Supplementary information

Engineering Thermostability of Lysine Hydroxylase for Scalable Production of (2*S*)-Hydroxy-1,5-pentanediamine from L-Lysine

Alei Zhang[†], Zhijie Zheng[†], Yangyang Li, Feifei Chen, Chaoqiang Wu, Kequan Chen[*]

College of Biotechnology and Pharmaceutical Engineering, State Key Laboratory of Materials-Oriented Chemical Engineering, Nanjing Tech University, Nanjing 211816, China

[*]Corresponding author: Kequan Chen

E-mail: kqchen@njtech.edu.cn.

[†]These authors contribute equally to this study

Table of Contents

1. Experimental section

1.1 Chemicals

Tryptone and yeast extract were purchased from Oxoid Co., Ltd. (Beijing, China). Fmoc chloride, SYPRO Orange dye and pyridoxal-5'-phosphate monohydrate (PLP) were purchased from Aladdin reagent Co., Ltd (Shanghai, China). All molecular reagents were purchased from Vazyme Biotech Co., Ltd (Nanjing, China). Other chemicals and reagents used in this study were purchased from local suppliers and analytical grade.

1.2 Strains, plasmids and primers

The PCR primers, strains, and plasmids used in this study were obtained from General Bio Co., Ltd. (Chuzhou, China) and summarized in Table S1 and Table S2.

1.3 Point mutation prediction and mutagenesis of K3H

Potential thermostable mutants were predicted using the online version of FireProt (https://loschmidt.chemi.muni.cz/fireprotweb/) and FoldX (https://foldxsuite.crg.eu/products). Multiple-sequence alignment was performed using ClustalW software (http://www.genome.jp/tools-bin/clustalw).

Single and combinatorial point mutants were constructed with the wild hydroxylase K3H gene from *Kineococcus radiotolerans* (Accession No. ABS05421) as a template using a QuickChange site-directed mutagenesis kit (Vazyme Biotech Co., Ltd). The mutagenic primers were shown in Table S1.

1.4 Construction of the plasmids and strains

The amplified PCR products of K3H and its mutant genes were purified by gel

electrophoresis, ligated into pRSFDuet1 vector with a C-terminal His×6-tag, and then transformed into competent *E. coli* Top 10. After sequenced by General Bio Co., Ltd. (Chuzhou, China), the positive recombinant plasmids were extracted and transformed into competent *E. coli* BL21(DE3) for protein expression.

E. coli BL21(DE3) harboring K3H together with lysine transporter argT and lysine decarboxylase CadA gene (Accession No. AM946981.) were respectively constructed according to our previous study[1].

1.5 Characterization of L-lysine 3-hydroxylase and its mutants

To investigate the temperature profile of K3H, purified protein and 10 mM L-lysine·hydrochloride was conducted under standard condition at 20, 25, 30, 35, 40, 45, 50, and 55°C for 30 min, respectively.

The kinetic parameters K_m and k_{cat} were determined using continuous kinetic assays. The protein concentration was 10 mg/L and the substrate concentration of L-lysine·hydrochloride varied from 0 to 25 mM. Data were subjected to nonlinear regression analysis according to the Michaelis-Menten equation using Origin 8.0 software.

1.6 Apparent melting temperature (T_m) and half-life ($t_{1/2}$) determination

A fluorescence-based thermal stability assay was used to determine apparent melting temperatures. The reaction solution (200 μL) contained 0.2–0.5 g/L purified enzyme, 20 μL 50-fold diluted SYPRO Orange dye and 50 mM PBS buffer (50 mM, pH 6.0) in a 96-well plate[3]. The plate was sealed with optical quality sealing tape and heated in a CFX 96 real-time polymerase chain reaction (PCR) system (BioRad, Hercules, CA,

USA) from 25 to 95°C at a heating rate of 1.25°C/min. All experiments were performed in triplicate.

The thermal inactivation half-life ($t_{1/2}$), measured the ability of an enzyme to retain 50% of its original activity. It was measured by calculating the residual enzyme activity at different temperatures for varied times.

1.7 Modelling, molecular docking and alanine scanning

To identify key sites for engineering from a structural perspective, AlphaFold 2.0 was utilized to construct a 3D model of K3H. The model demonstrated the presence of a double-stranded β-helix (DSBH) or jelly-roll core fold, and two lid regions, suggesting the possibility of divergent evolution from a common ancestor with other homologs. K3H contained eight antiparallel β folds and several α helices lying on the surface of the hydroxylase structure. A large number of random coils and loops were presented in the scaffold of its basic structure. The existence of a loop structure may play an important role in the catalysis of the enzyme. For instance, it has been reported that the loop region (lid region) of Fe/αKGD took part in the recognition of substrates, and the unbound substrate opened or the bound substrate closed the loop. However, sometimes the flexibility of a loop structure may lead to structural instability at high temperature, acidity, or alkalinity. Molecular docking was performed by AutoDock Vina. Alanine-scanning was carried out for amino acids within 5Å of the L-lysine within the complex. Molecular structures were visualized with PyMOL and VMD software. The modules of K3H and its mutants were repaired by the FoldX software.

1.8 Molecular dynamics (MD) simulation

The wild-type and mutant proteins were simulated in GROMACS 2018[4] using the CHARMM36 force field[5]. The proteins were placed in a cubic box distance 1.2 nm from the box edge and solvated using the TIP3P water model. 13 positive Na^+ ions were added to keep the systems neutral. All systems were minimized over 8000 steps of steepest descent minimization followed by 8,000 steps of conjugate gradient minimization. The Berendsen weak coupling method equilibrated the simulation for 300 ps at 320 K. The LINCS method was used to constrain the lengths of all bonds and long-range electrostatic interactions were calculated using the Particle Mesh Ewald method. Finally, 100 ns MD simulations were performed at 320 K.

1.9 Comparison of K3H and mutant D243A/S304V cells toward L-lysine hydroxylation

The recombinant K3H and mutant D243A/S304V cells were cultured via high-density fermentation in 5 L bioreactor according to our previous study[2]. The cells were harvested by centrifugation at 4°C and 5,000 rpm for 10 min. The 3 L reaction volume in 5 L bioreactor contained 100 mM phosphate buffer (pH 6.0), 600 mM L-lysine·hydrochloride (87.6 g/L L-lysine), 700 mM α-ketoglutaric acid (102.2 g/L), 0.1 mM $FeSO_4$, and *E. coli* whole cells (OD_{600}=15) was shaken with at 200 rpm and 30°C or 40°C, respectively.

1.10 Analytical methods

Recombinant protein samples were analyzed by 10% reductive sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with 20 mM β-mercaptoethanol incubation. A premixed protein marker (Takara Biotechnology Co., Ltd., Nanjing,

China) containing 180-, 140-, 100-, 75-, 60-, and 45-kDa protein bands was used as the molecular mass standard (Supplementary Figure S1, S2). Protein concentrations were determined by absorption at 595 nm using the Bradford method with bovine serum albumin as the standard.

The concentrations of L-lysine, 3-OH lysine, and 2-OH PDA were determined by Agilent 1260 series liquid chromatography system (Agilent Technologies, Santa Clara, USA) according to our previous study[2]. The gas chromatography of 2-OH PDA was performed with an Agilent 7890B system (Agilent Technologies, Santa Clara, USA) equipped with an HP-5 column (30 m×0.25 mm×0.5 μm, Agilent).

The viscosity and boiling point were analyzed by viscometer SNB-1 (Sunny Hengping Instrument, Shanghai, China) and petroleum products distillation tester NJ-SCLC-100 (Nuojin High-Speed Analyzer factory, Nanjing, China).

The molecular mass was analyzed using electrospray ionization mass spectrometry (ESI-MS, API 2000) with the ESI positive mode. The quadrupole scan mode was used under a capillary voltage of 2.8 kV, cone voltage of 30 V, desolvation gas temperature of 350°C and source temperature of 120°C and liquid chromatography quadrupole time-of-fight mass spectrometry (LC-Q-TOF-MS) analysis.

^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectra of the 2-OH-PDA were acquired in Chloroform-D1 ($CDCl_3$) solution on a Varian Mercury 400 NMR spectrometer (Bruker, MA, USA). ^{1}H NMR ($CDCl_3$, 400 MHz) δ: 3.42 (m, 1H), 2.86−2.77 (m, 1H), 2.69 (dd, J = 12.7, 3.7 Hz, 1H), 2.64−2.55 (m, 1H), 2.52 (dd, J = 12.7, 7.9 Hz, 1H), 2.11 (br s, 5H), 1.67−1.51 (m, 2H), 1.50−1.30 (m, 2H). ^{13}C NMR

($CDCl_3$, 100 MHz) δ: 71.6, 47.8, 41.7, 32.7, 29.5.

2. Supplementary tables and figures

Table S1. Primers used in this work

Number	Primer	Sequence (5'to3')
1	V259L-F	GAACGCctgCTGGAAATCTATGTGGCACATCG
2	V259L-R	ATTTCCAGcagGCGTTCCAGCAGGGCCTGGGC
3	K250P-F	ATGGCATTACCccgCATGCCCAGGCCCTGCTG
4	K250P-R	ATGcggGGTAATGCCATGCATCAGATCCTGAT
5	T237D-F	TGATCCGgatATGCTGGTTGATCAGGATCTGAT
6	T237D-R	CCAGCATatcCGGATCATCCACATCACCACTC
7	S227P-F	TCCGTTCccgATTCTGAGTGGTGATGTGGATGATC
8	S227P-R	TCAGAATcggGAACGGACCGCGCACTTCACGG
11	D210P-F	CCTGGTTccgGAATCATTCGCAGACTTCCTGG
12	D210P-R	ATGATTCcggAACCAGGGTGGTCCACAGCGGA
13	V192L-F	ATctgGACCCTACCGATGTGATGGAACTGTTC
14	V192L-R	ATCGGTAGGGTCcagATGGGCCAGCAGATCCAGTG
15	H191L-F	ATCTGCTGGCCctgGTGGACCCTACCGATGTGATGG
16	H191L-R	CACcagGGCCAGCAGATCCAGTGCGCGGAAGG
17	L186R-F	TTCCGCGCAcgtGATCTGCTGGCCCATGTGGA
18	L186R-R	AGATCacgTGCGCGGAAGGCATAGGTGGCGGC
19	R184S-F	TATGCCTTCagcGCACTGGATCTGCTGGCCCA
20	R184S-R	AGTGCgctGAAGGCATAGGTGGCGGCATCGGC
21	D161P-F	AACAGTGCTTCAGCccgCTGCGTCCGGATTATGTTGTG
22	D161P-R	CAGcggGCTGAAGCACTGTTCGGTATGTGCTTCCA
23	C158A-F	GAACAGgcaTTCAGCGATCTGCGTCCGGATTA
24	C158A-R	TCGCTGAAtgcCTGTTCGGTATGTGCTTCCAGTT
25	S144N-F	CCAGCAGaatCAGGGTAGCCGTGTTGAACTGG
26	S144N-R	TACCCTGattCTGCTGGGTGGCTGCCAGACGG
27	A139E-F	AATGCCCGTCTGgaaGCCACCCAGCAGAGTCAGG
28	A139E-R	GCttcCAGACGGGCATTCGGCACCATATCCTG
29	A136P-F	ATATGGTGCCGAATccgCGTCTGGCAGCCACCCAG
30	A136P-R	ACGcggATTCGGCACCATATCCTGCAGCAGATGAC
31	N135V-F	AGGATATGGTGCCGgttGCCCGTCTGGCAGCCACC
32	N135V-R	GCaacCGGCACCATATCCTGCAGCAGATGACCGT

Number	Primer	Sequence (5'to3')
33	H127A-F	TgcaCTGCTGCAGGATATGGTGCCGAATGCCC
34	H127A-R	TATCCTGCAGCAGtgcACCGTGGCCTTCTGCTGC
35	G119A-F	CATGGTGgcaTATGCAGCAGAAGGCCACGGTC
36	G119A-R	CTGCATAtgcCACCATGTGACCCAGGGCATGA
37	Q107L-F	GCCAGctgGCTATTGTTAGTCATGCCCTGGGT
38	Q107L-R	AACAATAGCcagCTGGCGTGCCAGCAGGGTGC
39	G95S-F	TAATACCagcGGCATTGGTGGTCGCACCCTGC
40	G95S-R	CAATGCCgctGGTATTATCTGCCGGGGTCGGC
41	E82P-F	GTCTGccgGTTGGCGAACTGCCGCCGACCCCG
42	E82P-R	TTCGCCAACcggCAGACCGCGCAGAACCAGAA
43	E67R-F	ATTCcgtGAACGTGGCAGCGATAGTGGTGTTC
44	E67R-R	TGCCACGTTCacgGAATGCCCACAGTGCTTCGG
45	T60R-F	GGAAGTTcgtGAAGCACTGTGGGCATTCGAAG
46	T60R-R	GTGCTTCacgAACTTCCACACTCAGTTCGCGT
47	V59L-F	TGTGGAActgACCGAAGCACTGTGGGCATTCG
48	V59L-R	CTTCGGTcagTTCCACACTCAGTTCGCGTGCC
49	S56P-F	AACTGccgGTGGAAGTTACCGAAGCACTGTGG
50	S56P-R	AACTTCCACcggCAGTTCGCGTGCCAGTAAGC
51	V264Y-F	CTATtatGCACATCGCCATGCAGTGGTTCTGC
52	V264Y-R	GGCGATGTGCataATAGATTTCCAGCACGCGTTCC
53	V233L-F	TGGTGATctgGATGATCCGACCATGCTGGTTG
54	V233L-R	GATCATCcagATCACCACTCAGAATACTGAACGG
55	E211P-F	GGTTGATccgTCATTCGCAGACTTCCTGGATAC
56	E211P-R	CGAATGAcggATCAACCAGGGTGGTCCACAGC
57	G170A-F	TTGTGCTGgcaTGTCTGCGCGGTGATGCCGAT
58	G170A-R	CAGACAtgcCAGCACAACATAATCCGGACGCA
59	Q157M-F	ACCGAAatgTGCTTCAGCGATCTGCGTCCGGA
60	Q157M-R	CTGAAGCAcatTTCGGTATGTGCTTCCAGTTCA
61	A140F-F	TCTGGCAtttACCCAGCAGAGTCAGGGTAGCC
62	A140F-r	GCTGGGTaaaTGCCAGACGGGCATTCGGCACC
63	S111A-F	TATTGTTgcaCATGCCCTGGGTCACATGGTGG
64	S111A-R	GGGCATGtgcAACAATAGCCTGCTGGCGTGCC
65	T60L-F	GGAAGTTctgGAAGCACTGTGGGCATTCGAAG
66	T60L-R	GTGCTTCcagAACTTCCACACTCAGTTCGCGT

Number	Primer	Sequence (5'to3')
67	I303L-F	TACCGATCGCTTCctgAGCCGCGGCTTCGTTGTT
68	I303L-R	TcagGAAGCGATCGGTACCATCGAAGCGCGGT
69	V277L-F	ATctgCTGCTGCTGGATAATCTGCGCGCAATG
70	V277L-R	ATCCAGCAGCAGcagATCGCCCGGCTGCAGAAC
71	L6P-F	TCTGTTCccgGATAGTAGTGCACATGTTCCGACC
72	L6P-R	TACTATCcggGAACAGACTGCTCATCGGATCC
73	T14P-F	ATGTTCCGccgCTGTTCGAACTGCCGGCCCCG
74	T14P-R	GAACAGcggCGGAACATGTGCACTACTATCCA
75	A20P-F	TGTTCGAACTGCCGccgCCGCAGCGTGCCGCATTA
76	A20P-R	cggCGGCAGTTCGAACAGGGTCGGAACATGTG
77	A25M-F	AGCGTGCCatgTTAGCTGCACTGGGTGCCCGC
78	A25M-R	AGCTAAcatGGCACGCTGCGGGGCCGGCAGTT
79	T39F-F	ATCCGGTGtttGAACCGGATGCCTTCGGCCGT
80	T39F-R	CGGTTCaaaCACCGGATCTGCGGTCAGGCGGG
81	V57P-F	ACTGAGTccgGAAGTTACCGAAGCACTGTGGG
82	V57P-R	TAACTTCcggACTCAGTTCGCGTGCCAGTAAG
83	A62M-F	TACCGAAatgCTGTGGGCATTCGAAGAACGTG
84	A62M-R	CCCACAGcatTTCGGTAACTTCCACACTCAGTTC
85	A65F-F	GCACTGTGGtttTTCGAAGAACGTGGCAGCGA
86	A65F-R	TCGAAaaaCCACAGTGCTTCGGTAACTTCCAC
87	V77I-F	TGTTCTGattCTGCGCGGTCTGGAAGTTGGCG
88	V77I-R	CGCGCAGaatCAGAACACCACTATCGCTGCCA
89	Q106M-F	CACGCatgCAGGCTATTGTTAGTCATGCCCTG
90	Q106M-R	AATAGCCTGcatGCGTGCCAGCAGGGTGCGAC
91	T206I-F	CTGTGGattACCCTGGTTGATGAATCATTCGC
92	T206I-R	ACCAGGGTaatCCACAGCGGACGGAACAGTTC
93	T206M-F	TGTGGatgACCCTGGTTGATGAATCATTCGCA
94	T206M-R	AACCAGGGTcatCCACAGCGGACGGAACAGTT
95	T206F-F	CTGTGGtttACCCTGGTTGATGAATCATTCGC
96	T206F-R	ACCAGGGTaaaCCACAGCGGACGGAACAGTTC
97	T206W-F	tggACCCTGGTTGATGAATCATTCGCAGACTT
98	T206W-R	TCATCAACCAGGGTccaCCACAGCGGACGGAACAG
99	T206V-F	TGTGGgttACCCTGGTTGATGAATCATTCGCA
100	T206V-R	AACCAGGGTaacCCACAGCGGACGGAACAGTT

Number	Primer	Sequence (5'to3')
101	T206P-F	ccgACCCTGGTTGATGAATCATTCGCAGACTT
102	T206P-R	TCATCAACCAGGGTcggCCACAGCGGACGGAACAG
103	S304I-F	CTTCATTattCGCGGCTTCGTTGTTCGTGATC
104	S304I-R	AGCCGCGaatAATGAAGCGATCGGTACCATCG
105	S304V-F	CTTCATTgttCGCGGCTTCGTTGTTCGTGATC
106	S304V-R	AGCCGCGaacAATGAAGCGATCGGTACCATCG
107	G306A-F	TTAGCCGCgcaTTCGTTGTTCGTGATCTGCGT
108	G306A-R	AACGAAtgcGCGGCTAATGAAGCGATCGGTAC
109	E211S-F	GGTTGATagcTCATTCGCAGACTTCCTGGATAC
110	E211S-R	CGAATGAgctATCAACCAGGGTGGTCCACAGC
111	E211T-F	GGTTGATaccTCATTCGCAGACTTCCTGGATAC
112	E211T-r	CGAATGAggtATCAACCAGGGTGGTCCACAGC
113	T141Q-F	CAGCCcagCAGCAGAGTCAGGGTAGCCGTGTT
114	T141Q-R	ACTCTGCTGctgGGCTGCCAGACGGGCATTCG
115	T141F-F	CAGCCtttCAGCAGAGTCAGGGTAGCCGTGTT
116	T141F-R	ACTCTGCTGaaaGGCTGCCAGACGGGCATTCG
117	D243A-F	TGATCAGgcaCTGATGCATGGCATTACCAAAC
118	D243A-R	GCATCAGtgcCTGATCAACCAGCATGGTCGGA
119	D243H-F	TGATCAGcatCTGATGCATGGCATTACCAAAC
120	D243H-r	GCATCAGatgCTGATCAACCAGCATGGTCGGA
121	D243L-F	TGATCAGctgCTGATGCATGGCATTACCAAAC
122	D243L-R	GCATCAGcagCTGATCAACCAGCATGGTCGGA
123	E123A-F	TATGCAGCAgcaGGCCACGGTCATCTGCTGCA
124	E123A-R	TGGCCtgcTGCTGCATAGCCCACCATGTGACC
125	M132A-F	AGGATgcaGTGCCGAATGCCCGTCTGGCAGCC
126	M132A-R	ATTCGGCACtgcATCCTGCAGCAGATGACCGT
127	Q142A-F	CgcaCAGAGTCAGGGTAGCCGTGTTGAACTGG
128	Q142A-R	TACCCTGACTCTGtgcGGTGGCTGCCAGACGGGC
129	Q143A-F	CCAGgcaAGTCAGGGTAGCCGTGTTGAACTGG
130	Q143A-R	TACCCTGACTtgcCTGGGTGGCTGCCAGACGG
131	S144A-F	CCAGCAGgcaCAGGGTAGCCGTGTTGAACTGG
132	S144A-R	TACCCTGtgcCTGCTGGGTGGCTGCCAGACGG
133	Q145A-F	AGCAGAGTgcaGGTAGCCGTGTTGAACTGGAA
134	Q145A-R	GCTACCtgcACTCTGCTGGGTGGCTGCCAGAC

Number	Primer	Sequence (5'to3')
135	L151A-F	TGTTGAAgcaGAAGCACATACCGAACAGTGCTT
136	L151A-R	GTGCTTCtgcTTCAACACGGCTACCCTGACTC
137	H154A-F	TGGAAGCAgcaACCGAACAGTGCTTCAGCGAT
138	H154A-R	TTCGGTtgcTGCTTCCAGTTCAACACGGCTAC
139	E156A-F	ACATACCgcaCAGTGCTTCAGCGATCTGCGTC
140	E156A-R	AGCACTGtgcGGTATGTGCTTCCAGTTCAACACG
141	Q157A-F	CACATACCGAAgcaTGCTTCAGCGATCTGCGTCCGG
142	Q157A-R	TGAAGCAtgcTTCGGTATGTGCTTCCAGTTCAACA
143	F159A-F	AACAGTGCgcaAGCGATCTGCGTCCGGATTAT
144	F159A-R	ATCGCTtgcGCACTGTTCGGTATGTGCTTCCA
145	L167A-F	GGATTATgcaGTGCTGGGTTGTCTGCGCGGTG
146	L167A-R	CCAGCACtgcATAATCCGGACGCAGATCGCTG
147	L169A-F	TGTTGTGgcaGGTTGTCTGCGCGGTGATGCCG
148	L169A-R	GACAACCtgcCACAACATAATCCGGACGCAGA
149	T180A-F	TGCCGCCgcaTATGCCTTCCGCGCACTGGATC
150	T180A-r	AGGCATAtgcGGCGGCATCGGCATCACCGCGC
151	D210A-F	CCTGGTTgcaGAATCATTCGCAGACTTCCTGG
152	D210A-R	ATGATTCtgcAACCAGGGTGGTCCACAGCGGA
153	E211A-F	GGTTGATgcaTCATTCGCAGACTTCCTGGATAC
154	E211A-R	CGAATGAtgcATCAACCAGGGTGGTCCACAGC
155	S212A-F	TGATGAAgcaTTCGCAGACTTCCTGGATACCC
156	S212A-R	CTGCGAAtgcTTCATCAACCAGGGTGGTCCAC
157	F213A-F	TGAATCAgcaGCAGACTTCCTGGATACCCGTG
158	F213A-R	AGTCTGCtgcTGATTCATCAACCAGGGTGGTC
159	D241A-F	TGCTGGTTgcaCAGGATCTGATGCATGGCATT
160	D241A-R	ATCCTGtgcAACCAGCATGGTCGGATCATCCA
161	D243A-F	TGATCAGgcaCTGATGCATGGCATTACCAAAC
162	D243A-R	GCATCAGtgcCTGATCAACCAGCATGGTCGGA
163	L244A-F	GTTGATCAGGATgcaATGCATGGCATTACCAAACATG
164	L244A-R	CATGCATtgcATCCTGATCAACCAGCATGGTCGGAT
165	L272A-F	ATGCAGTGGTTgcaCAGCCGGGCGATGTTCTG
166	L272A-R	CTGtgcAACCACTGCATGGCGATGTGCCACAT
167	L280A-F	TCTGCTGgcaGATAATCTGCGCGCAATGCATG
168	L280A-R	GATTATCtgcCAGCAGAACATCGCCCGGCTGC

Number	Primer	Sequence (5'to3')
169	H287A-F	ATCTGCGCGCAATGgcaGGCCGTAGCCCGTTCGCA
170	H287A-R	tgcCATTGCGCGCAGATTATCCAGCAGCAGAA
171	R289A-F	AATGCATGGCgcaAGCCCGTTCGCACCGCGCT
172	R289A-R	GGCTtgcGCCATGCATTGCGCGCAGATTATCC
173	F292A-F	ATGGCCGTAGCCCGgcaGCACCGCGCTTCGATGGTA
174	F292A-R	tgcCGGGCTACGGCCATGCATTGCGCGCAGATTAT
175	R301A-F	GGTACCGATgcaTTCATTAGCCGCGGCTTCGT
176	R301A-R	ATGAAtgcATCGGTACCATCGAAGCGCGGTGC
177	I303A-F	TACCGATCGCTTCgcaAGCCGCGGCTTCGTTGTT
178	I303A-R	TtgcGAAGCGATCGGTACCATCGAAGCGCGGT
179	R305A-F	TCATTAGCgcaGGCTTCGTTGTTCGTGATCTG
180	R305A-R	GAAGCCtgcGCTAATGAAGCGATCGGTACCAT
181	F307A-F	GCgcaGTTGTTCGTGATCTGCGTCGCAGTCGC
182	F307A-R	ATCACGAACAACtgcGCCGCGGCTAATGAAGCG

Notes: Lowercase letters in primers indicate the mutation sites.

Table S2. Strains and plasmids used in this work

Strain or plasmid	Description	Source
Strain		
E.coli Top10	F- ψ80dlacZ△M15 △(lacZYAargF)U169 DEOr recA1 endA1 hsdR17(rk-mk+)phoA supE44 λ- thi-1 gyrA96 relA1	General Bio Co., Ltd.
E.coli BL21(DE3)	F- ompT gal dcm lon hsdSb(Rb-Mb-) λ(DE3 [lacI lac lacUV5-T7 gene 1 ind1 sam7 nin5])	General Bio Co., Ltd.
Plasmid		
pRSFDuet1	*ColE1* ori, KanR, *E. coli* expression vector	This lab

Table S3. The optimum temperatures (T_{opt}) of various iron (II) αKG-dependent dioxygenases

Substrate	Proteins ID	Organism	Product	T_{opt} (℃)	Ref.
L-lysine	C7QJ42	*Catenulispora acidiphila* DSM44928	3-OH-lysine	30	[6]
	ABS05421	*Kineococcus radiotolerans* SRS30216		40	This study
	EAR24255	*Marine actinobacterium* PHSC20C1		15	[7]
	A5FF23	*Flavobacterium johnsoniae* DSM2064		NA	[6]
	ABQ06186	*Flavobacterium johnsoniae* UW101		40	[7]
L-lysine	ACU60313	*Chitinophaga pinensis* DSM2588	4-OH-lysine	35	[7]
	AEV99100	*Niastella koreensis* GR20-10		40	[7]
	EFK34737	*Chryseobacterium gleum* ATCC35910		25	[7]
	CAL80820	*Schlegelella brevitalea*		NA	[8]
L-proline	BAA22406	*Streptomyces* sp. TH1	cis-3-OH proline	35	[9]
	BAB52605	*Mesorhizobium japonicum* MAFF303099	cis-4-OH proline	25	[10]
	CAC47686	*Sinorhizobium meliloti* 1021		25	[10]
	AGS49661	uncultured bacterium esnapd13	trans-4-OH proline	37	[11]
L-isoleucine	AAP08048	*Bacillus thuringiensis* AKU0251	4-OH-isoleucine	30	[12]
L-leucine	ABA21237	*Anabaena variabillis* ATCC29413	5-OH-isoleucine	25	[13]

Substrate	Proteins ID	Organism	Product	T_{opt} (℃)	Ref.
L-pipecolate	CBL93717	*Streptomyces* sp. L-49973	cis-3-OH-pipecolate	30	[14]
	BAU50539	*Fusarium oxysporum* C8D	trans-4-OH-pipecolate	15	[15]
	XP_659994	*Aspergillus nidulans* FGSCA4		37	[15]
L-glutamine	I1R9B5	*Fusarium graminearum* PH-1	4-OH-glutamine	NA	[16]
L-citrulline	CBL93719	*Streptomyces* sp. L-49973	4-OH-citrulline	NA	[17]
L-arginine	AHW56367	*Streptomyces lavendulae* BCRC 12163	3-,4-OH-arginine	26	[18]
N-succinyl -L-leucine	YP_777923	*Burkholderia ambifaria* AMMD	N-succinyl 3-OH-leucine	30	[19]

Notes: The substrates are amino acids and their derivatives, which are highly selective substrates for hydroxylase. The above dioxygenases different in selectivity and activity also catalyzed other amino acids and their analogues to produce hydroxyl amino acids.

Table S4. Green metrics for 300 L-scale production of 2-OH-PDA

Process Stage	Input Mass (kg)	Notes
1. Biocatalyst Production	172.95	Includes 155.95 kg fermentation water, 17.0 kg solids (LB + feed)
2. Biotransformation	376.90	300 kg reaction water + 76.90 kg chemicals (substrates, cofactors, cells, PBS salts)
3. Downstream Processing	53.60	5.0 kg NaOH + 48.6 kg net n-butanol (The recovery rate is 90%)
Total Input	603.45	
Product (2-OH-PDA)	14.80	98.6% purity, 70.3% isolated yield
Green Metrics		
PMI (water-inclusive)	40.77	Per guidelines
MI (water-exclusive)	9.97	Chemicals only
E-factor	8.97	= MI − 1, per Sheldon

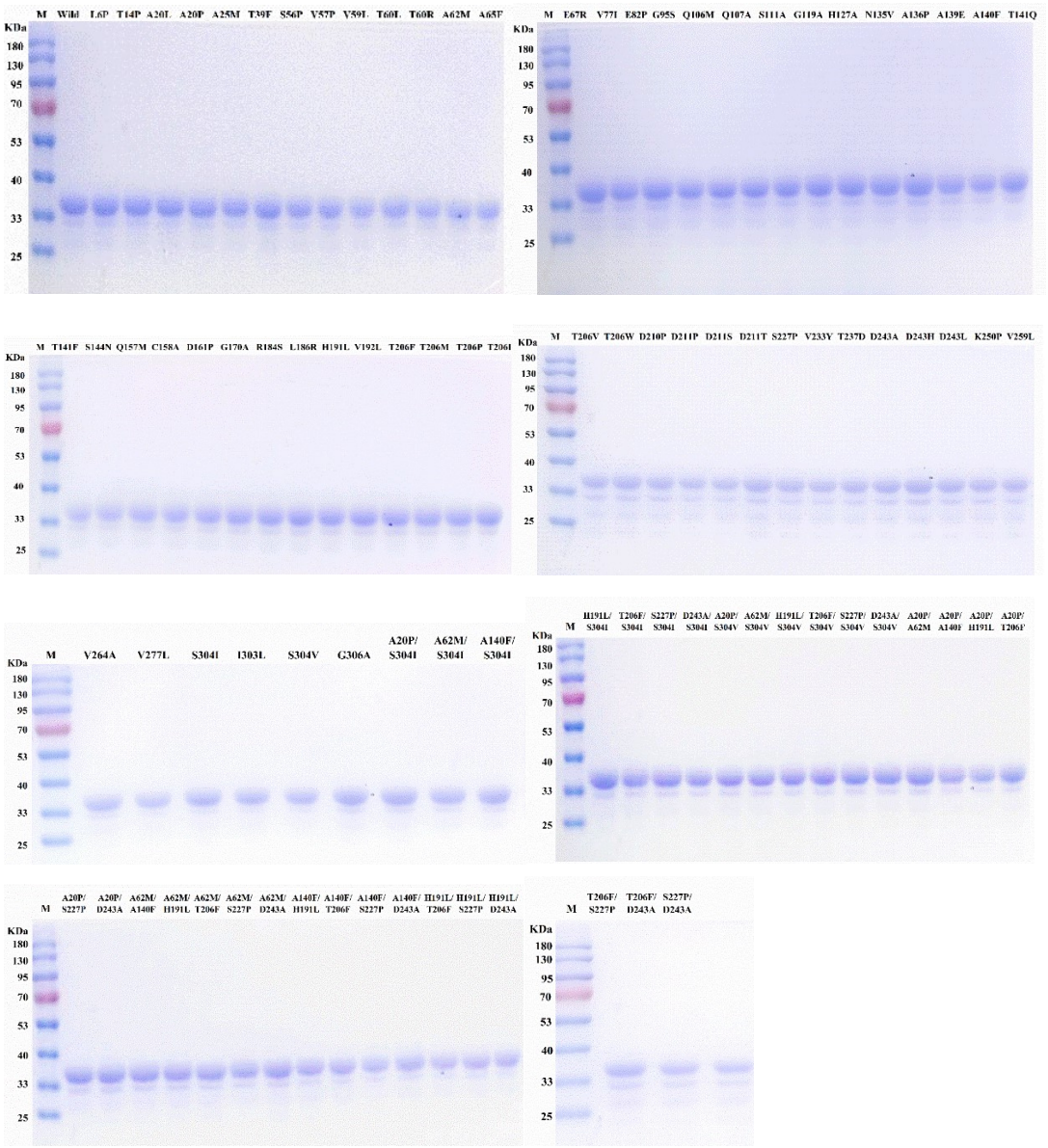
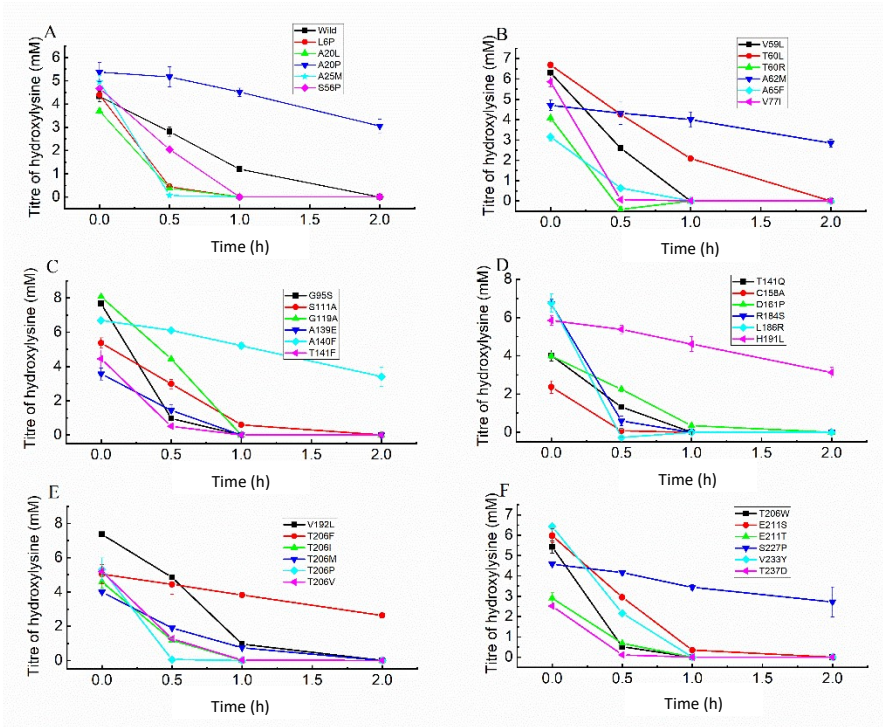
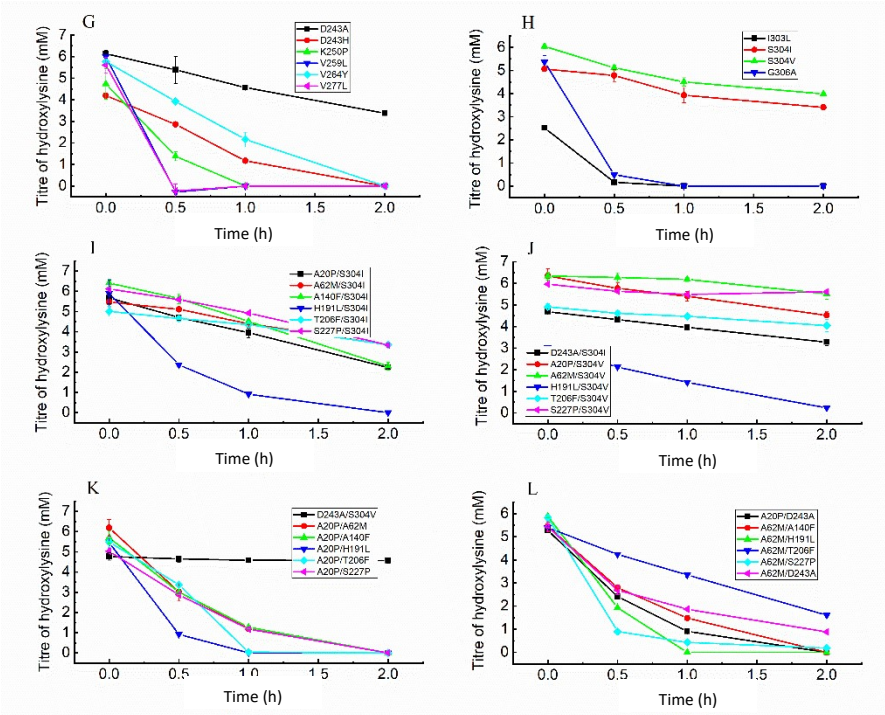


Figure S1. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of K3H and its mutants.

The soluble proteins (96) were initially purified on a Ni-NTA affinity column and analyzed by SDS-PAGE, except the insoluble protein (A140F/S304I).


A
Wild
L6P
A20L
A20P
A25M
S56P
Titre of hydroxylysine (mM)
Time (h)
B
V59L
T60L
T60R
A62M
A65F
V77I
C
G95S
S111A
G119A
A139E
A140F
T141F
D
T141Q
C158A
D161P
R184S
L186R
H191L
E
V192L
T206F
T206I
T206M
T206P
T206V
F
T206W
E211S
E211T
S227P
V233Y
T237D


G
D243A
D243H
K250P
V259L
V264Y
V277L
Titre of hydroxylysine (mM)
Time (h)
H
I303L
S304I
S304V
G306A
I
A20P/S304I
A62M/S304I
A140F/S304I
H191L/S304I
T206F/S304I
S227P/S304I
J
D243A/S304I
A20P/S304V
A62M/S304V
H191L/S304V
T206F/S304V
S227P/S304V
K
D243A/S304V
A20P/A62M
A20P/A140F
A20P/H191L
A20P/T206F
A20P/S227P
L
A20P/D243A
A62M/A140F
A62M/H191L
A62M/T206F
A62M/S227P
A62M/D243A

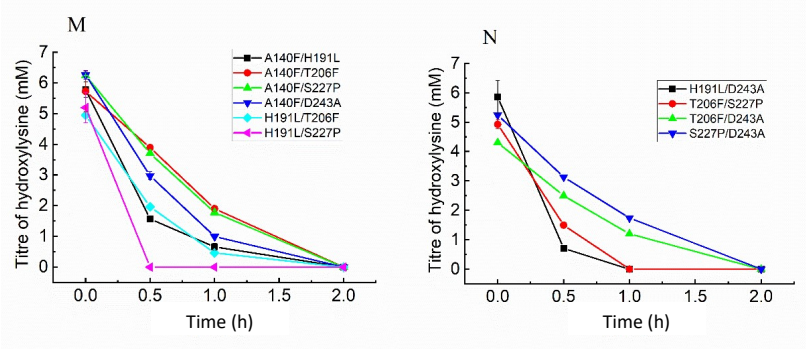


Figure S2. Thermostability screening of single-site mutants and combinational mutants.

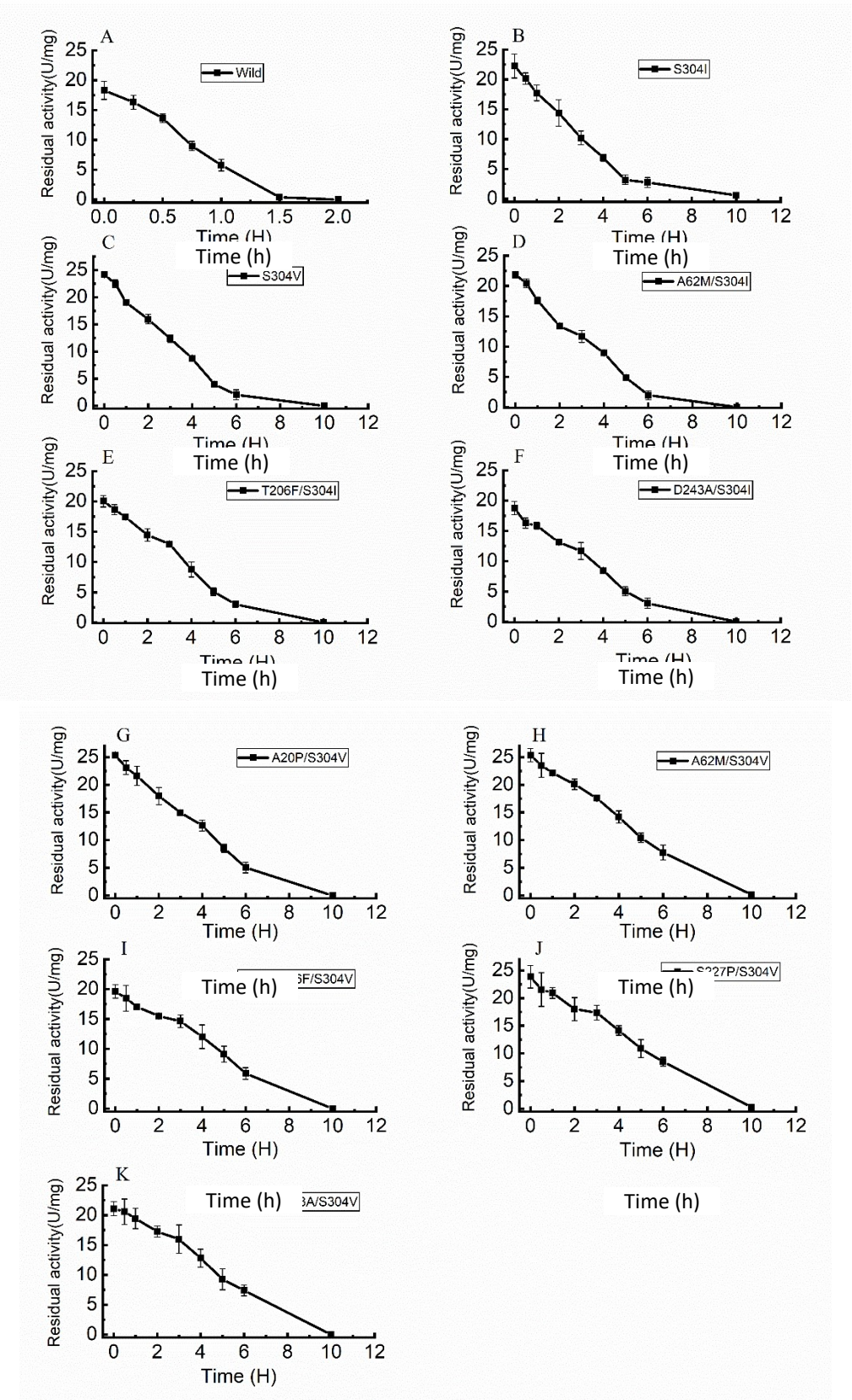


Figure S3. Analy Time (h) al inactivation half-lives ($t_{1/2}$) of the wild-type and mutants.

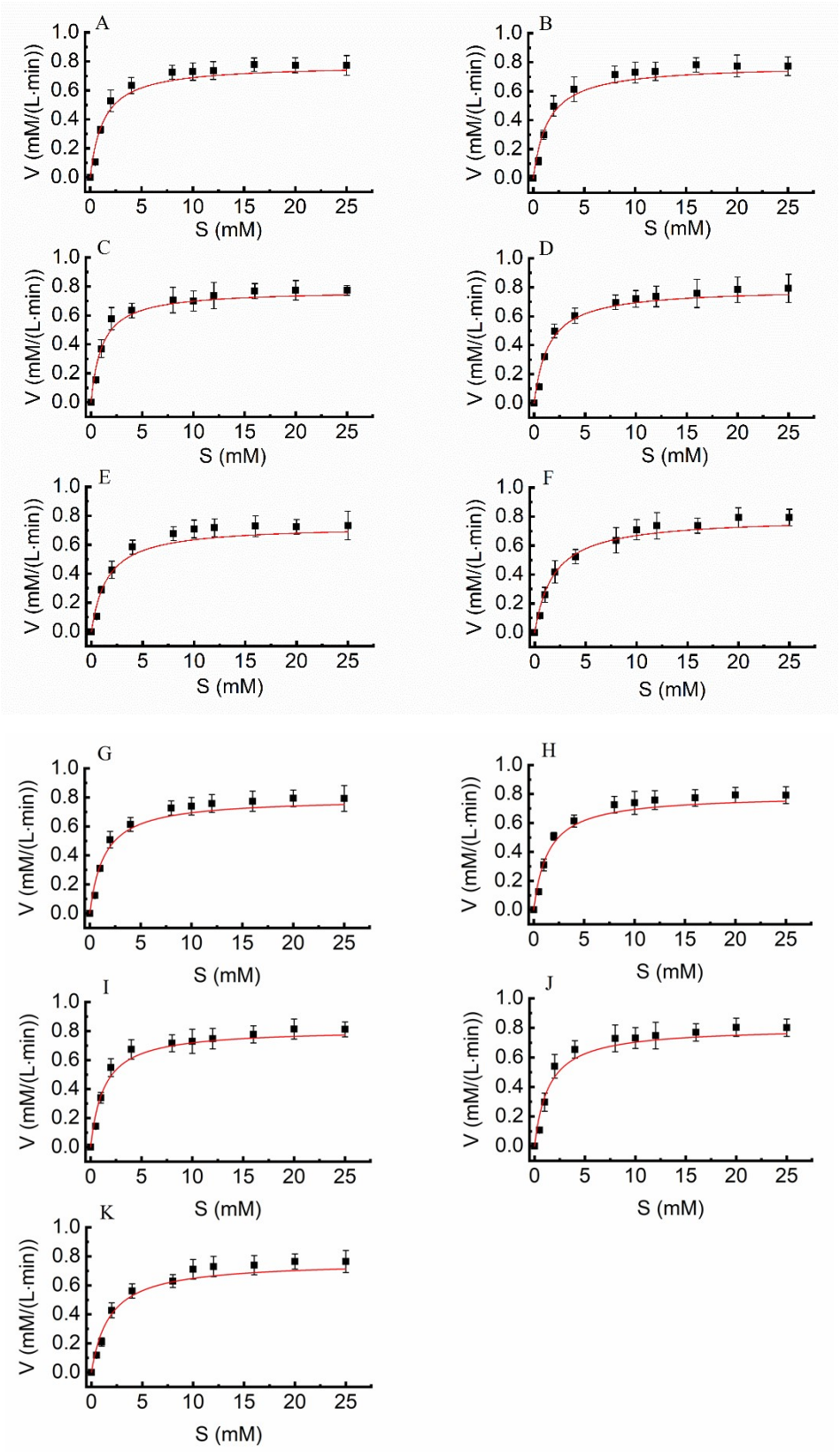


Figure S4. Kinetics parameters of the wild-type and mutants.

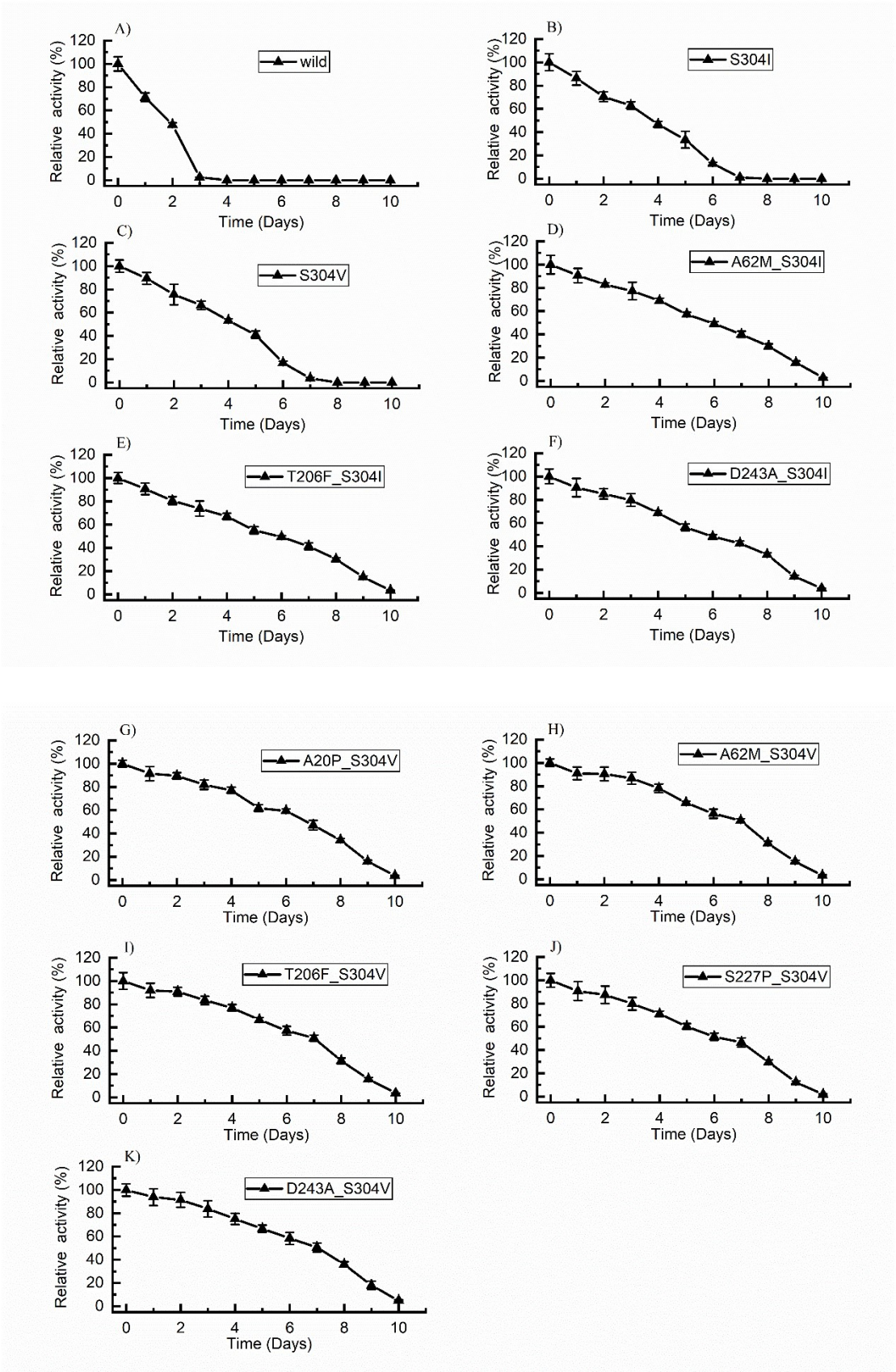


Figure S5. The relative activity of the wild-type and mutants at 30°C.

The wild-type and mutants were heated at 30°C water bath for 10 days. Then the residual enzyme activity was measured by enzyme activity assay and the initial relative activity was defined as

100%.

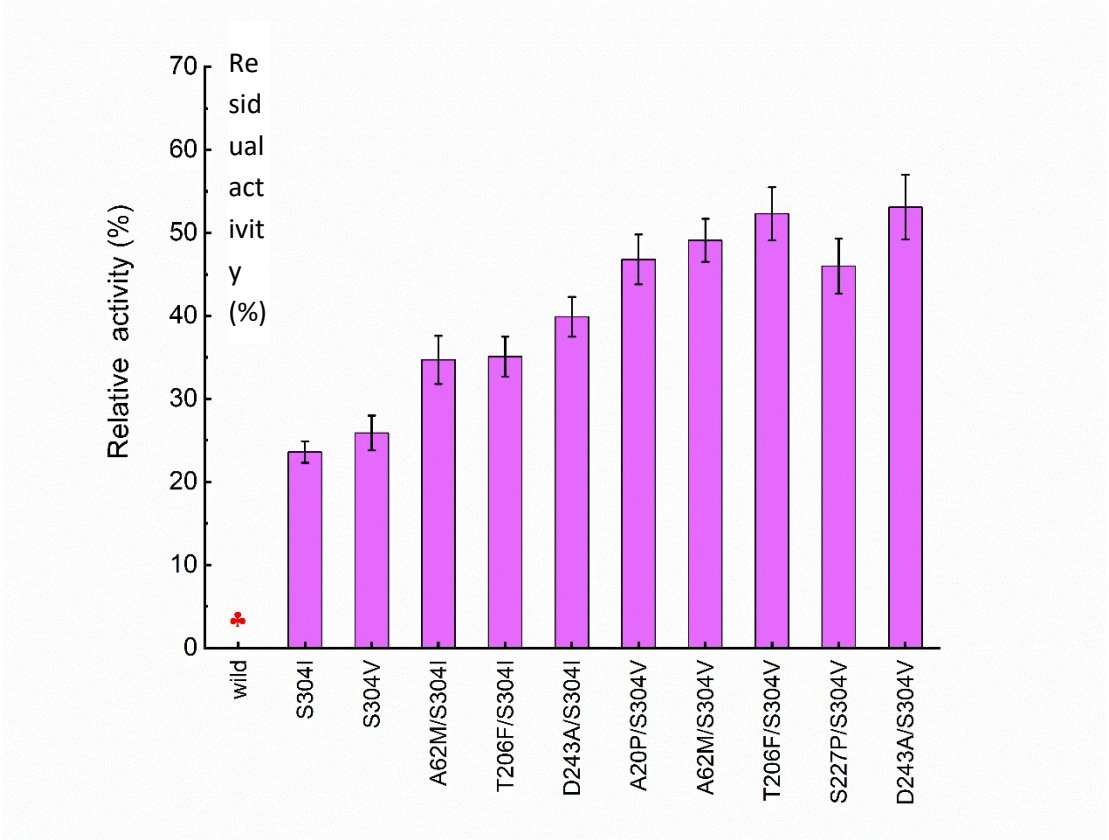


Figure S6. Comparison of the residual activity of the wild-type and mutants.
The wild and mutants were heated at 45°C for 30 min. Then the residual enzyme activity was measured with the initial relative activity normalized as 100%.

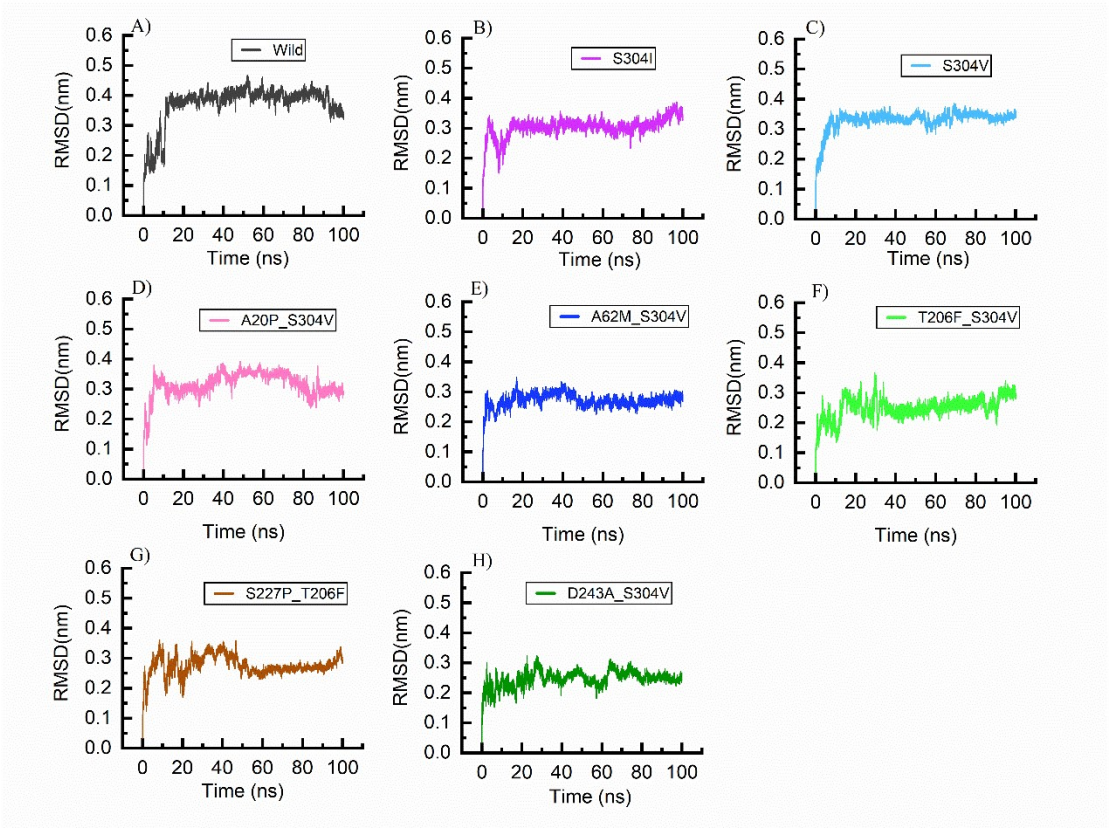


Figure S7. The average Cα root mean square deviation (RMSD) values.

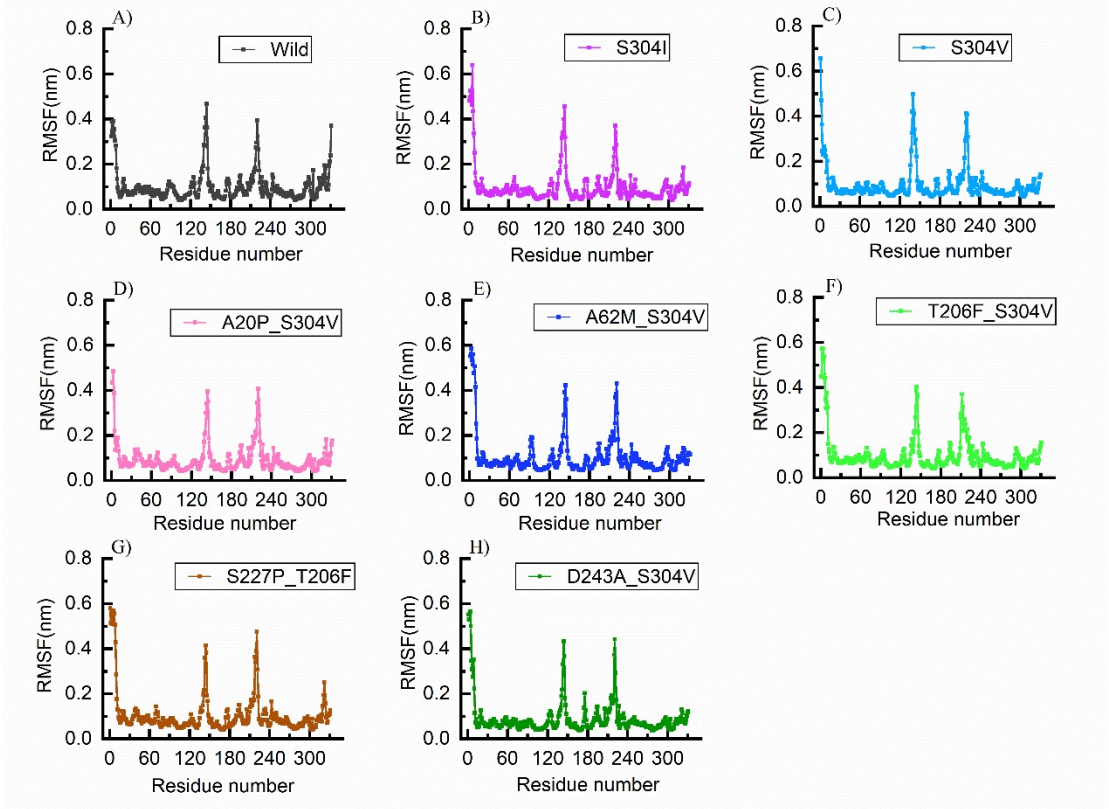


Figure S8. The average Cα root mean square fluctuation (RMSF) values.

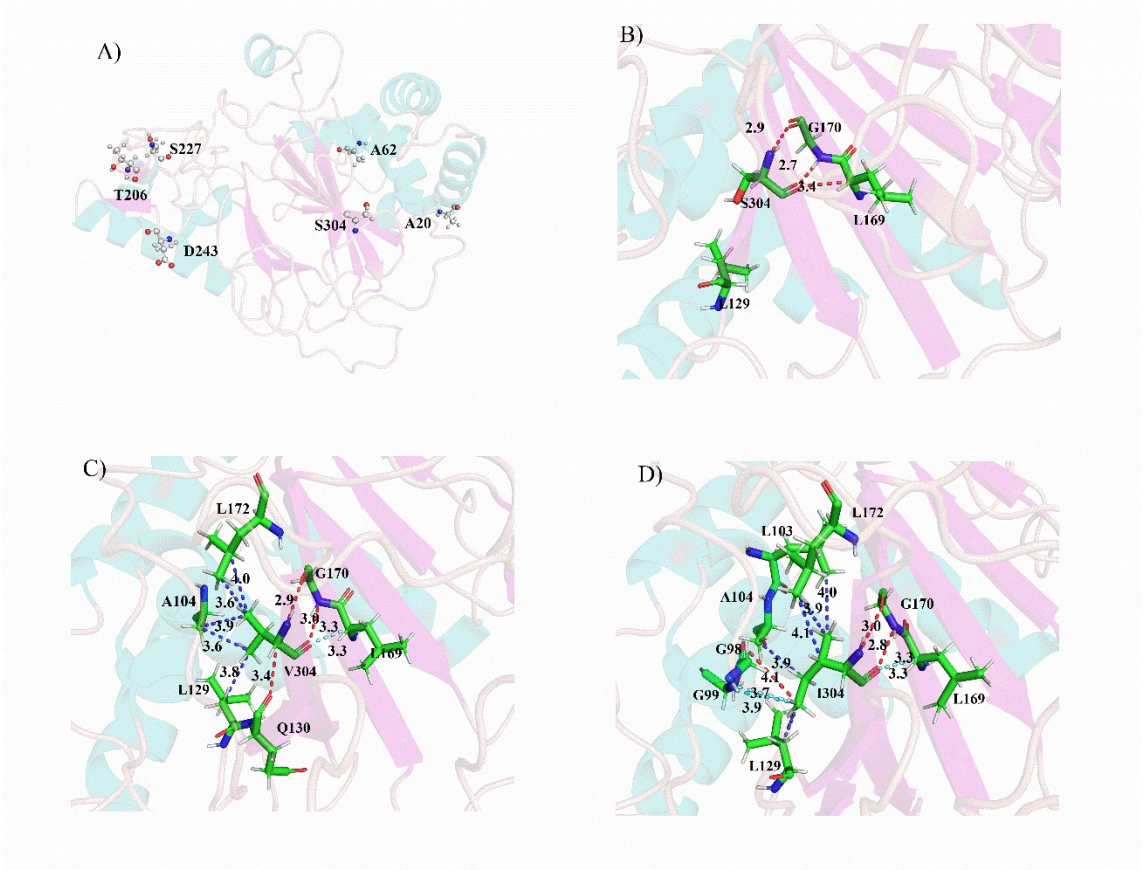


Figure S9. The mutation sites and interaction forces of S304V, S304I and wild-type. A) distribution of positive mutation sites, A20, A62, T206, D243, S227 and S304. B) interactions forces in wild-type. C) interactions forces in S304V. D) interactions forces in S304I. Polar force, red dotted lines. VDW force, cyans dotted lines. Hydrophobic force, tv_blue dotted lines. The numbers above dotted lines represented interaction forces distance (Å).

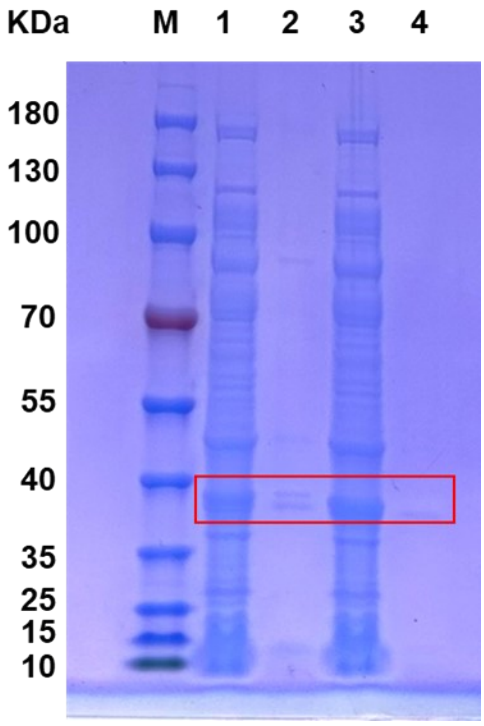


Figure S10. SDS-PAGE analysis of protein expression and solubility of wild-type K3H and D243A/S304V mutant in *E. coli* BL21(DE3). Lane M: Protein marker (kDa); Lane 1: Soluble fraction (supernatant) of cells expressing wild-type K3H; Lane 2: Insoluble fraction (pellet) of cells expressing wild-type K3H; Lane 3: Soluble fraction (supernatant) of cells expressing D243A/S304V mutant; Lane 4: Insoluble fraction (pellet) of cells expressing D243A/S304V mutant.

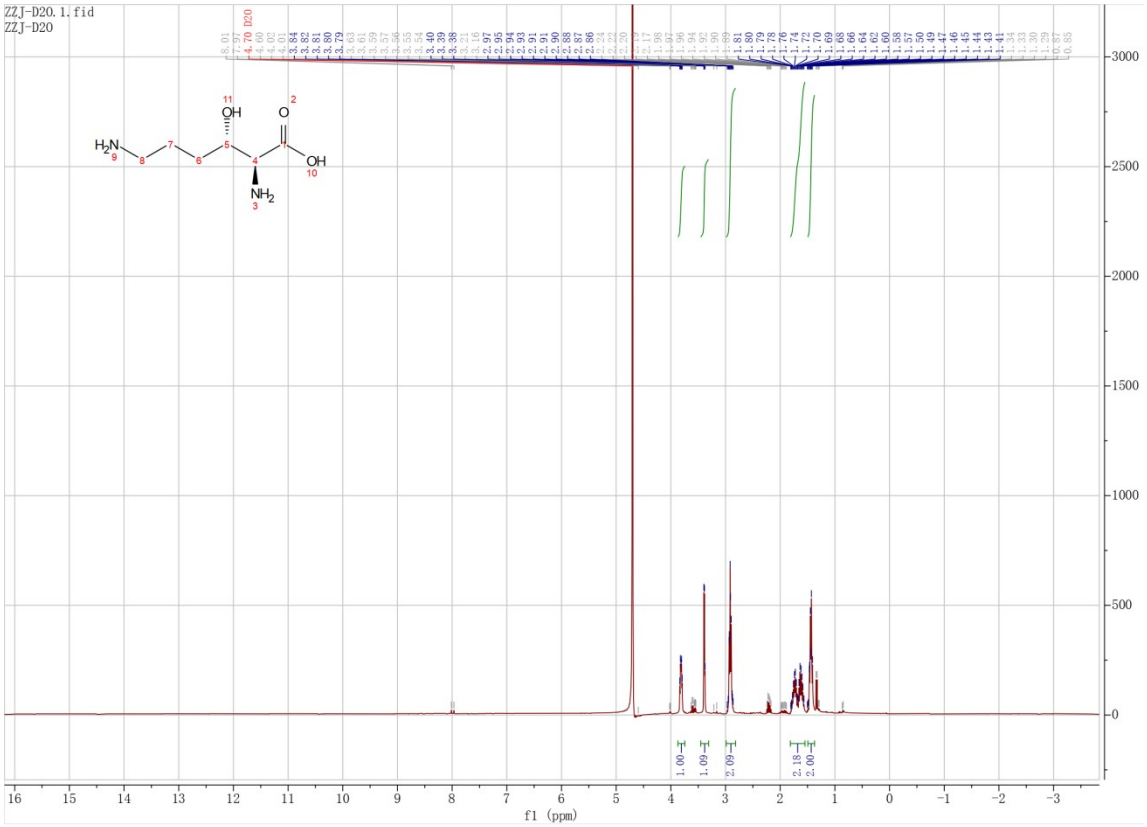


Figure S11. ^{1}H NMR spectrum of purified 3-OH-lysine intermediate.

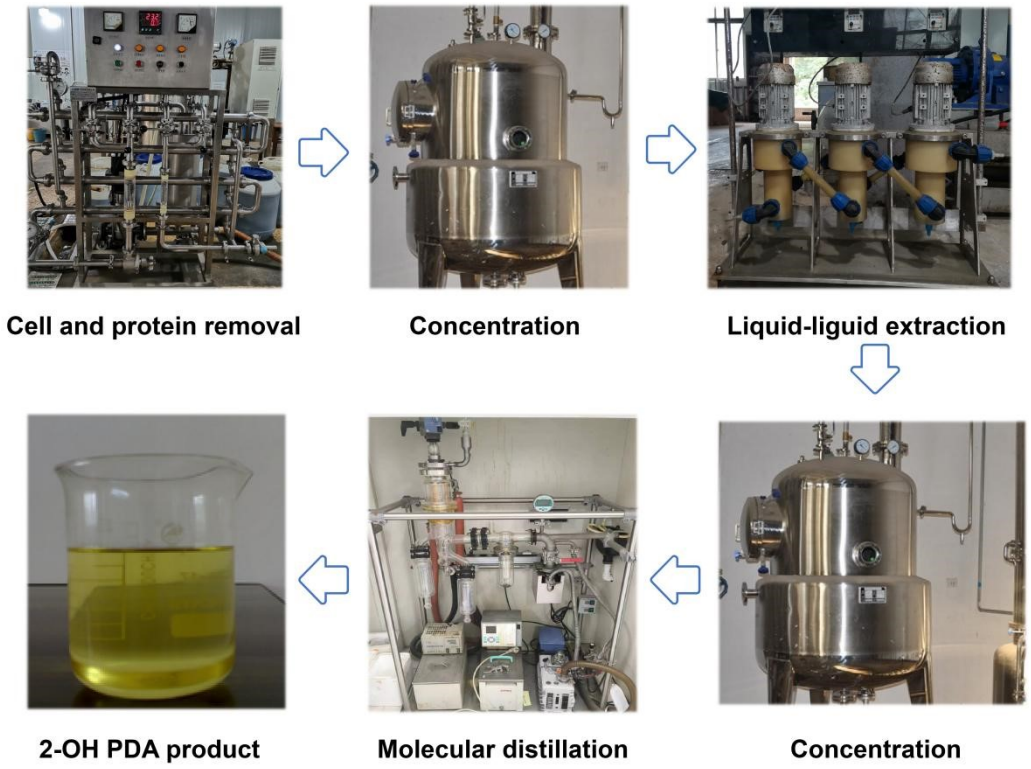


Figure S12. The workflow of purification of 2-OH PDA.

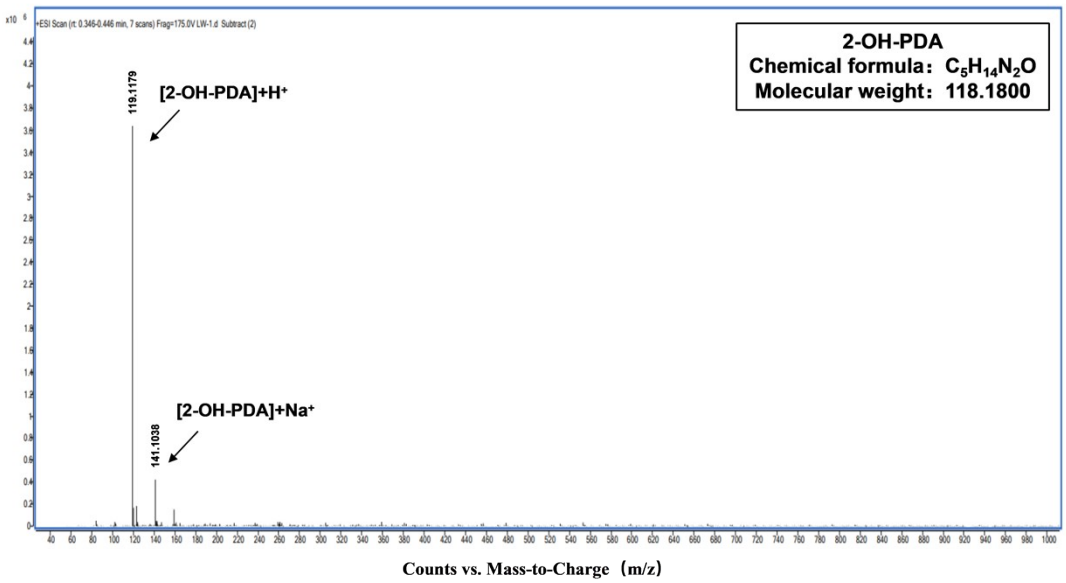


Figure S13. HRMS (ESI) spectrum of purified 2-OH PDA product.

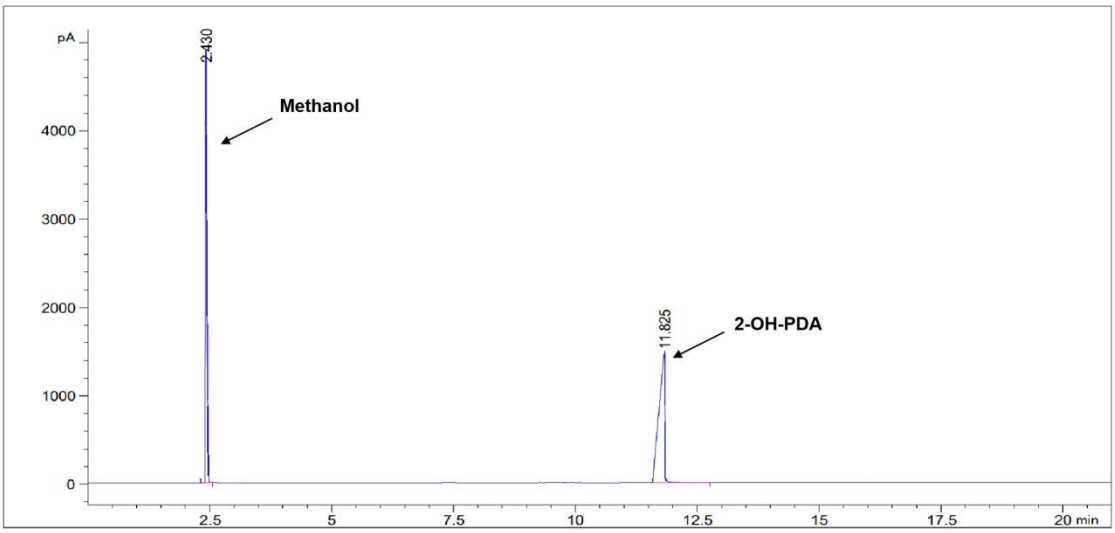


Figure S14. GC spectrum of purified 2-OH PDA product.

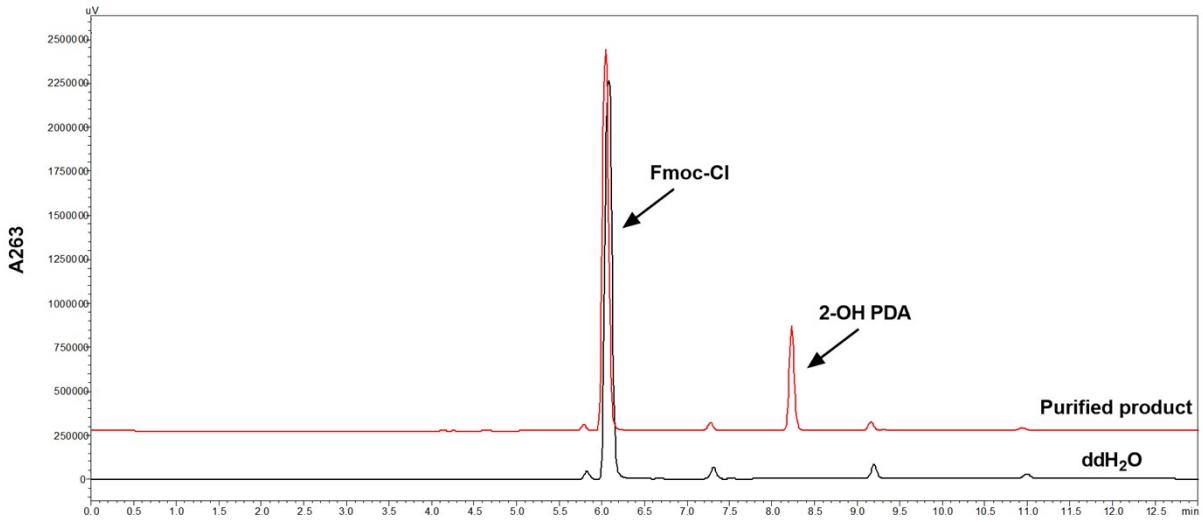
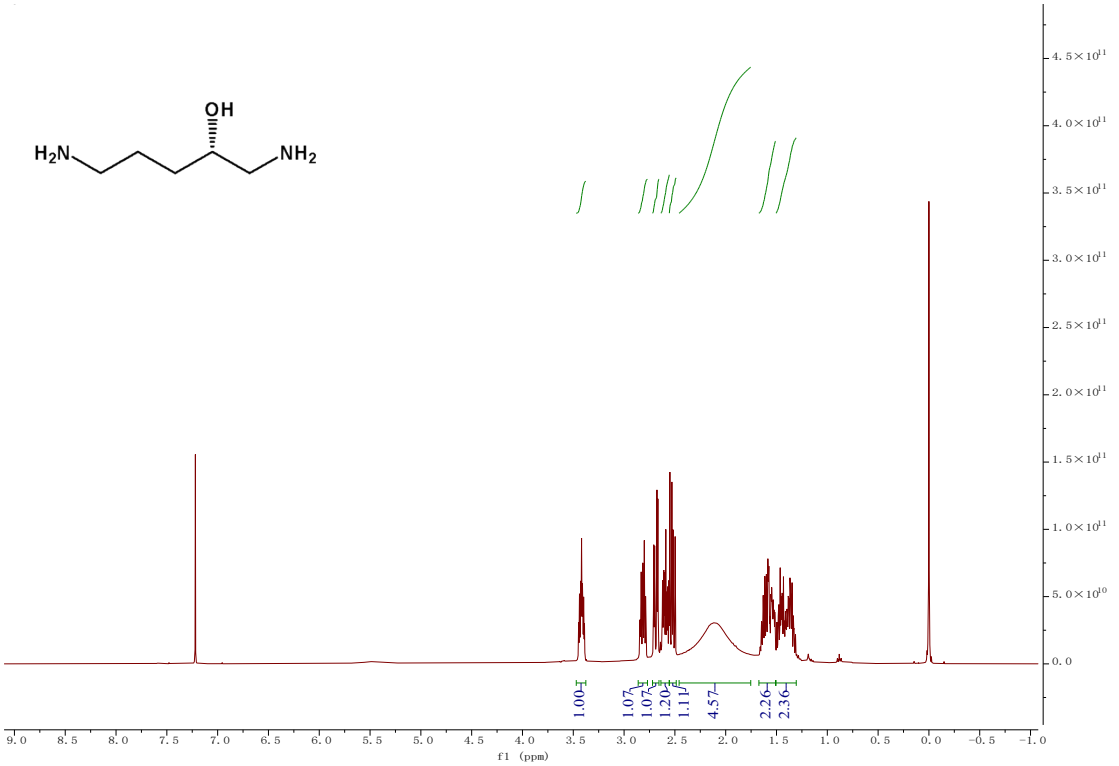


Figure S15. HPLC spectrum of purified 2-OH PDA product.



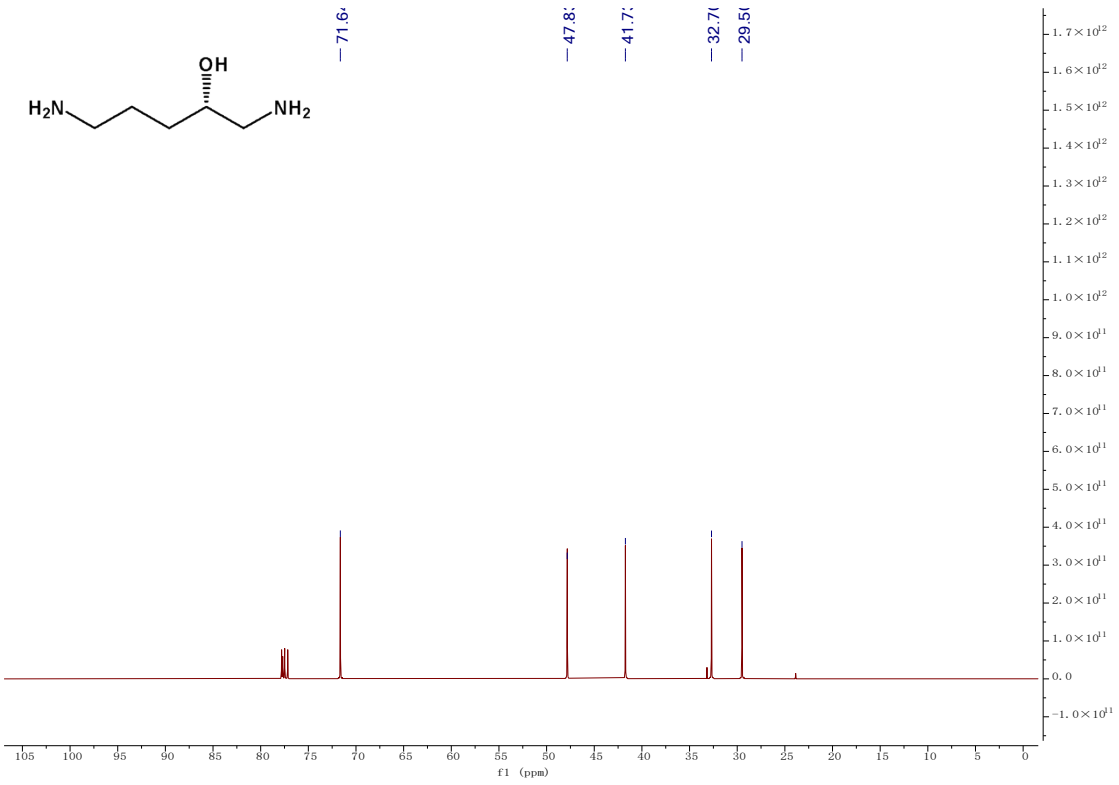


Figure S16. ^{1}H and ^{13}C NMR spectrum of purified 2-OH PDA product.

Reference

1. Y. Li, A. Zhang, S. Hu, K. Chen and P. Ouyang, *Microb Cell Fact*, 2022, 21, 142.
2. R. Xu, W. Jiang, Y. Liu, J. Hu, D. Liu, S. Zhou, Y. Zhong, F. Zhang and M. Zhao, *Biochem Biophys Res Commun*, 2022, 602, 156-162.
3. Y. L. Cui, Y. C. Chen, X. Y. Liu, S. J. Dong, Y. E. Tian, Y. X. Qiao, R. Mitra, J. Han, C. L. Li, X. Han, W. D. Liu, Q. Chen, W. Q. Wei, X. Wang, W. B. Du, S. Y. Tang, H. Xiang, H. Y. Liu, Y. Liang, K. N. Houk and B. Wu, *Acs Catal*, 2021, 11, 1340-1350.
4. J. A. Lemkul, *Living J. Comp. Mol. Sci*, 2018, 1, 5068.
5. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller and A. D. MacKerell, Jr., *Nat Methods*, 2017, 14, 71-73.
6. D. Baud, P.-L. Saaidi, A. Monfleur, M. Harari, J. Cuccaro, A. Fossey, M. Besnard, A. Debard, A. Mariage, V. Pellouin, J.-L. Petit, M. Salanoubat, J. Weissenbach, V. de Berardinis and A. Zaparucha, *ChemCatChem*, 2014, 6, 3012-3017.
7. K. Y. Ryotaro Hara, Ryoma Miyake, Hiroshi Kawabata, Hisatoshi Uehara, Kuniki Kino, *Applied and Environmental Microbiology*, 2017, 83, e00693-00617.
8. A. Amatuni and H. Renata, *Org Biomol Chem*, 2019, 17, 1736-1739.
9. H. Mori, T. Shibasaki, K. Yano and A. Ozaki, *J Bacteriol*, 1997, 179, 5677-5683.
10. R. Hara and K. Kino, *Biochem Biophys Res Commun*, 2009, 379, 882-886.
11. C. Liu, J. Zhao, J. Liu, X. Guo, D. Rao, H. Liu, P. Zheng, J. Sun and Y. Ma, *Appl Microbiol Biotechnol*, 2019, 103, 265-277.
12. Z. Qiao, M. Xu, M. Shao, Y. Zhao, M. Long, T. Yang, X. Zhang, S. Yang, H. Nakanishi and Z. Rao, *Eng Life Sci*, 2020, 20, 7-16.
13. R. S. Correia Cordeiro, J. Enoki, F. Busch, C. Mugge and R. Kourist, *J Biotechnol*, 2018, 284, 68-74.
14. S. Hu, P. Yang, Y. Li, A. Zhang, K. Chen and P. Ouyang, *Enzyme and microbial technology*, 2022, 154, 109958-109958.
15. M. Hibi, R. Mori, R. Miyake, H. Kawabata, S. Kozono, S. Takahashi and J. Ogawa, *Appl Environ Microbiol*, 2016, 82, 2070-2077.
16. H. Renata, E. Shimizu and C. R. Zwick, 3rd, *Tetrahedron*, 2021, 90.
17. C. R. Zwick, 3rd, M. B. Sosa and H. Renata, *Angew Chem Int Ed Engl*, 2019, 58, 18854-18858.
18. C. Y. Chang, S. Y. Lyu, Y. C. Liu, N. S. Hsu, C. C. Wu, C. F. Tang, K. H. Lin, J. Y. Ho, C. J. Wu, M. D. Tsai and T. L. Li, *Angew Chem Int Ed Engl*, 2014, 53, 1943-1948.
19. M. Hibi, T. Kawashima, T. Kasahara, P. M. Sokolov, S. V. Smirnov, T. Kodera, M. Sugiyama, S. Shimizu, K. Yokozeki and J. Ogawa, *Lett Appl Microbiol*, 2012, 55, 414-419.