Supplementary Information for:

Protein Dynamics-assisted Engineering of a Selective Debenzylase for Replacing Pd/C-catalyzed Debenzylation in Statin Precursor Synthesis

Jianfeng Lin,[a,b,c] Xuri Wu,[a,c] and Yijun Chen[*a,b,c]

[a] State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing, Jiangsu 211198, China

[b] Laboratory of Chemical Biology, School of Life Science and Technology, China Pharmaceutical University, Nanjing, Jiangsu 211198, China

[c] Department of Microbiology and Synthetic Biology, School of Life Science and Technology, China Pharmaceutical University, Nanjing, Jiangsu 211198, China

Correspondence: yjchen@cpu.edu.cn

Table of contents

Experimental procedures

HPLC and LC-MS analysis

Liquid chromatography–mass spectrometry (LC-MS) analysis was conducted with a Shimadzu LCMS-8045 spectrometer. LC-MS analysis was performed using a Phenomenex Gemini® C18 column (5 μm, 250 × 4.6 mm) with mobile phase A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) at 30 °C. The detection wavelength was set at 220 nm, with a flow rate of 1 mL/min. The gradient elution program for the preliminary screening of enzyme activity and selectivity was detailed in Table S5, and the gradient elution program for the preparation of the standard curve and product quantification was shown in Table S6 in order to separate the internal standard testosterone and the products. The MS parameters were set as Table S7. A multiple reaction monitoring (MRM)-mass spectrometry (MS) method was developed for product quantification. Precursor ion was initially screened in both positive and negative ionization modes using Q1 scan mode. Subsequently, product ion scan mode was selected for analysis of precursor ion in order to determine specific m/z values of produced fragments. This was followed by optimization of specific precursor ion to

fragment ion transitions selected for the MRM method. Then, optimization was performed for collision energy and input voltage of Q1 and Q3.

Measurement of protein concentration

The concentration of CYP102A1 and its mutants was determined by carbon monoxide difference spectrophotometry.[1] Purified CYP102A1 and mutant solutions were diluted tenfold, and 1 mL aliquots were transferred into two cuvettes. The cuvettes were placed in the spectrophotometer sample chamber, and the baseline spectrum was recorded from 400 to 500 nm. Carbon monoxide gas was then bubbled into the sample cuvette at a rate of one bubble per second for a total of 60 bubbles. Subsequently, 20 μL of freshly prepared sodium dithionite solution (50 mg/mL) was added to each cuvette, followed by gentle inversion mixing. The absorption spectrum from 400 to 500 nm was immediately recorded. The reaction was monitored continuously until the absorbance at 450 nm reached a plateau. The characteristic absorbance values at 450 nm and 490 nm were recorded, and the P450 concentration was calculated using the equation mentioned in the protocol.[1] Each sample was measured in triplicate.

Determination of kinetic parameters and selectivity

Kinetic parameters were determined using purified enzymes. Reactions were performed in 200 μL of 100 mM phosphate buffer (pH 7.0) at 30 °C and at 350 rpm for 30 min in the presence of CYP102A1, M1, M2, M3, M4 and M5 at concentration of 2.89, 2.42, 0.38, 0.40, 0.37, and 0.29 μM, respectively, 0.5 μM G6PDH, different concentrations of substrate **1b** varied from 0.05 mM to 25 mM, 2 mM $NADP^+$, and 10 mM G6P. When turnover number was measured, the reactions were performed in 200 μL of 100 mM phosphate buffer (pH 7.0) at 30°C and at 500 rpm in the presence of CYP102A1, M1, M2, M3, M4 and M5 at concentration of 0.1 μM, 0.5 μM G6PDH, 3 mM substrate **1b**, 2 mM $NADP^+$, and 10 mM G6P. Reactions for TTN measurement were conducted until no further product formation was observed. The reaction times were as follows: WT, 15 h; M1, 8 h; M2, 6 h; M3, 5 h; M4, 5 h; and M5, 5 h. An equal volume of acetonitrile was added to quench the reactions. The mixtures were

centrifuged, and 190 μL of the supernatant was collected and mixed with 10 μL of a 200 μM testosterone internal standard solution prior to LC-MS/MS analysis. Kinetic parameters k_{cat} and K_M were calculated based on the Michaelis-Menten equation. All experiments were performed in triplicate.

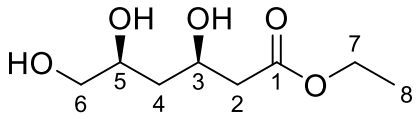
Preparation of ethyl 2-((4R,6S)-6-(hydroxymethyl)-2-phenyl-1,3-dioxan-4-yl)acetate by enzymatic cascade reaction and protection of the 1,3-diol

The crude enzyme solutions were prepared by resuspending the cell pellets in pH 7 phosphate buffer solution at a 1:10 ratio, followed by sonication and centrifugation. This preparation method was applied uniformly to all enzymes used in the cascade reaction. Enzymatic activity of the crude enzyme solutions was assayed at 30 °C and pH 7.0, with one unit (U) of enzyme activity defined as the amount of enzyme that converts 1 mM substrate per minute in a 1 mL reaction volume. The enzymatic cascade reaction was performed in a 5 mL system containing 100 mM phosphate buffer (pH 7.0), 1 mM substrate **1a**, and the following components: 0.7 U DKR, 2 U glucose dehydrogenase, 0.2 U mutant M5, 5 U G6PDH, 5 mM glucose, 8 mM G6P, 0.1 mM NAD^+, 0.2 mM $NADP^+$, and 0.2% (v/v) DMSO. The reaction was conducted at 30 °C with shaking at 80 rpm. To maintain reaction efficiency, enzyme and chemicals were added at specific time points. At 10 min, 0.066 U of M5, 1 U of G6PDH, 4 mM G6P, and 0.2 mM $NADP^+$ were added. Furthermore, 0.033 U of M5, 1 U of G6PDH, and 1 mM G6P was supplemented at 40 min, 70 min, 100 min, and 130 min after the initiation of the reactions. The total reaction time was 310 min, and samples were taken at 10 min, 40 min, 70 min, 100 min, 130 min, 190 min, 250 min, and 310 min for the analysis. At each time point, 50 μL of the reaction mixture was sampled and mixed with 250 μL of acetonitrile to quench the reaction. After centrifugation, 190 μL of the supernatant was collected and mixed with 10 μL of a 200 μM testosterone internal standard solution prior to LC-MS/MS analysis. The protection of the 1,3-diol moiety in compound **2a** was performed in a 4 mL reaction system containing 2 mM **2a**, 2.5 mol% p-toluenesulfonic acid, 2 equivalents of benzaldehyde, using dichloromethane as solvent. The reaction was carried out at 28 °C, 220 rpm for 3 h. The reaction mixture was diluted

20-fold with a 50% (v/v) aqueous methanol solution containing testosterone at a final concentration of 10 μM as internal standard. The product Ethyl 2-((4R,6S)-6-(hydroxymethyl)-2-phenyl-1,3-dioxan-4-yl)acetate **2c** was quantitated by LC-MS/MS with the pair of m/z 298 and m/z 87 (Fig. S7). All experiments were performed in triplicate.

Supplementary Tables

Table S1 NMR data of compound **2a**

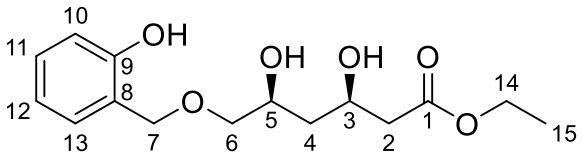


(2a)

Position	δ_C, type	δ_H, mult. (*J* in Hz)
1	171.32, C	-
2	42.34, CH_2	2.47-2.21, m
3	65.61, CH	4.08-4.00, overlapped
4	40.67, CH_2	1.57-1.40, m
5	69.25, CH	3.56, m
6	65.85, CH_2	3.29-3.23, m
7	59.57, CH_2	4.08-4.00, overlapped
8	14.12, CH_3	1.18, t (7.1)
3-OH	-	4.72, s
5-OH	-	4.53, s
6-OH	-	4.48, s

^{1}H NMR:600 MHz; ^{13}C NMR: 150 MHz (in DMSO-d_6)

Table S2 NMR data of compound **2b**

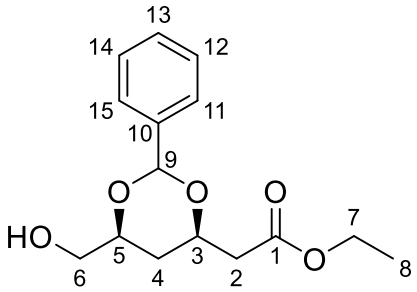


(2b)

Position	δ_C, type	δ_H, mult. (*J* in Hz)
1	171.26, C	-
2	42.31, CH_2	2.26-2.46, m
3	65.37, CH	4.04, overlapped
4	40.98, CH_2	1.60-1.46, m
5	67.03, CH	3.77, m
6	74.69, CH_2	4.56-4.42, overlapped
7	67.44, CH_2	4.72, s
8	124.71, C	-
9	154.82, C	-
10	114.89, CH	7.24, dd (7.5, 1.7)
11	128.14, CH	7.08, td (7.7, 1.8)
12	118.62, CH	6.76, t (7.4)
13	128.68, CH	7.35, dd (8.3, 5.9)
14	59.54, CH_2	4.04, q (7.1)
15	14.08, CH_3	1.17, t (7.1)
3-OH	-	6.79, d (8.0)
5-OH	-	4.45, d (3.2)
9-OH	-	9.44, s

^{1}H NMR:600 MHz; ^{13}C NMR: 150 MHz (in DMSO-d_6)

Table S3 NMR data of compound **2c**



(2c)

Position	δ_C, type	δ_H, mult. (*J* in Hz)
1	170.17, C	-
2	40.53, CH_2	2.63-2.51, m
3	72.70, CH	4.27, m
4	32.39, CH_2	1.70-1.32, m
5	76.96, CH	3.91, m
6	64.04, CH_2	3.45, m
7	59.89, CH_2	4.07, m
8	14.06, CH_3	1.17, t (7.2)
9	99.79, CH	5.57, s
10	138.63, C	-
11	126.18, CH	7.43-7.39, m
12	127.86, CH	7.38-7.32, overlapped
13	128.51, CH	7.38-7.32, overlapped
14	127.86, CH	7.38-7.32, overlapped
15	126.19, CH	7.40-7.41, m
6-OH	-	4.75, t (5.8)

^{1}H NMR:600 MHz; ^{13}C NMR: 150 MHz (in DMSO-d_6)

Table S4 Regions with different movement patterns and the bulky residues involved

Region	Residues within region	Bulky amino acids
1	204-231 and 246-276;	Q204, F205, Q206, R223, Q229, R255, Y256, Q257, F261, H266, F275;
2	134-163 and 180-188, 200-218;	H138, R147, F158, Y160, R161, F162, R203, (Q204, F205, Q206 are included above);
3	76-87 and 203-228;	R79, F77, F81, F87 (R203, Q204, F205, Q206, R223 are included above);
4	3-46 and 272-285;	Q7, F11, Q27, F40, F42, Y278, F279, H285 (F275 is included above);
5	3-46 and 423-436, 442-450;	F423, H426, Y429, F444 (Q7, F11, Q27, F40, F42 are included above);
6	180-203, 219-232 and 433-438, 443-448;	Q189, R190, Y198 (R203, F444 are included above);

Table S5 HPLC gradient elution program used in preliminary screening of enzyme activity and selectivity

Time (min)	Mobile phase A	Mobile phase B
0	91.0%	9.0%
9	91.0%	9.0%
19	0.0%	100.0%
25	0.0%	100.0%

Table S6 HPLC gradient elution program for preparing standard curves and quantitative analysis of products

Time (min)	Mobile phase A	Mobile phase B
0	88.0%	12.0%
6	88.0%	12.0%
6.01	80.0%	20.0%
26	55.0%	45.0%
32	48.0%	52.0%
32.01	0.0%	100.0%
36	0.0%	100.0%

Table S7 MS parameters used in this study

Parameter	Value
MS Min Range	100
MS Max Range	500
MS scan rate (u/sec)	2142
Drying gas Flow (L/min)	10
Nebulizing gas flow (L/min)	3
Heating gas flow (L/min)	10
Heat block temperature (℃)	400
Interface temperature (℃)	300
Desolvation line temperature (℃)	250
Ion source	ESI
Capillary voltage (+) (kV)	4
Capillary voltage (-) (kV)	-3

Supplementary Figures

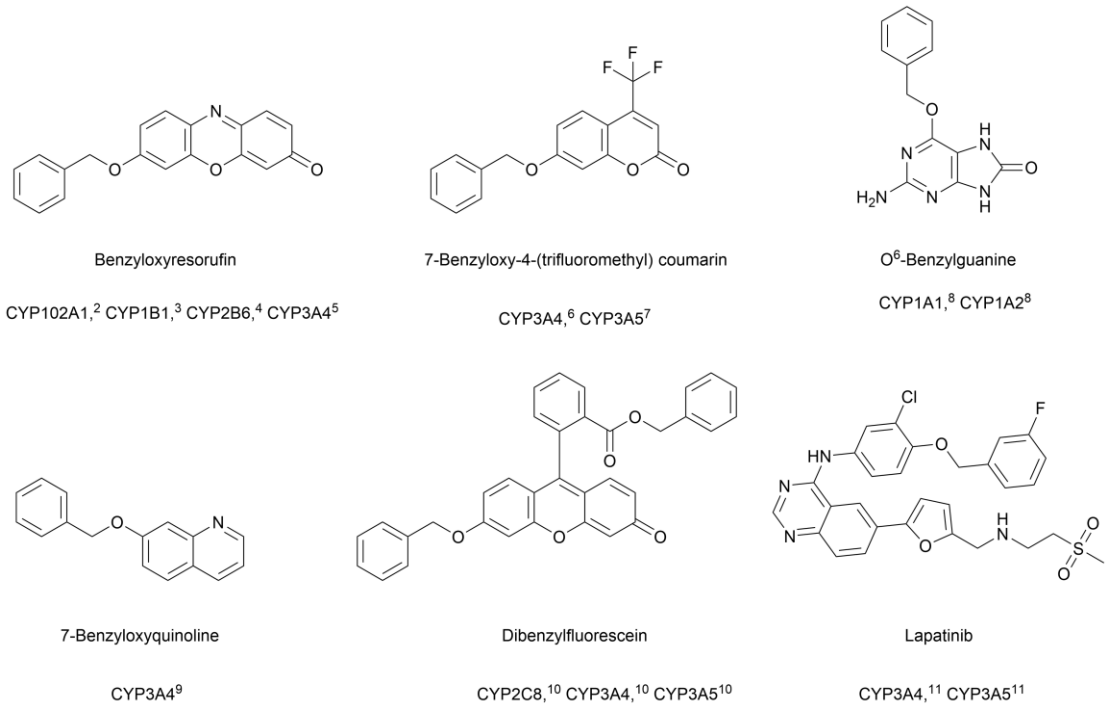


Fig. S1 Compounds that have been reported to be debenzylated by P450 and the specific P450s that catalyze the reaction.[2-11]

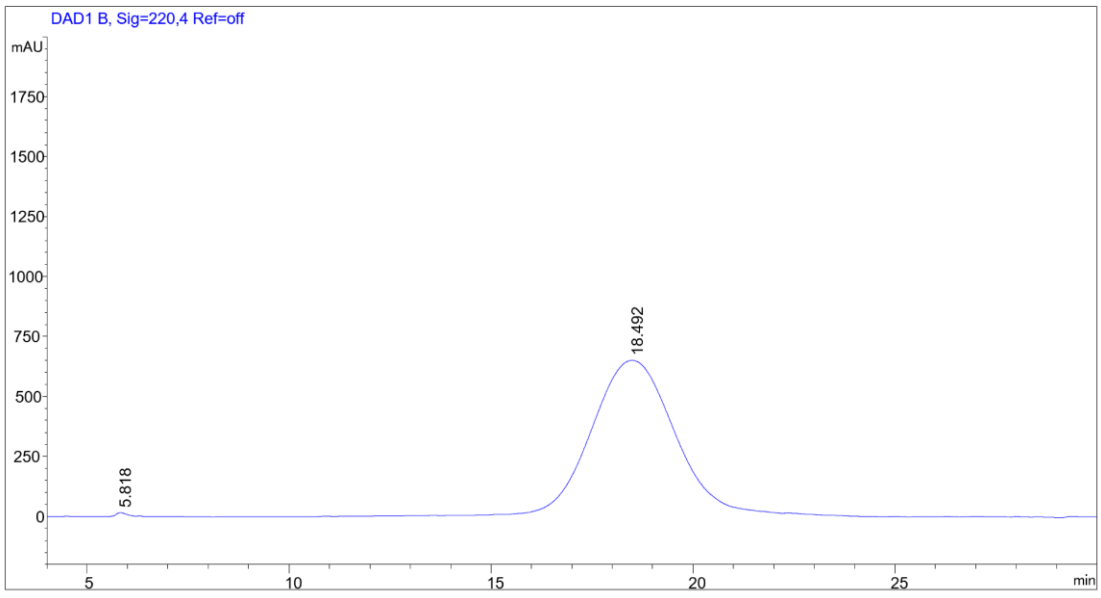


Fig. S2 HPLC analysis for **1a**. Separation was achieved on a Phenomenex Gemini® C18 column (5 μm, 250 × 4.6 mm). The mobile phase consisted of (A) 0.1% formic acid in water and (B) acetonitrile with the following gradient program: 0-30 min, 45%

B to 45% B. The flow rate was 1.0 mL/min and the detection wavelength was set at 220 nm.

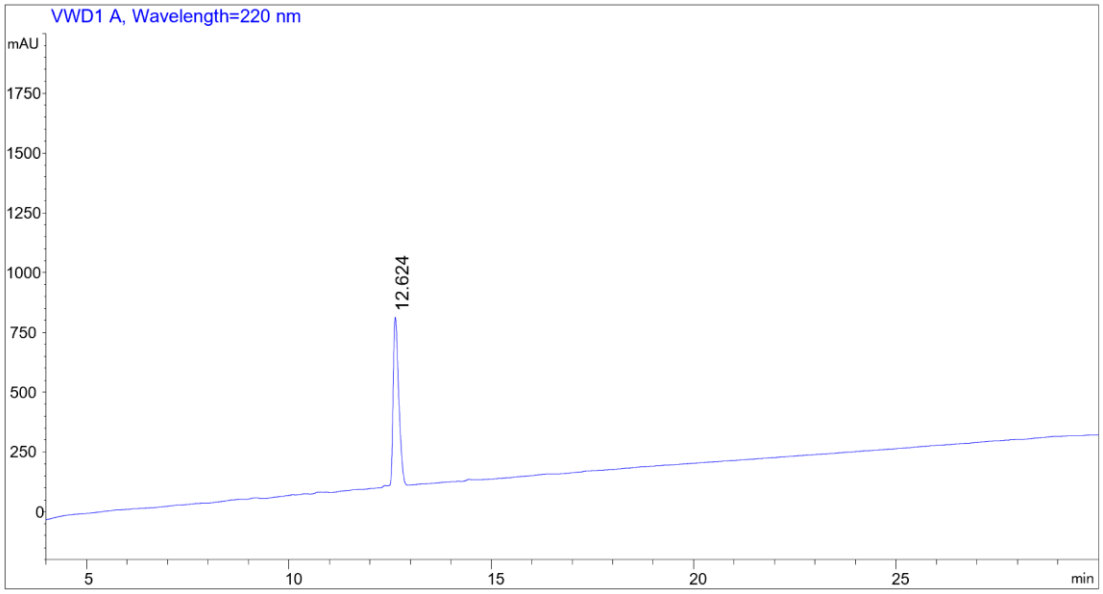


Fig. S3 HPLC analysis for **1b**. Separation was achieved on a Phenomenex Gemini® C18 column (5 μm, 250 × 4.6 mm). The mobile phase consisted of (A) 0.1% formic acid in water and (B) acetonitrile with the following gradient program: 0-30 min, 20% B to 100% B; The flow rate was 1.0 mL/min and the detection wavelength was set at 220 nm.

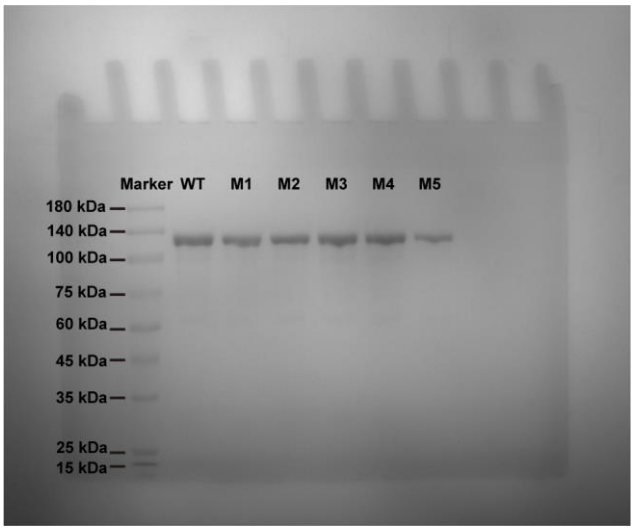


Fig. S4 SDS-PAGE of CYP102A1 and its mutants. Lane 1 is the marker and lane 2 to 7 are purified enzymes WT, M1, M2, M3, M4 and M5 respectively.

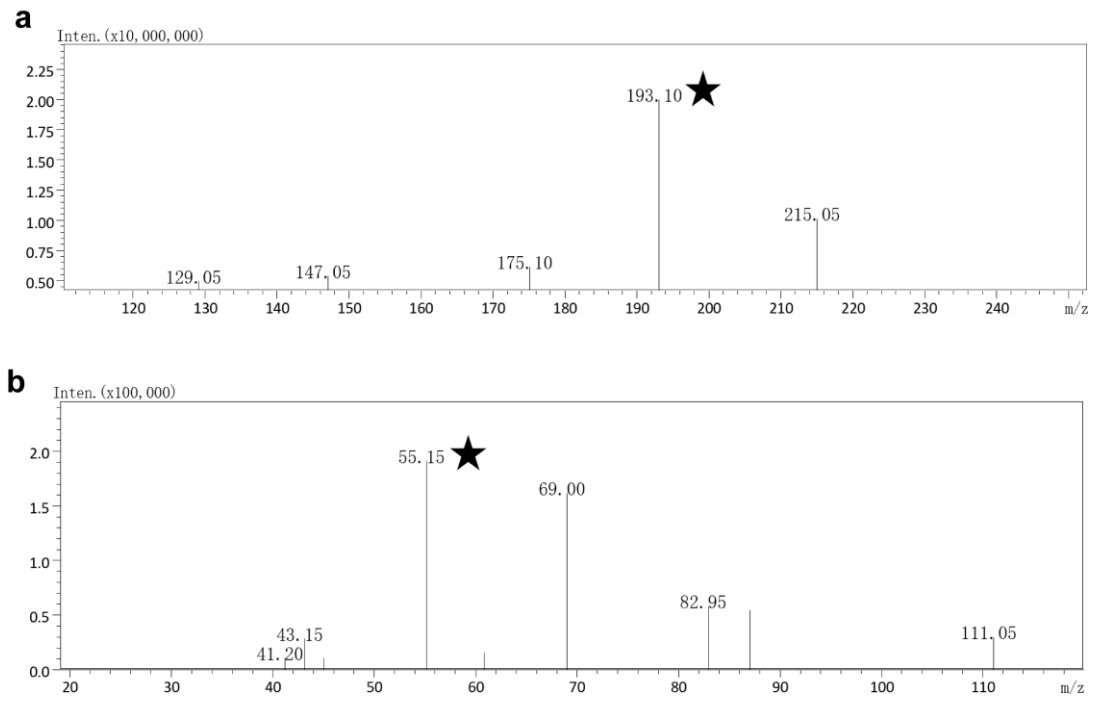


Fig. S5 LC-MS/MS analysis for **2a** in positive mode. a) Full-scan MS spectrum. The molecular ion at m/z 193 chosen for quantification is indicated by a star. b) MS/MS spectrum of m/z 193 ion. The daughter ion at m/z 55 chosen for quantification is indicated by a star.

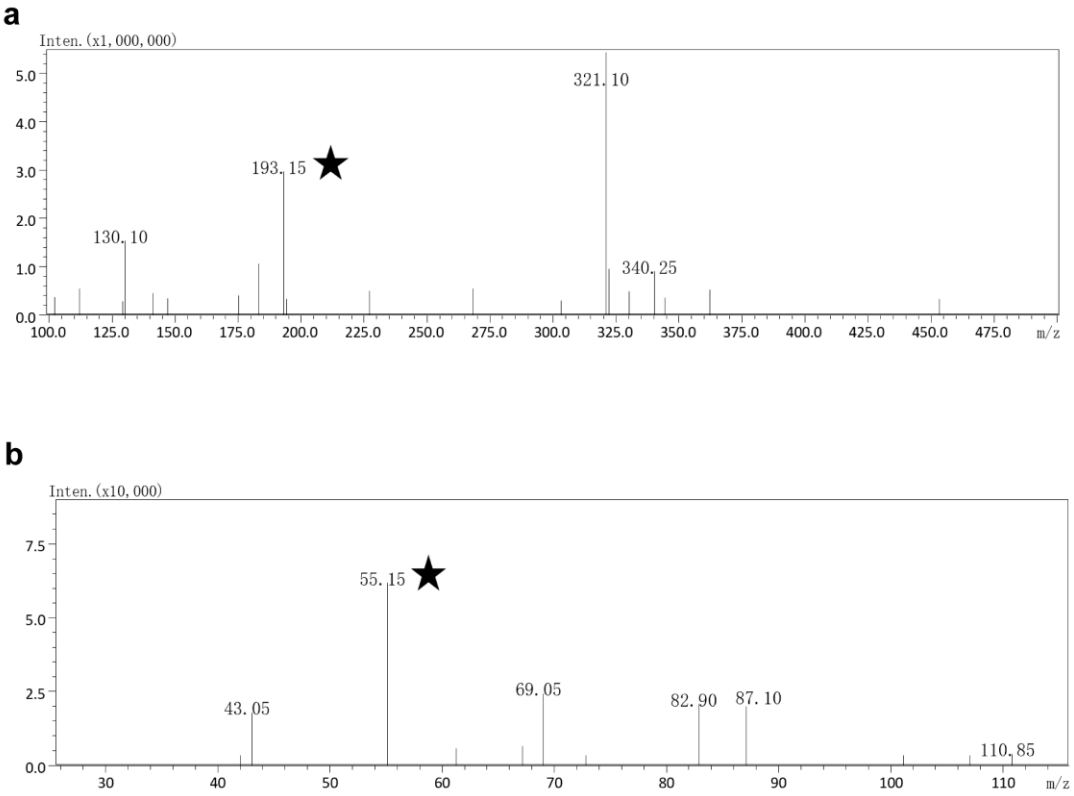


Fig. S6 LC-MS/MS analysis for **2b** in positive mode. a) Full-scan MS spectrum. The fragment ion chosen for quantification at m/z 193 is indicated by a star. b) MS/MS spectrum of m/z 193 ion. The daughter ion at m/z 55 chosen for quantification is indicated by a star.

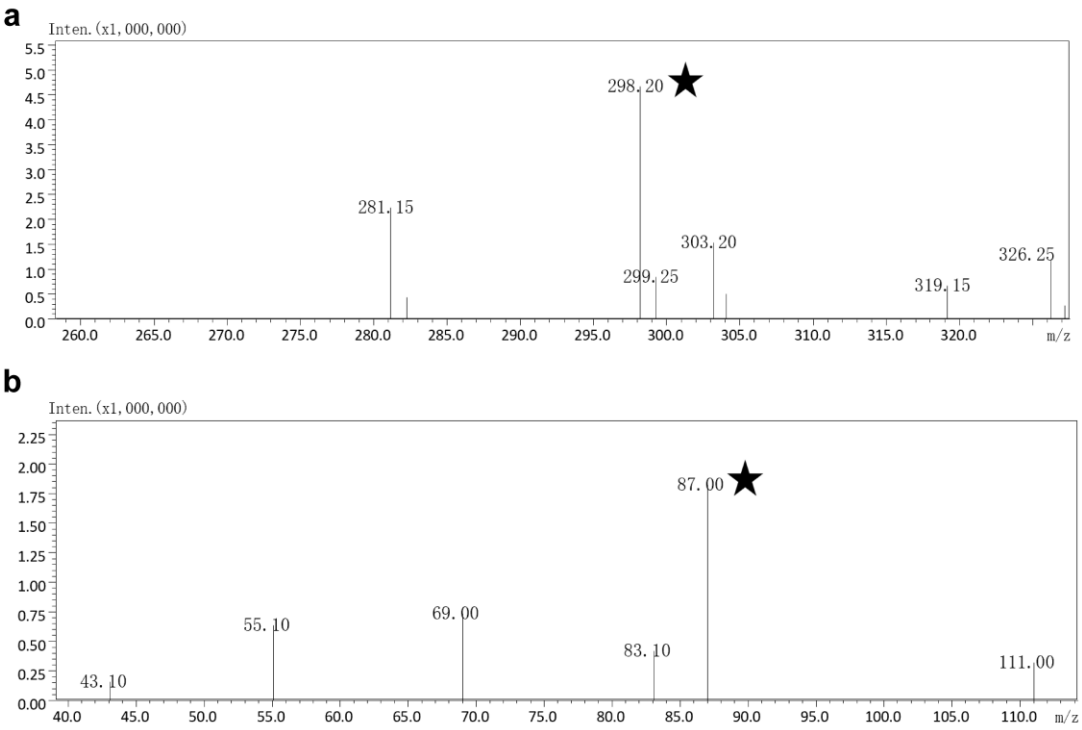


Fig. S7 LC-MS/MS analysis for **2c** in positive mode. a) Full-scan MS spectrum. The fragment ion chosen for quantification at m/z 298 is indicated by a star. b) MS/MS spectrum of m/z 298 ion. The daughter ion chosen for quantification at m/z 87 is indicated by a star.

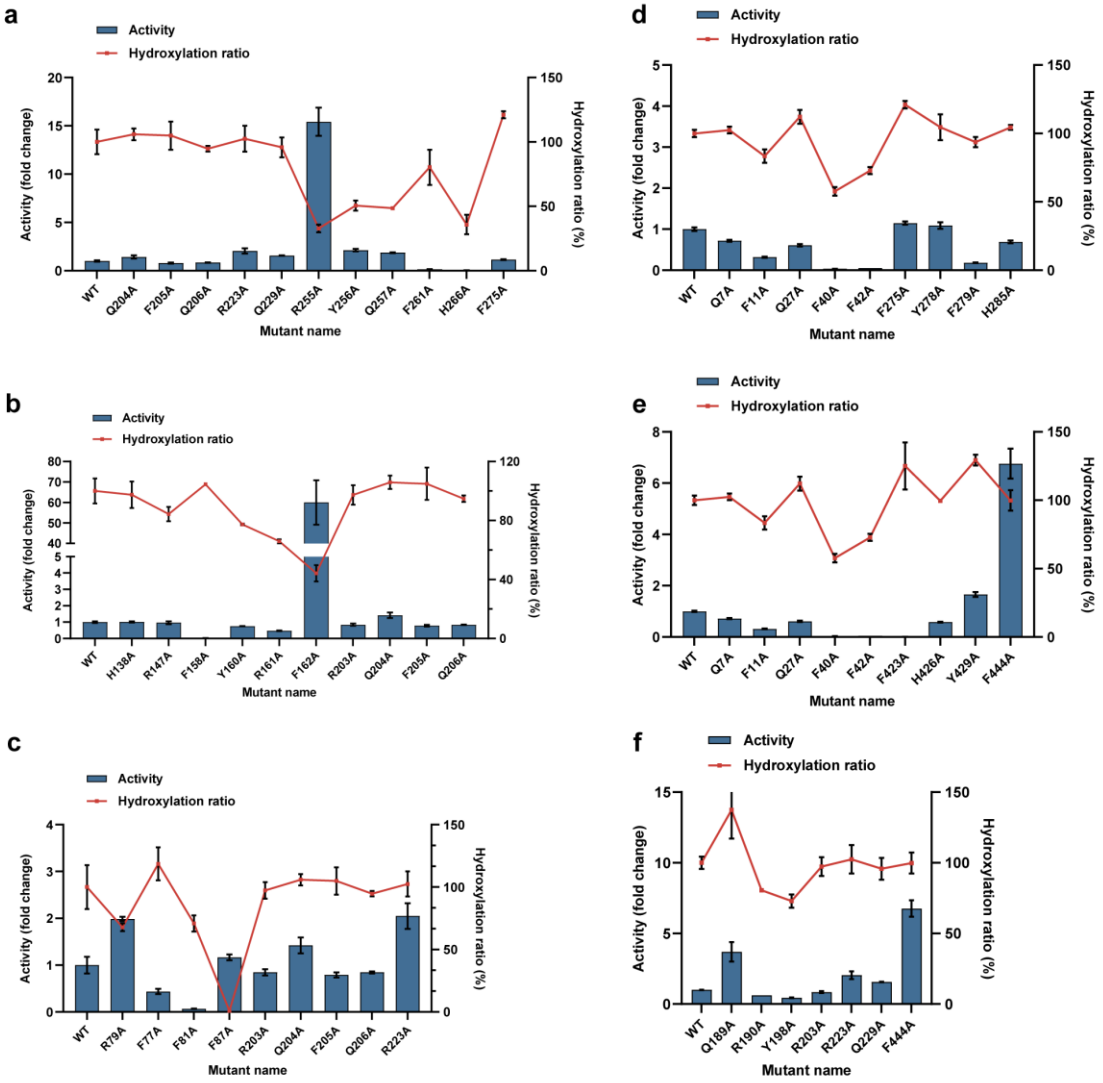


Fig. S8 Alanine scan of sterically hindered amino acids in regions with differences identified by DCCM. a) Region 1 Alanine Scan Results. b) Region 2 Alanine Scan Results. c) Region 3 Alanine Scan Results. d) Region 4 Alanine Scan Results. e) Region 5 Alanine Scan Results. f) Region 6 Alanine Scan Results.

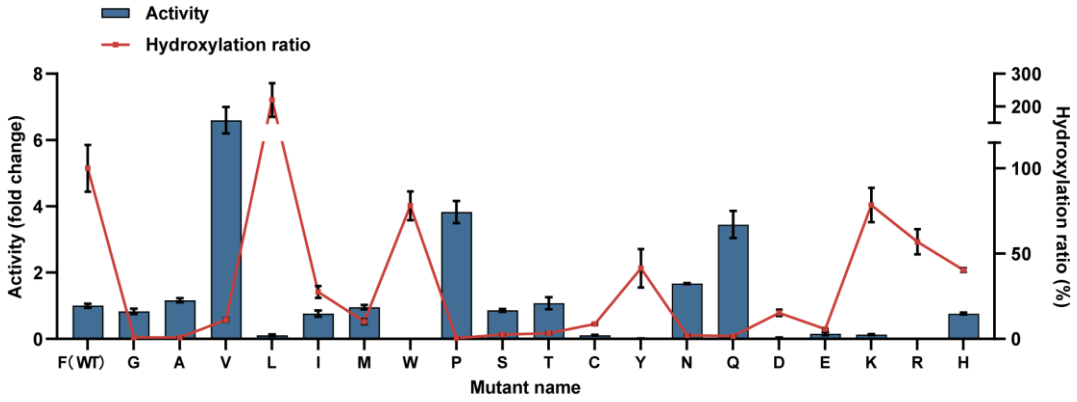


Fig. S9 Saturation mutagenesis at position F87.

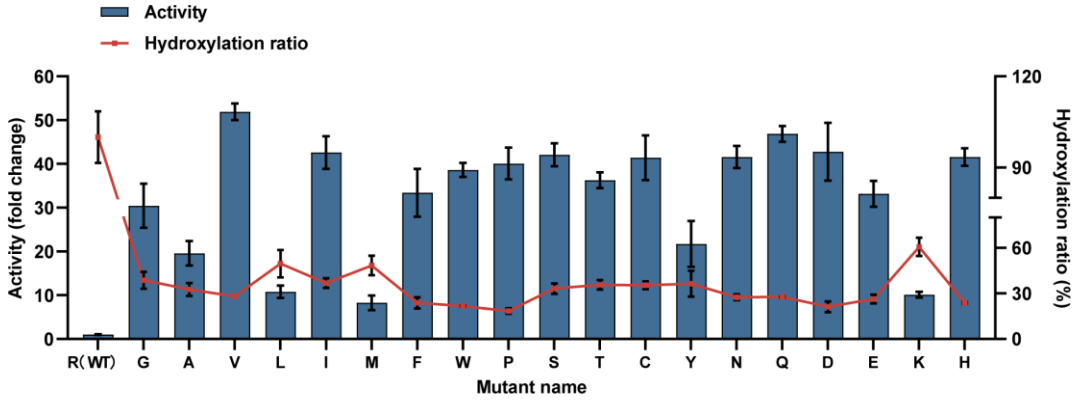


Fig. S10 Saturation mutagenesis at position R255.

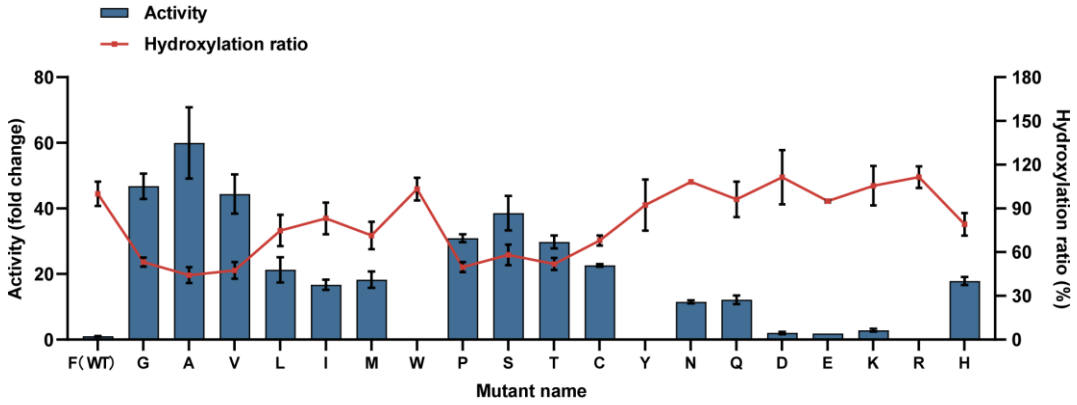


Fig. S11 Saturation mutagenesis at position F162.

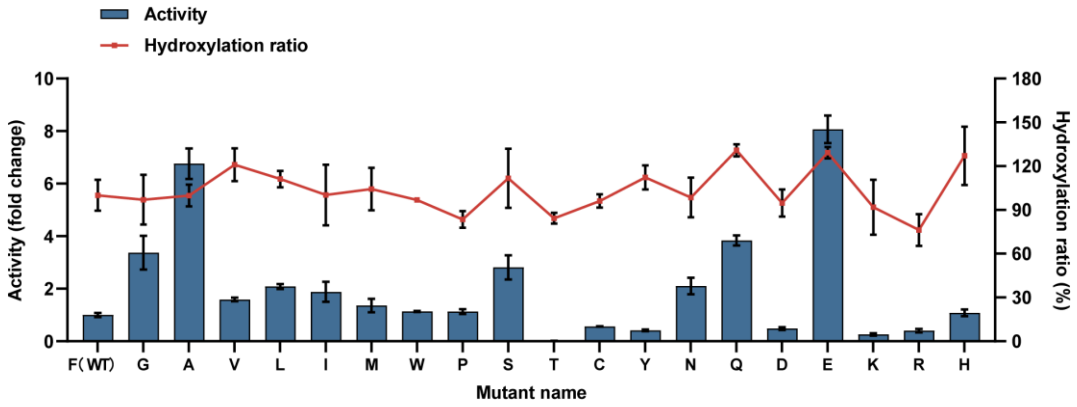


Fig. S12 Saturation mutagenesis at position F444.

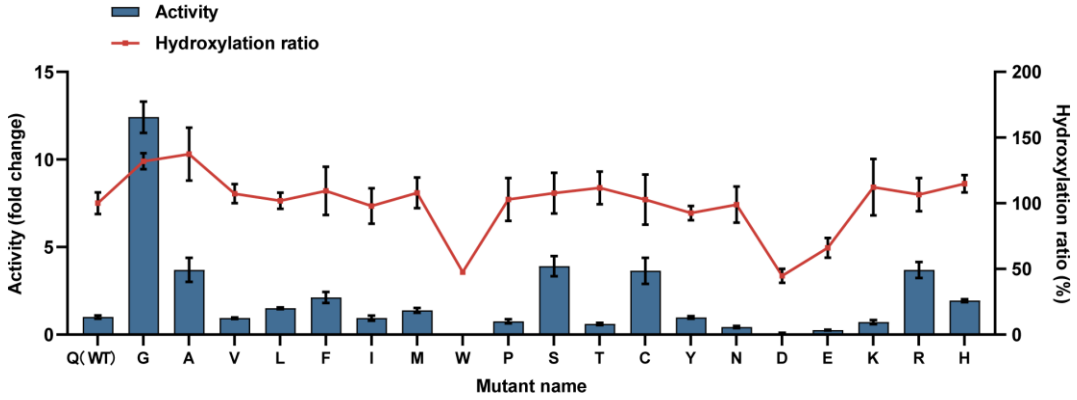


Fig. S13 Saturation mutagenesis at position Q189.

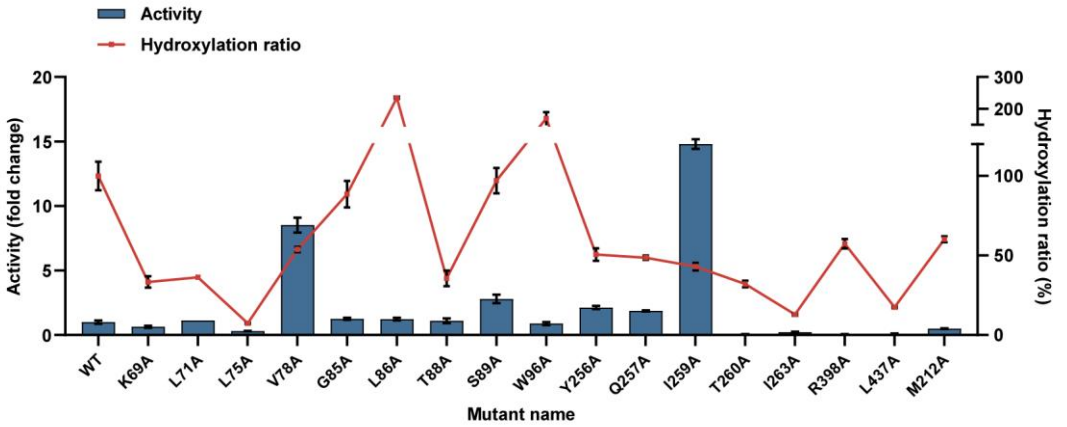


Fig. S14 Alanine scan of amino acids within 5 Å around F87.

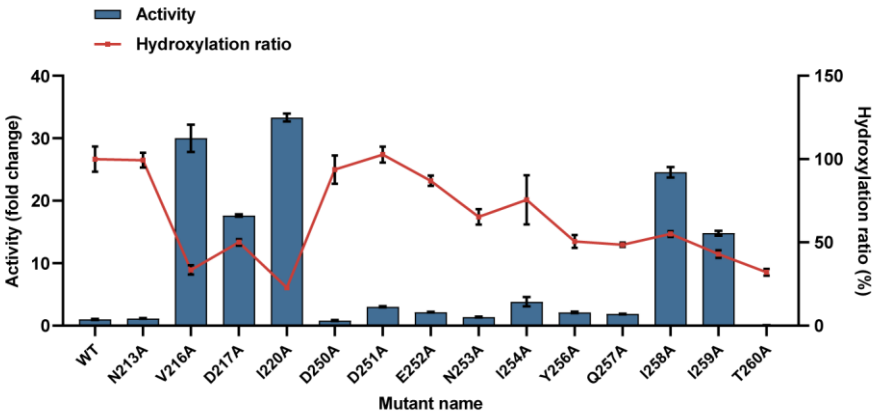


Fig. S15 Alanine scan of amino acids within 5 Å around R255.

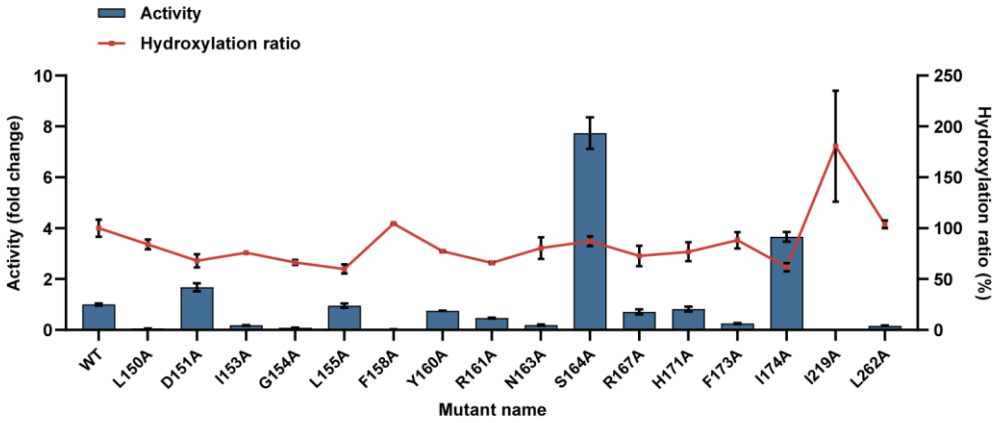


Fig. S16 Alanine scan of amino acids within 5 Å around F162.

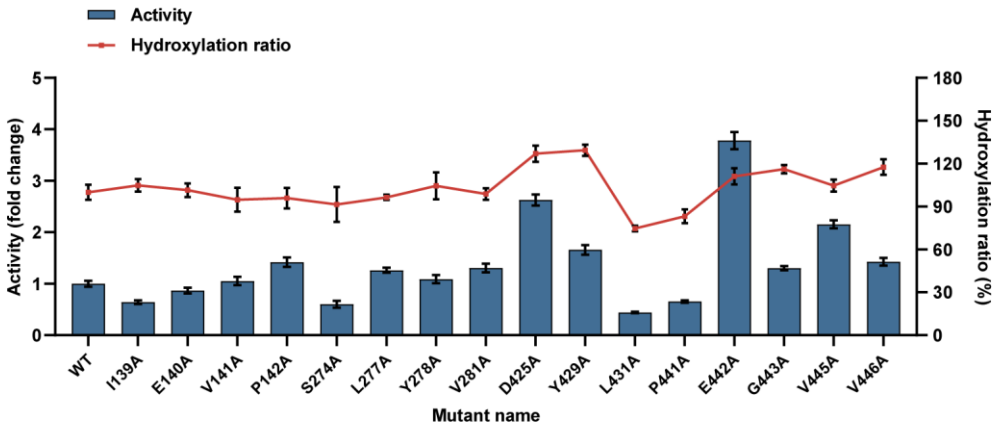


Fig. S17 Alanine scan of amino acids within 5 Å around F444.

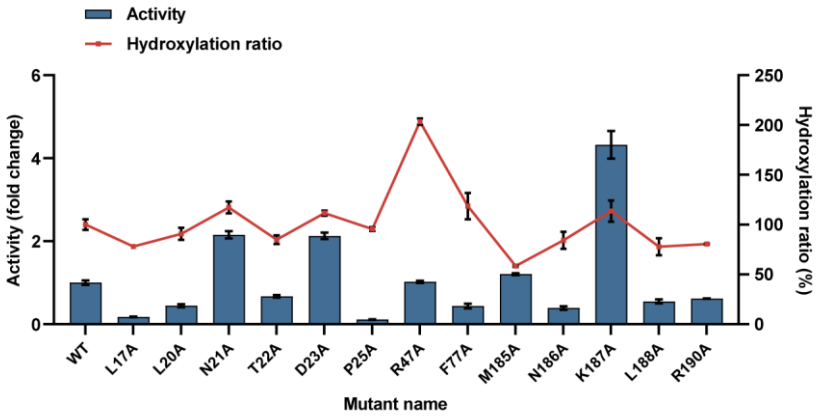


Fig. S18 Alanine scan of amino acids within 5 Å around Q189.

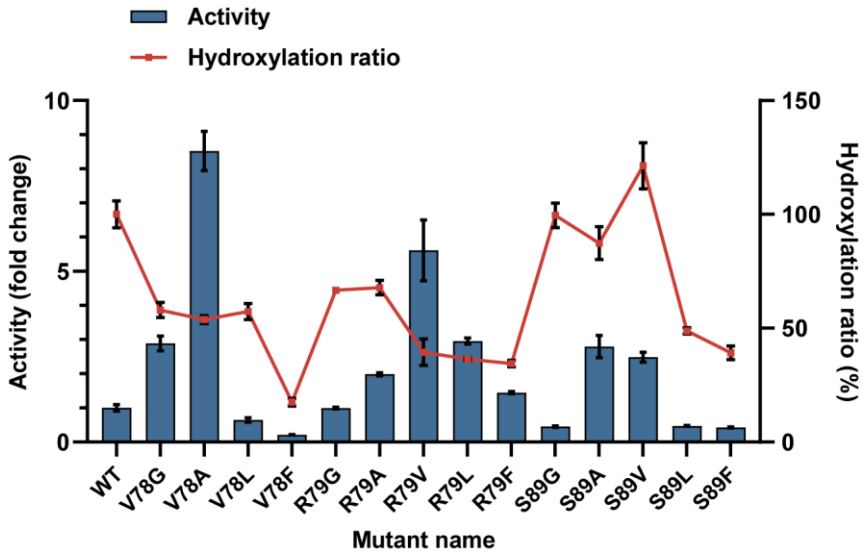


Fig. S19 GAVLF Scan of amino acids around F87.

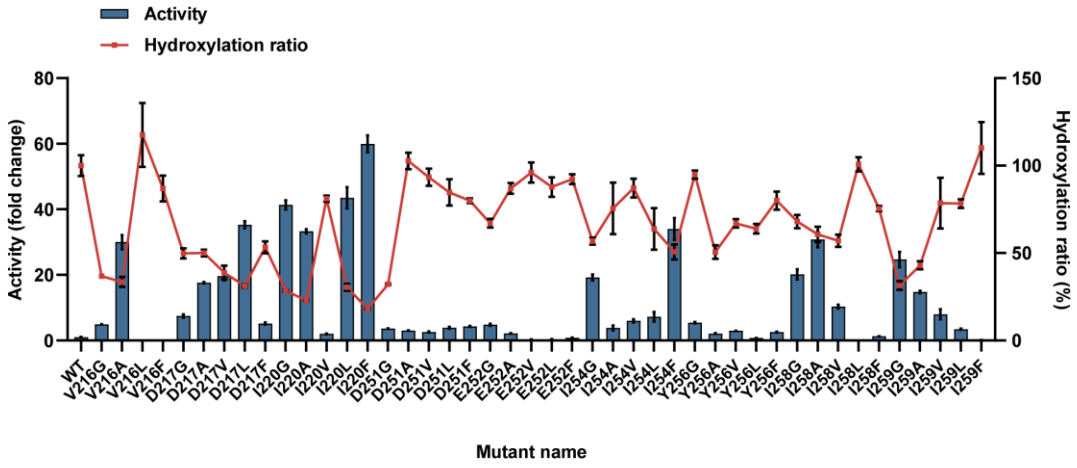


Fig. S20 GAVLF Scan of amino acids around R255.

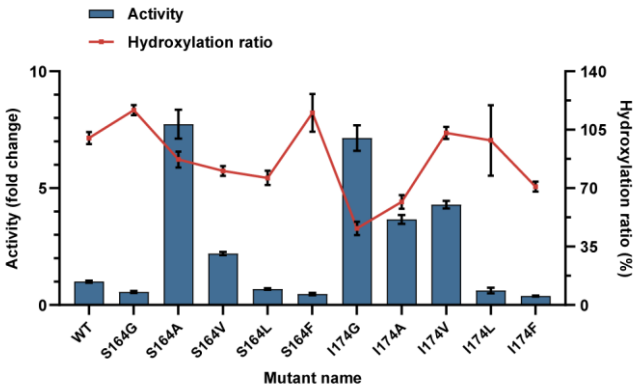


Fig. S21 GAVLF Scan of amino acids around F162.

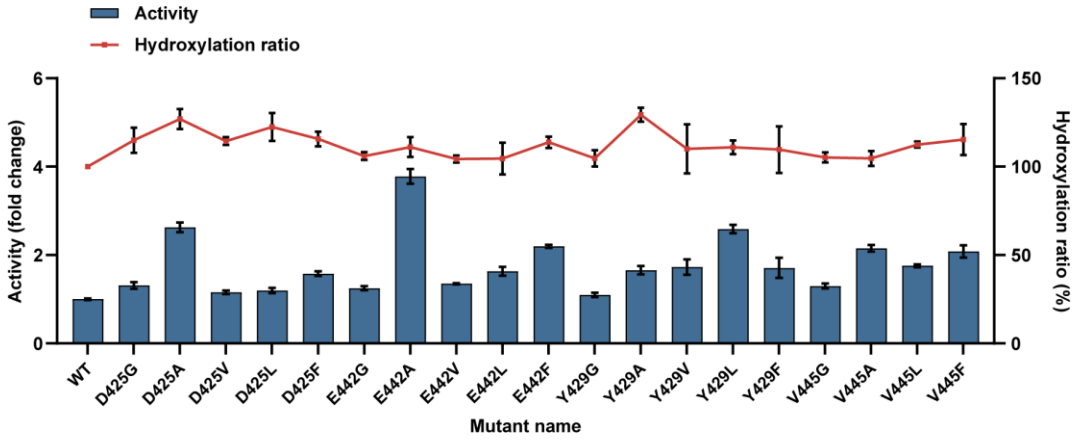


Fig. S22 GAVLF Scan of amino acids around F444.

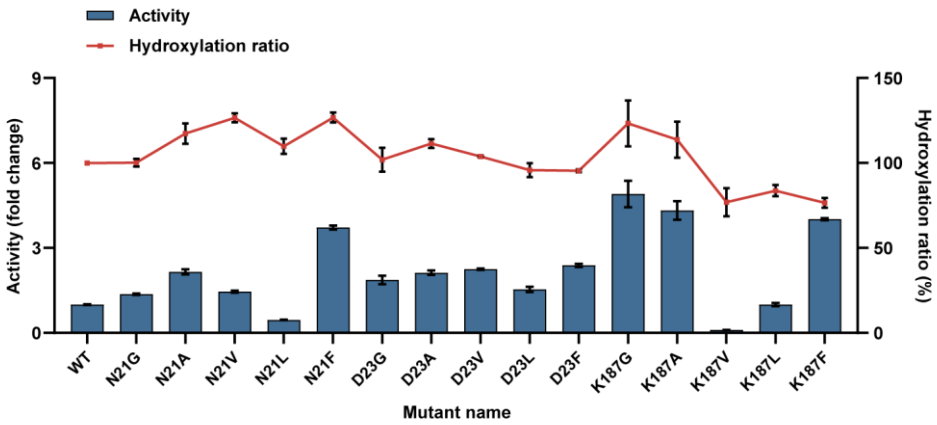


Fig. S23 GAVLF Scan of amino acids around Q189.

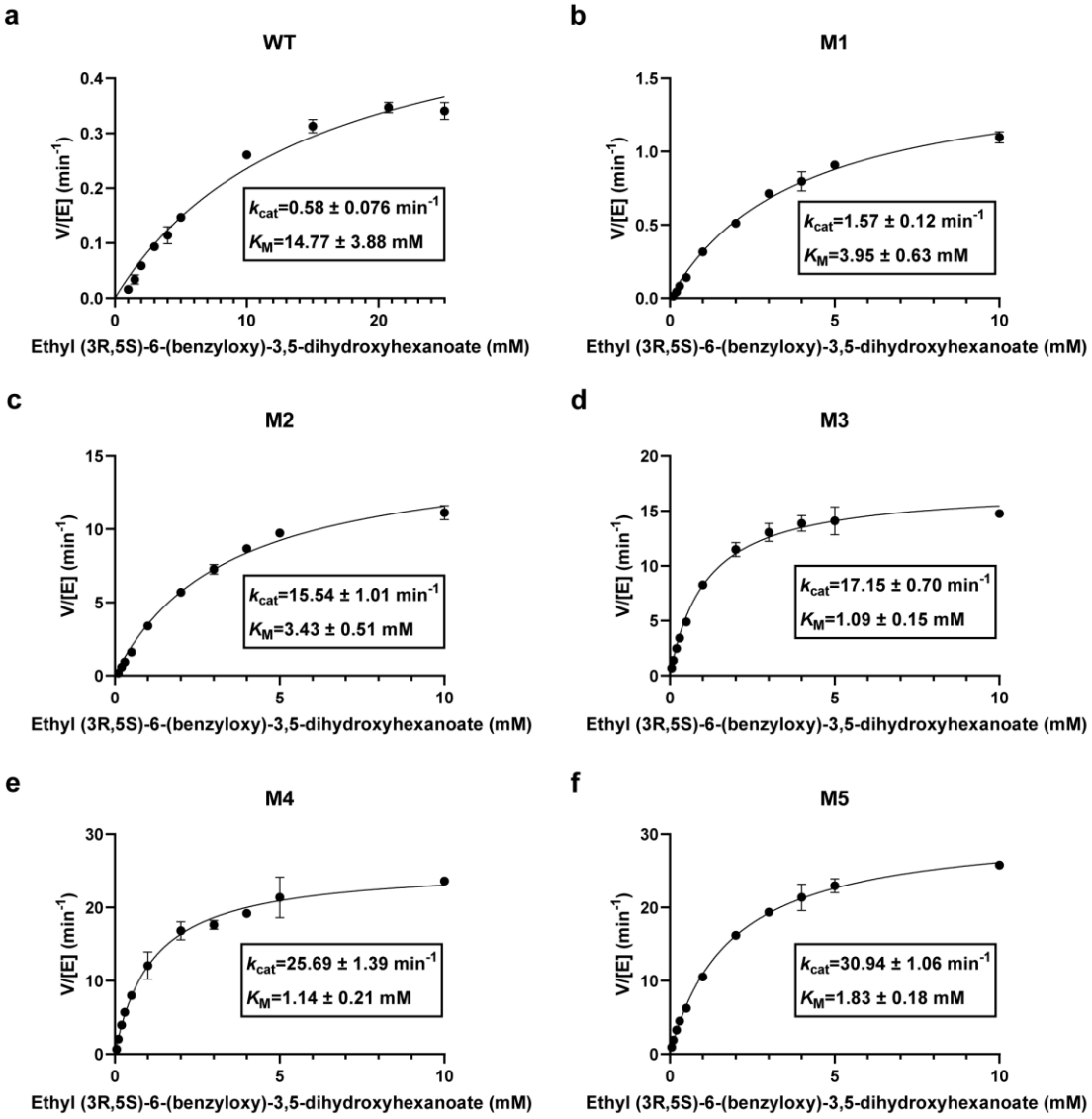


Fig. S24 Michaelis-Menten kinetics for CYP102A1 and its mutants. a) wild type, b) M1 variant, c) M2 variant, d) M3 variant, e) M4 variant and f) M5 variant.

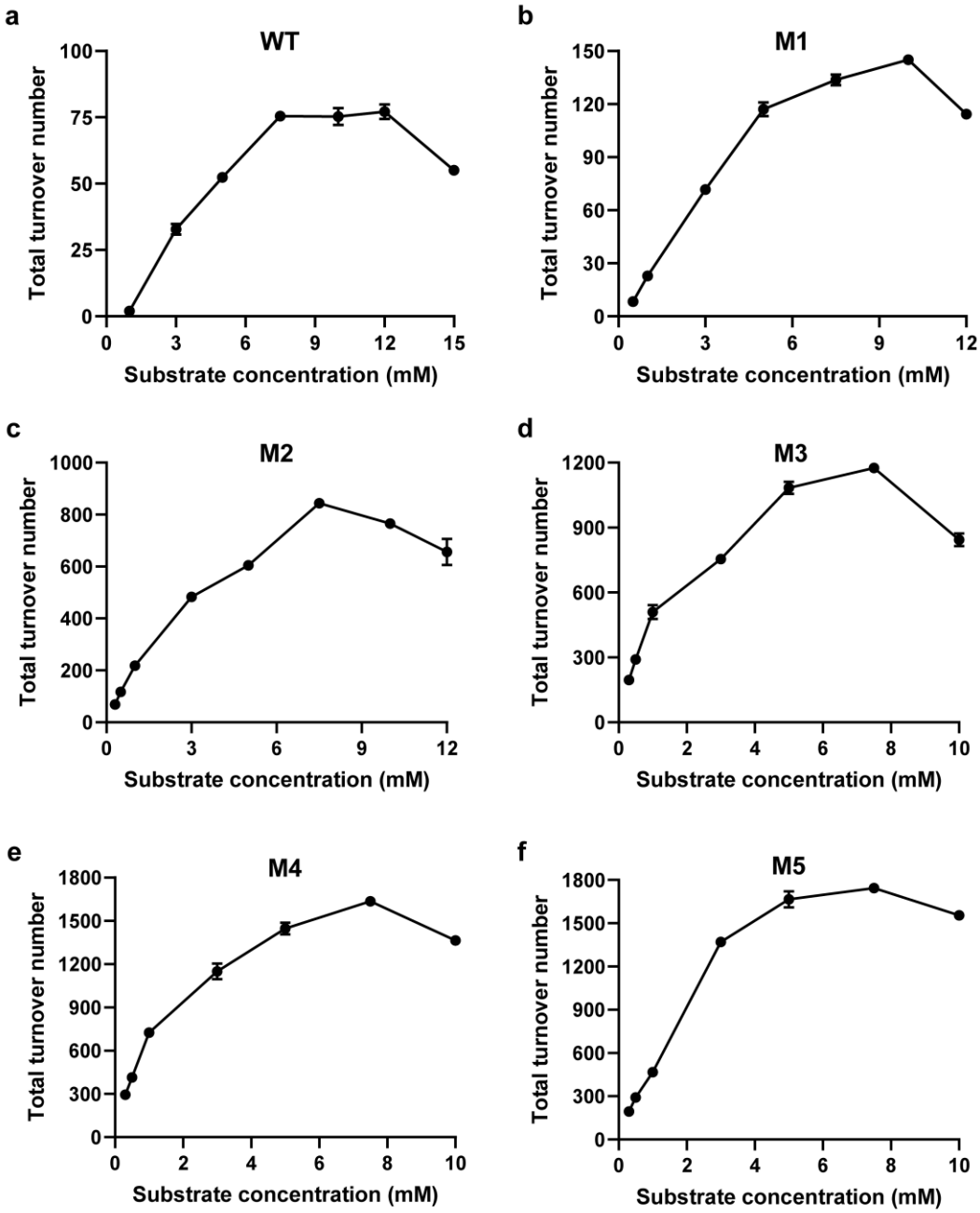


Fig. S25 Total turnover numbers (TTNs) of CYP102A1 and its mutants at different substrate concentrations. a) wild type, b) M1 variant, c) M2 variant, d) M3 variant, e) M4 variant and f) M5 variant.

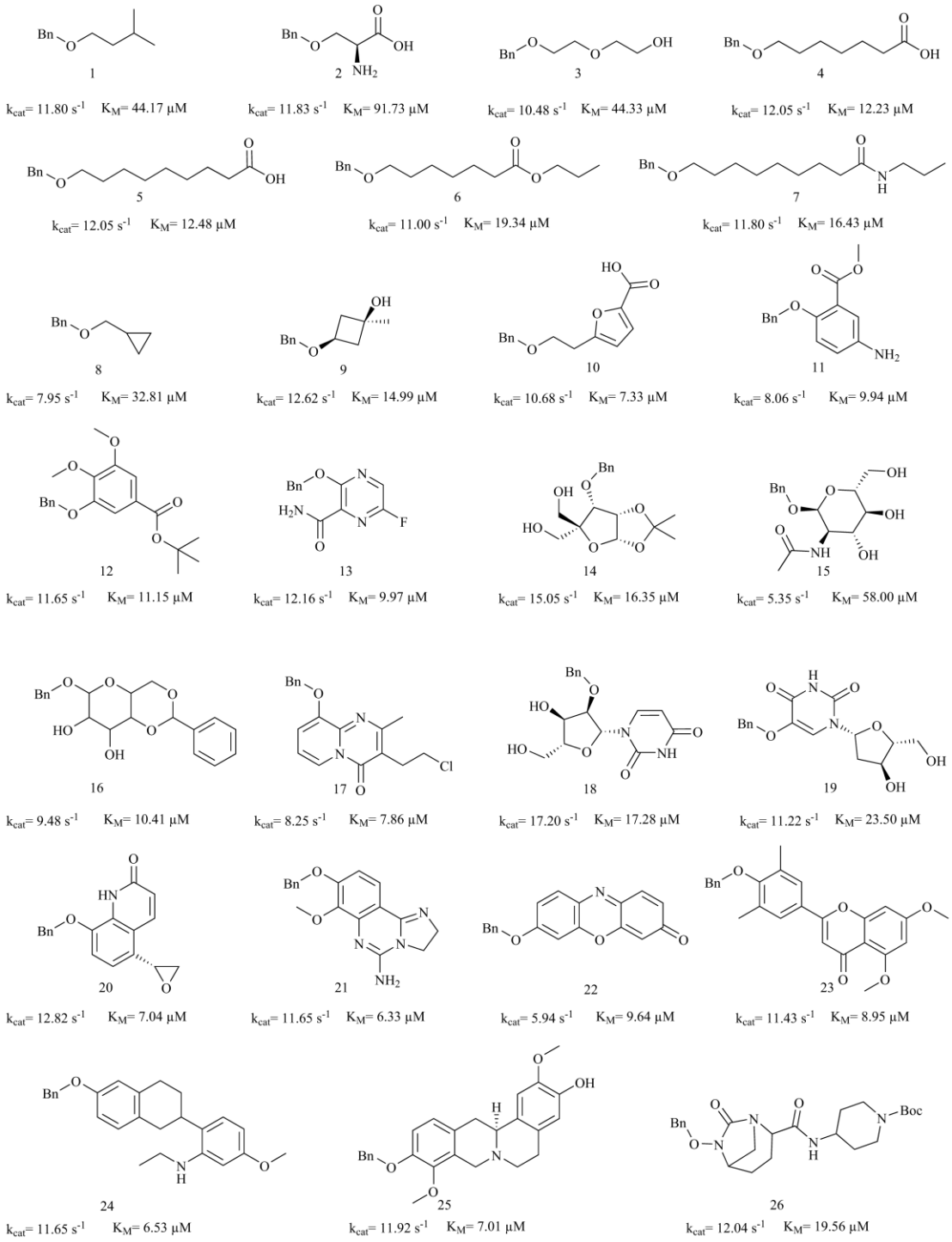


Fig. S26 Substrate scope of mutant M5 predicted using the enzyme–substrate kinetic parameter artificial intelligence model available on the DeepMolecules platform.

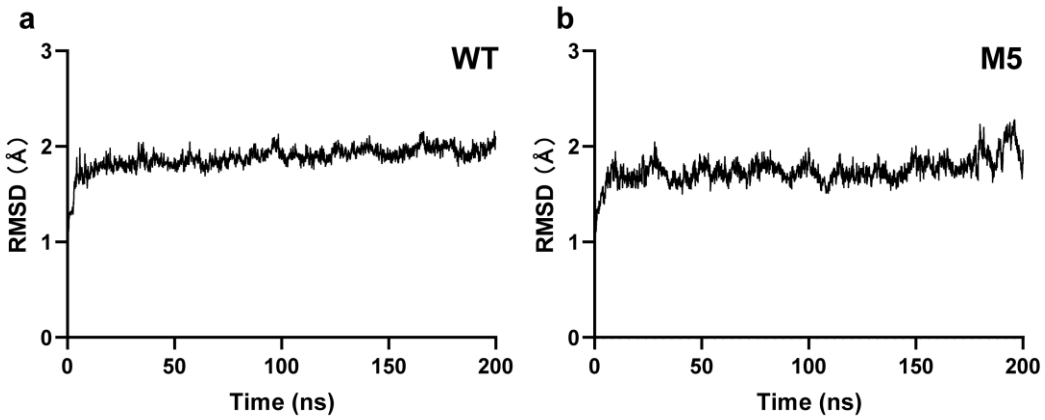


Fig. S27 a) RMSD analysis for WT. b) RMSD analysis for M5.

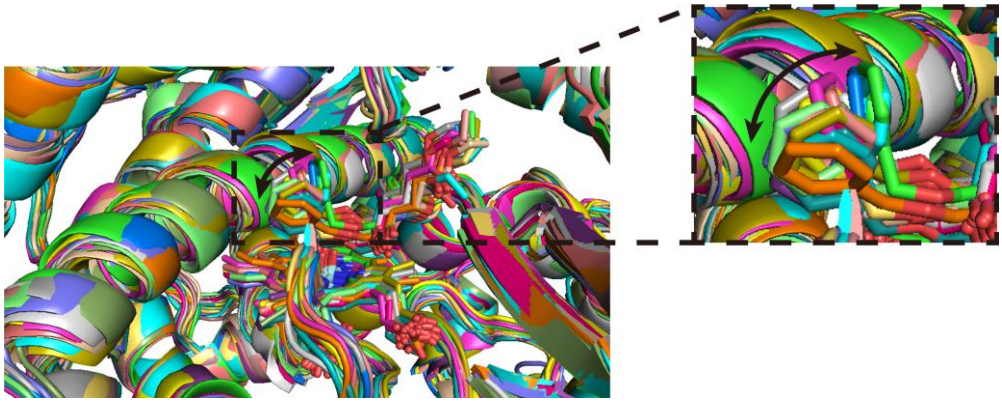


Fig. S28 Superposition of the structures from extracting snapshots at fixed time intervals in MD trajectory of M5. The boxed region highlights the fluctuating phenyl ring of the substrate, which is shown in an enlarged view. The arrows indicate a predominant direction of motion.

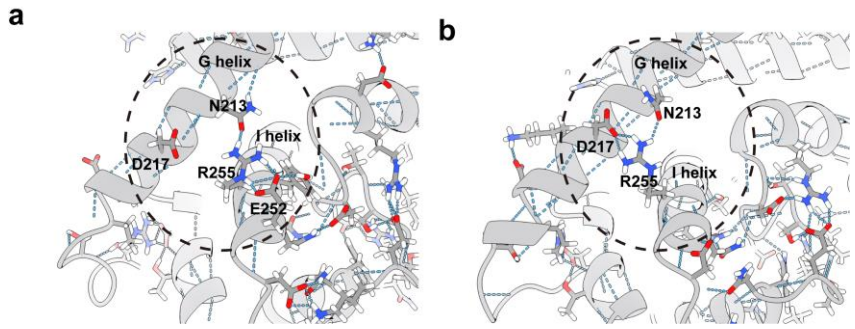


Fig. S29 a) Local hydrogen bond network (hydrogen bonds indicated by blue dashed lines) between I-helix and G-helix in WT. b) Local hydrogen bond network between I-helix and G-helix in M5 mutant.

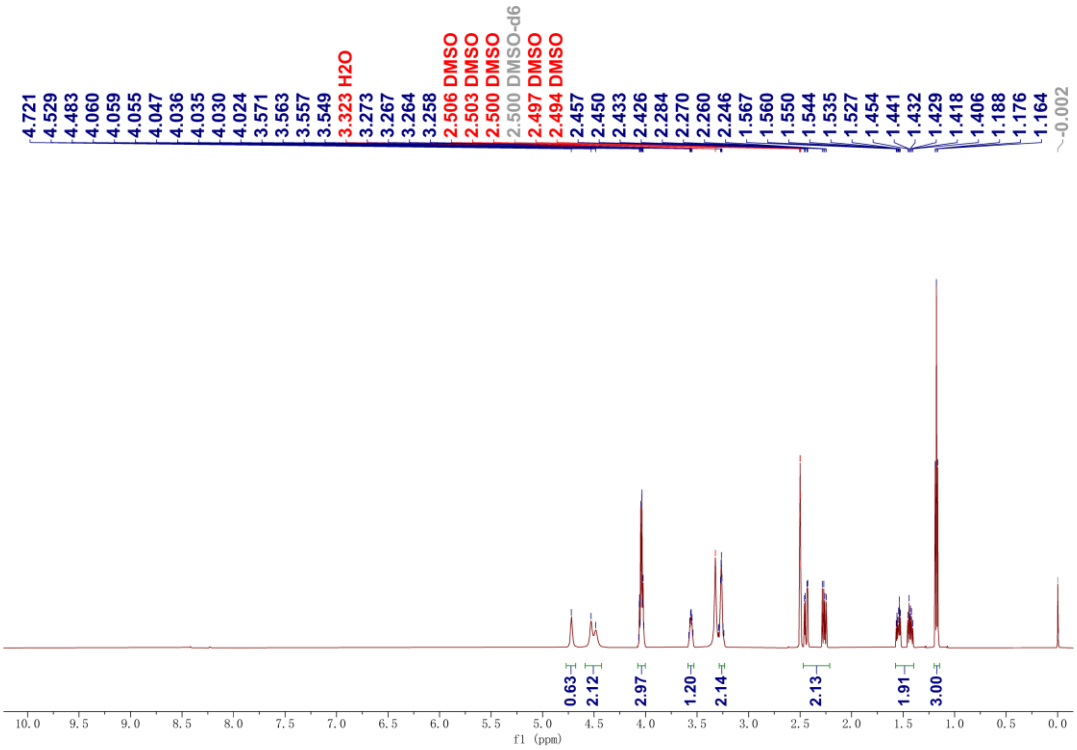


Fig. S30 ^{1}H NMR spectrum of **2a** in DMSO-d_6 (600 MHz).

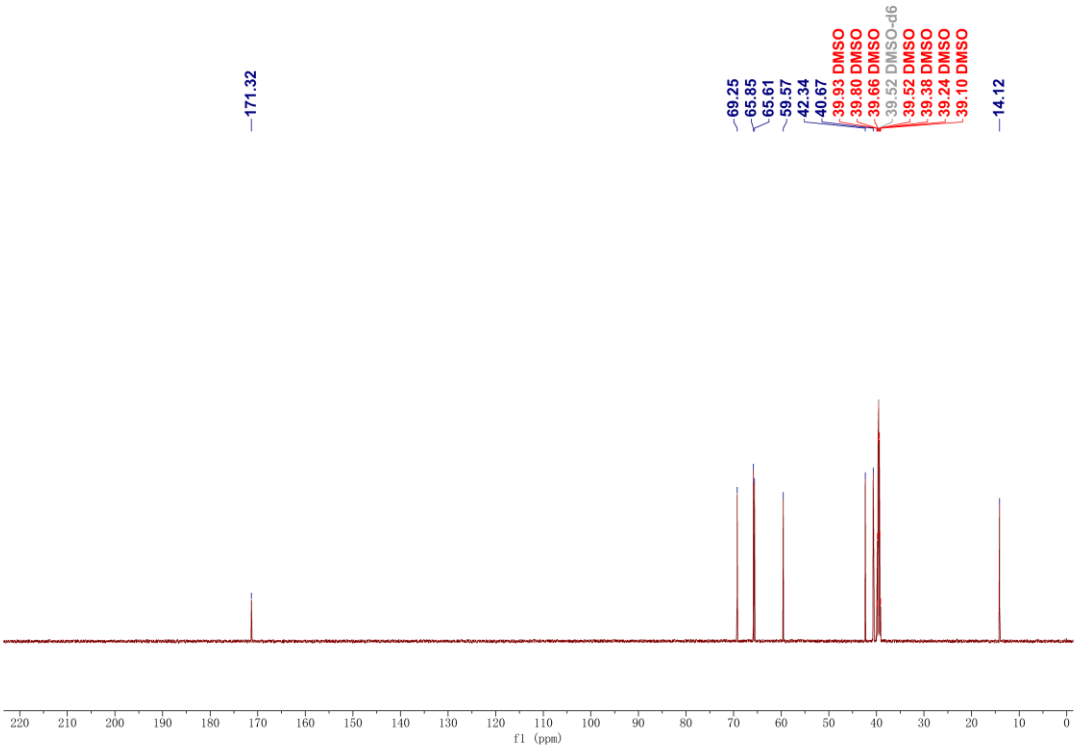


Fig. S31 ^{13}C NMR spectrum of **2a** in DMSO-d_6 (150 MHz).

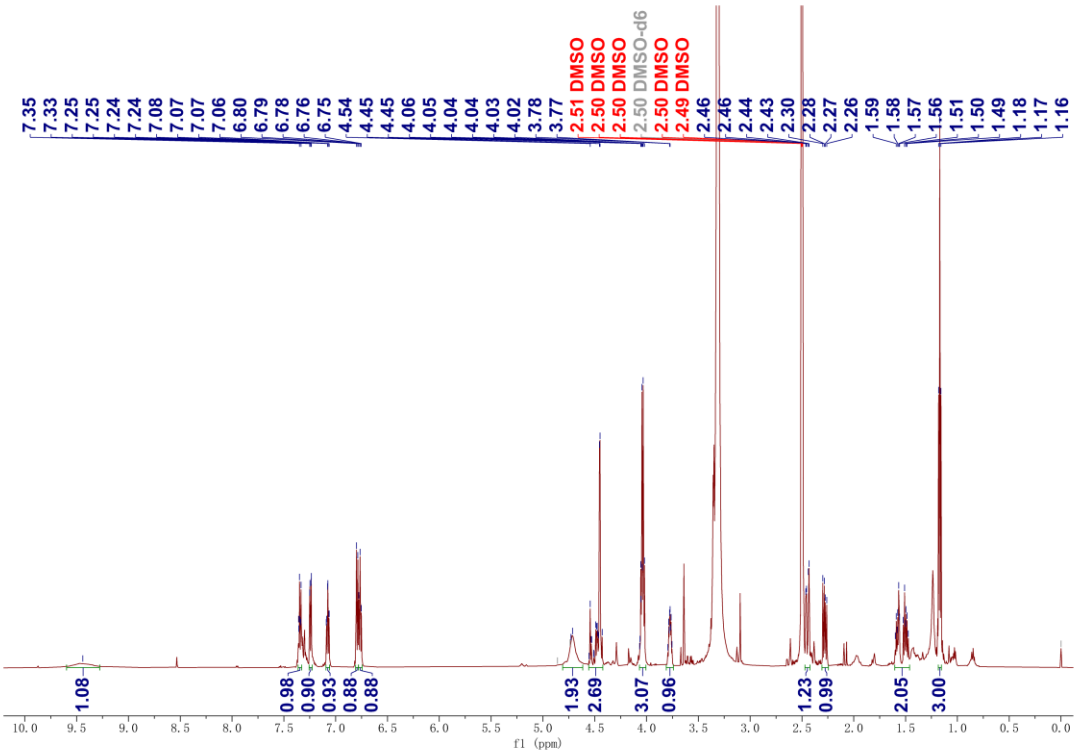


Fig. S32 ^{1}H NMR spectrum of **2b** in DMSO-d_6 (600 MHz).

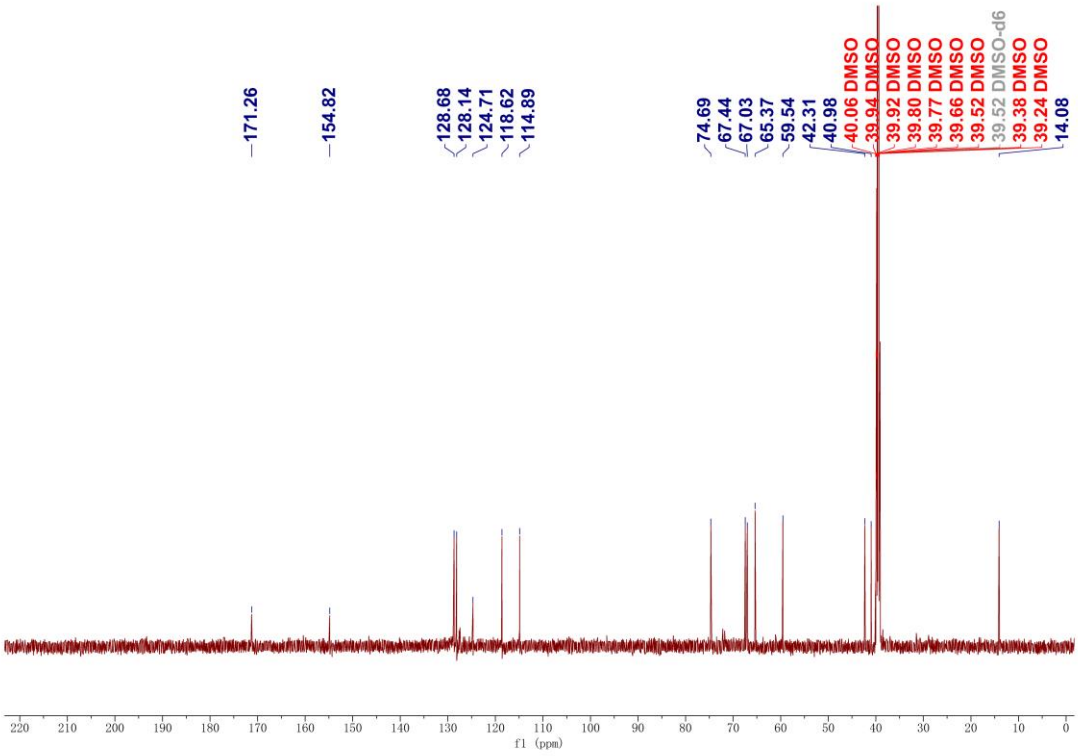


Fig. S33 ^{13}C NMR spectrum of **2b** in DMSO-d_6 (150 MHz).

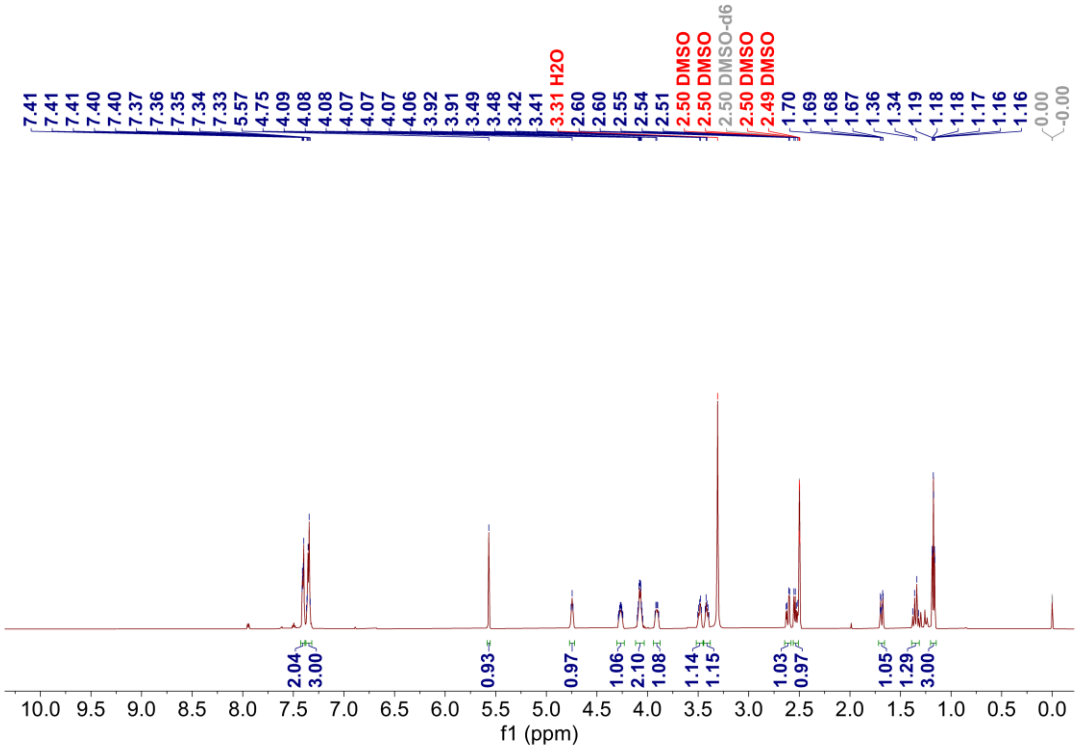


Fig. S34 ^{1}H NMR spectrum of **2c** in DMSO-d_6 (600 MHz).

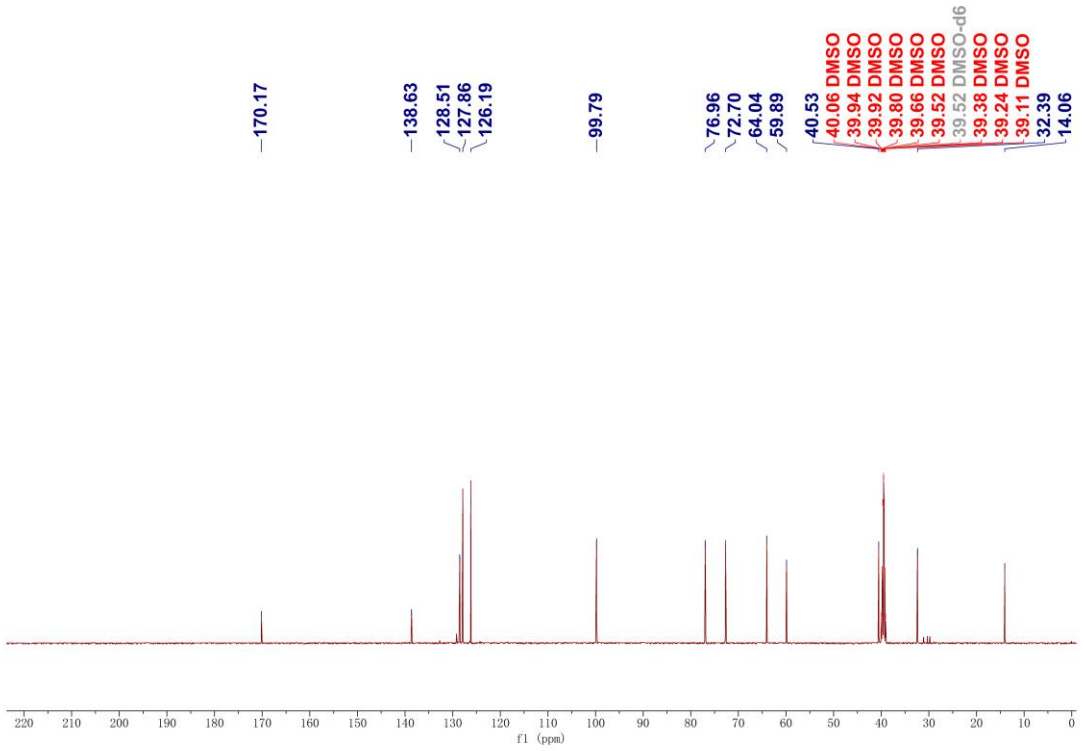


Fig. S35 ^{13}C NMR spectrum of **2c** in DMSO-d_6 (150 MHz).

Reference

1. F. P. Guengerich, M. V. Martin, C. D. Sohl and Q. Cheng, *Nat. Protoc.*, 2009, **4**, 1245-1251.
2. B. M. A. Lussenburg, L. C. Babel, N. P. E. Vermeulen and J. N. M. Commandeur, *Anal. Biochem.*, 2005, **341**, 148-155.
3. D. Y. Hwang, J. S. Cho, J. H. Oh, S. B. Shim, S. W. Jee, S. H. Lee, S. J. Seo, H. G. Kang, Y. Y. Sheen, S. H. Lee and Y. K. Kim, *Int. J. Toxicol.*, 2005, **24**, 157-164.
4. K. Umehara, Y. Shimokawa and G. Miyamoto, *Biol. Pharm. Bull.*, 2002, **25**, 682-685.
5. C. L. Crespi, V. P. Miller and B. W. Penman, *Anal. Chem.*, 1997, **248**, 188-190.
6. T. L. Domanski, Y.-A. He, K. K. Khan, F. Roussel, Q. Wang, and J. R. Halpert, *Biochemistry*, 2001, **40**, 10150-10160.
7. C. Sridar, U. M. Kent, K. Noon, A. McCall, B. Alworth, M. Foroozesh and P. F. Hollenberg, *Drug Metab. Dispos.*, 2008, **36**, 2234-2243.
8. L. Long and M. E. Dolan, *Clin. Cancer Res.*, 2001, **7**, 4239-4244.
9. S. Kumar, D. R. Davydov and J. R. Halpert, *Drug Metab. Dispos.*, 2005, **33**, 1131-1136.
10. D. Moutinho, C. C. Marohnic, S. P. Panda, J. Rueff, B. S. Masters and M. Kranendonk, *Drug Metab. Dispos.*, 2012, **40**, 754-760.
11. J. E. Bissada, V. Truong, A. A. Abouda, K. J. Wines, R. D. Crouch and K. D. Jackson, *Drug Metab. Dispos.*, 2019, **47**, 1257-1269.