


Figure S4. Determination of T_m with Nano-DSF

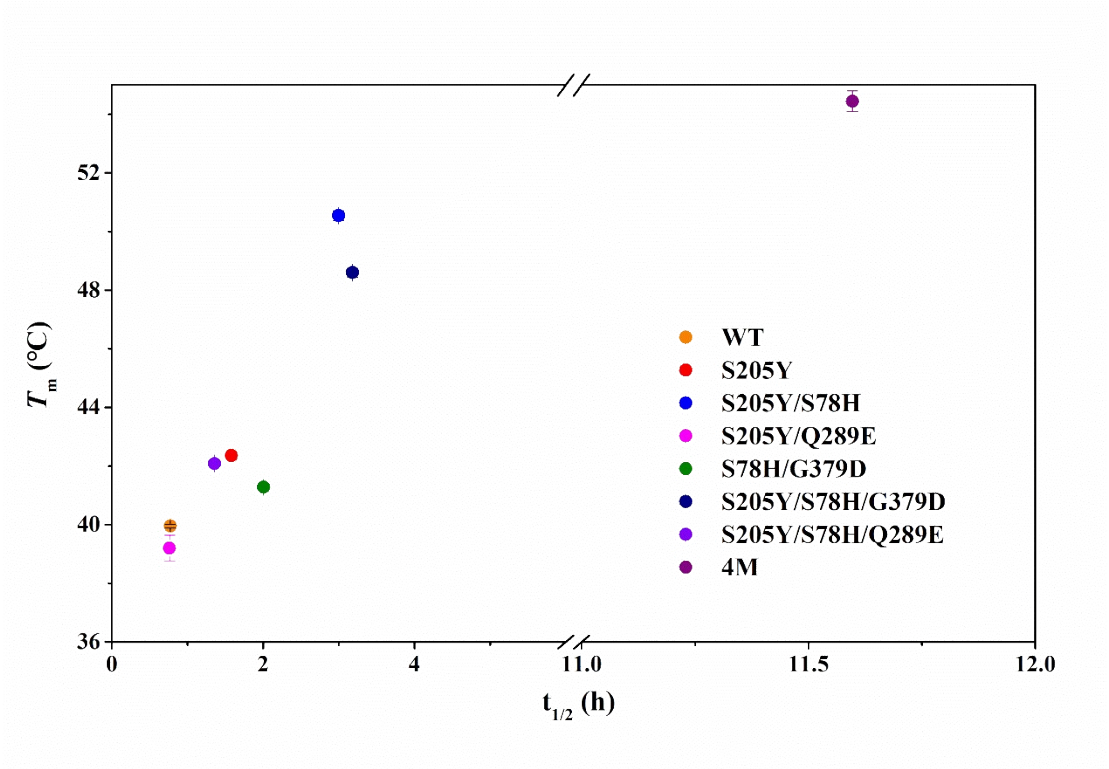


Figure S5. The correlation between T_m and $t_{1/2}$.

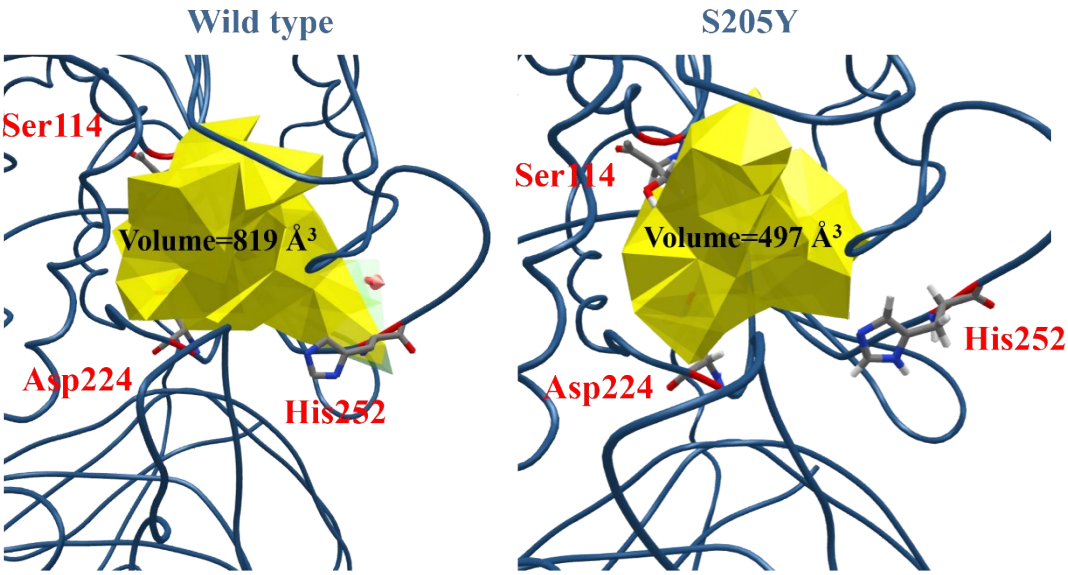


Figure S6. The volume of active center. Wild type: 819 Å^3; S205Y: 497 Å^3. Decreased flexibility of the active center decrease the volume, which might limit the binding of substrate or release of product

Analysis of molecular dynamics simulation trajectories

Table S3. Salt bridges with occupancy higher than 20% in the MD simulations trajectories

Rank[a]	Salt bridges	WT		4M	
		Average distance (Å)	Occupancy	Average distance (Å)	Occupancy
1	ASP15-ARG70	2.8	100%	2.8	100%
2	ASP89-ARG70	3.7	100%	3.7	99.9%
3	ASP95-ARG129	3.6	99.2%	3.5	99.8%
4	GLU455-LYS435	3.5	97.1%	3.7	95.3%
5	GLU266-LYS295	3.1	88.0%	3.2	90.5%
6	ASP127-ARG129	3.9	80.0%	3.9	85.2%
7	ASP398-ARG396	3.9	68.9%	3.9	70.6%
8	ASP276-LYS280	3.9	68.6%	3.9	72.9%
9	ASP451-ARG404	4.5	61.7%	5.1	25.4%
10	ASP273-ARG60	4.1	50.7%	4.0	64.5%
11	ASP234-ARG453	5.4	43.9%	4.8	56.4%
12	ASP287-ARG288	5.3	26.9%	5.5	31.0%

[a] Rank based on salt bridge occupancy of wild type

Table S4 Hydrogen bonds identified by PIC server for wile type and 4M variant

Position	WT	4M
205	S205(N)-P202(O)[a]	Y205(N)-A203(O)
	S205(O)-E208(N)	Y205(O)-D207(N)
	S205(O)-I209(N)	Y205(O)-E208(N)
	S205(O)-P202(O)	Y205(OH)-A125(N)[b]
		Y205(OH)-L121(O)
		Y205(OH)- L122(O)
289	Q289(N)-I286(O)	E289(N)-G285(O)
	Q289(O)-Q244(N)	E289(N)-I286(O)
	Q289(O)-Q270(N)	E289(O)-A291(N)
	Q289(NE2)-G285(O)	E289(OE1)-Q270(NE2)[c]
		E289(OE1)-Q293(NE2)
379	G379(N)-F377(O)	D379(N)-F377(O)
	G379(O)-S381(N)	D379(O)-S381(N)
	G379(O)-A382(N)	D379(O)-A382(N)
	G379(O)-A383(N)	D379(O)-A383(N)
		D379(OD1)-S381(N)
		D379(OD1)-S381(OG)
		D379(OD2)-S381(OG)
78	S78(O)-D80(N)	

[a] Hydrogen bonds formed between main-chain and main-chain

[b] Hydrogen bonds formed between side-chain and main chain

[c] Hydrogen bonds formed between side-chain and side-chain

The topological characteristics and Correlation coefficients (C_{ij}) of mutation points

We annotated and abstracted the four mutation positions in a three-dimensional structure into a ‘pyramid’, with residue 379 (∠C) as the apex and residues 78 (∠B), 205 (∠A) and 289 (∠D) as the bottom triangle (Figure S5). The last 10-ns MD simulation trajectory was applied to analyze the distances between each two residues, angles formed by lines linking residues and dynamics cross correlation coefficients between two residues. The arrow of pyramid in the figure was determined by the direction of movement in the DCCM, and the coloring was determined by the value of heat map (Figure S7 and Table S5)

Table S5. Correlation coefficients (C_{ij}) of mutation points.

Residue	WT	S205Y	2M-1	2M-2	2M-3	3M-1	3M-2	4M
289-78	0.0835	0.1488	0.0688	0.075	0.1239	0.0646	0.1012	0.1123
289-205	0.0808	0.1254	0.0479	0.016	0.00975	0.1136	0.0528	0.0560
289-379	0.1837	0.0254	0.0893	0.053	0.0081	0.0790	0.1043	0.0632
205-78	0.0260	0.0836	0.0747	0.150	0.0528	0.0357	0.0846	0.1373
205-379	-0.085	0.0275	0.0666	0.026	0.0237	0.0369	0.0329	0.0379
78-379	0.1075	0.1744	0.0949	0.050	0.1373	0.1222	0.1372	0.0544

*2M-1: S205Y/S78H; 2M-2: S205Y/Q289E; 2M-3: S78H/G379D; 3M-1: S205Y/S78H/G379D; 3M-2: SS205Y/S78H/Q289E

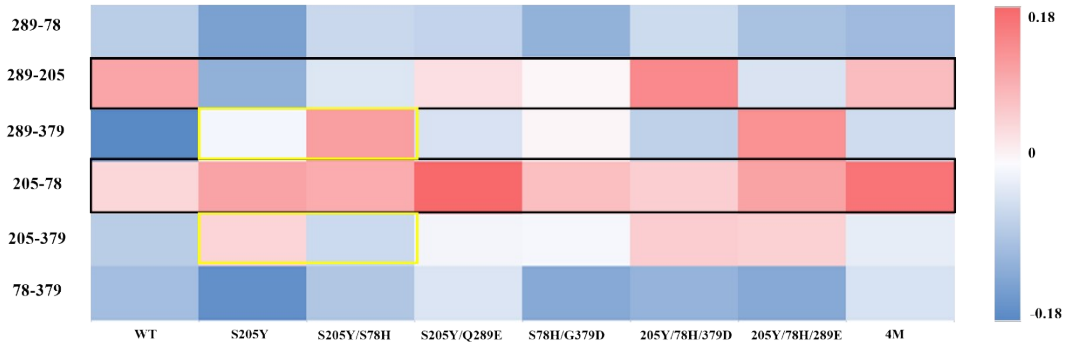


Figure S7. Heat map of Correlation coefficients (C_{ij}) of mutation points.

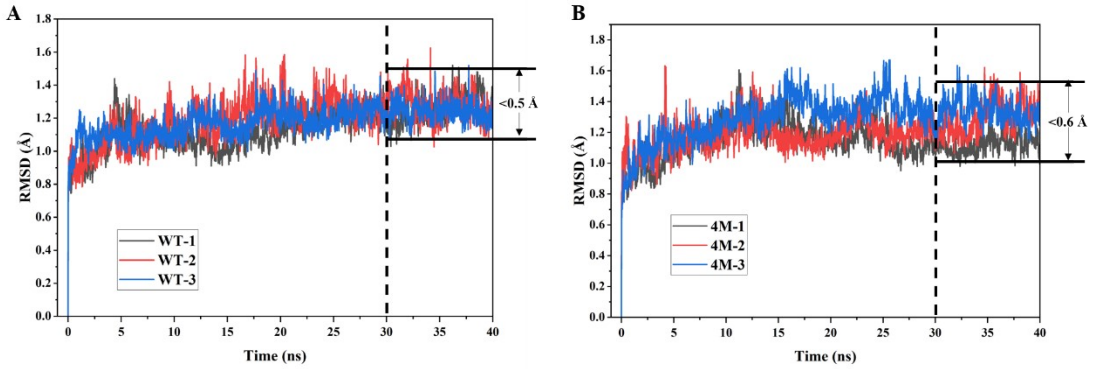


Figure S8. RMSD of WT and 4M Variants. All the simulated dynamic frames in the last 10 ns were stable, and the fluctuations were almost within 0.6 Å.

Reference

1. Yu, Z. et al. Site-specifically Incorporate Non-Canonical Amino Acids into Pseudomonas alcaligenes Lipase to Hydrolyze *L*-menthol propionate among the Eight Isomers. *Chemcatchem*, (2021).