
参赛队员姓名：张鹭飞、殷岂墨、郁雅涵

中学：南京外国语学校

省份：江苏

国家/地区：中国

指导教师姓名：曹飞

指导教师单位：南京工业大学

论文题目：Improvement of regioselectivity for L-pipecolic acid hydroxylation by site-directedly mutated L-proline *cis*-4-hydroxylase

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。以本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：张鹭飞、郁雅涵、殷岂墨

指导老师：曹飞

2023 年 9 月 12 日

Improvement of regioselectivity for L-pipecolic acid hydroxylation by site-directedly mutated L-proline *cis*-4-hydroxylase

Lufei Zhang¹, Qimo Yin¹, Yahan Yu¹, Fei Cao²

(1, 2022 grand, Nanjing Foreign Language School, Nanjing, Jiangsu, 210008;

2, College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University

30 South Puzhu Road, Nanjing, 211816)

Abstract

The non-heme Fe²⁺/ α -ketoglutarate-dependent dioxygenases (KDDs) have wide range of applications in the synthesis of chiral pharmaceutical building blocks and the functionalization of C(sp³)-H, due to its inexpensive cofactors and high efficiency on amino acids hydroxylation. However, for amino acids with multiple reaction sites (e.g., L-pipecolic acid, L-PA), L-proline *cis*-4-hydroxylase (*cis*-P4H), because of its low regioselectivity, usually outputs almost equal *cis*-5/*cis*-3 hydroxylated isomers which are hard to separate and purify. Here, we obtained a hydroxylase from *Kordia jejudonensis* (KjPH) with higher hydroxylation regioselectivity for C5 of L-PA (*cis*-5/*cis*-3 ratio about 20:1). Together with the comparisons of sequence, structure, and catalytic chamber of KjPH, Smp4H (previously studied, *cis*-5/*cis*-3 ratio around 1:1), and MIP4H (*cis*-5/*cis*-3 = 1:7, from *Mesorhizobium loti*). F95 in KjPH was found to be a key residue affecting regioselectivity. Site-directed mutagenesis on F95 showed that the regioselectivity of F95Y mutant was further improved, and the *cis*-5/*cis*-3 ratio reached 55:1. Combined with the results of molecular dynamics (MD) simulations, the regioselectivity improvement of F95Y variant is due to the formation of electrostatic repulsive interaction between the phenol hydroxyl group (Y95) and the carboxyl group (L-PA), which results in the distance increasing between substrate C3 and the Fe²⁺ catalytic core.

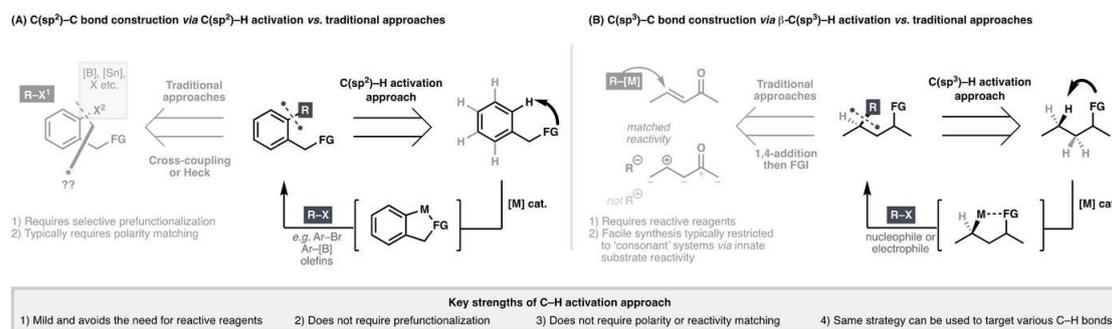
Keywords: L-proline *cis*-4-hydroxylase, *cis*-5-hydroxy-L-pipecolic acid, regioselectivity, electrostatic repulsion effect

TABLE OF CONTENTS

1. INTRODUCTION	5
2. MATERIALS AND METHODS.....	9
2.1 STRAINS, PLASMIDS AND REAGENTS	9
2.2 STRUCTURAL BIOINFORMATICS ANALYSIS OF THE KJPH AND HOMOLOGY MODELING.....	9
2.3 SITE-DIRECTED MUTAGENESIS, EXPRESSION AND PURIFICATION OF RECOMBINANT KJPH.....	9
2.4 <i>IN VITRO</i> HYDROXYLATION OF L-PA.....	12
2.5 MOLECULAR DYNAMICS SIMULATIONS.....	12
2.6 DERIVATIZATION AND DETECTION OF L-PA, CIS-5-HPA AND CIS-3-HPA BY HPLC.....	14
3. RESULTS AND DISCUSSION	15
3.1 IDENTIFICATION OF REGIOSELECTIVITY DIFFERENCE FOR KJPH AND SMP4H	15
3.2 STRUCTURE-GUIDED SCREENING FOR KEY RESIDUES	16
3.3 REGIOSELECTIVITY IMPROVEMENT VIA SITE-DIRECTED MUTAGENESIS.....	19
3.4 MOLECULAR DYNAMICS SIMULATIONS.....	21
4. CONCLUSION.....	24
5. REFERENCE	25
6. ACKNOWLEDGEMENTS.....	26
7. SUPERVISOR AND STUDENTS INFORMATION	28

1. Introduction

The selective functionalization of inert C–H bond under mild conditions is a significant but challenging reaction in chiral compounds synthesis due to the high strength and weak polarity of C–H bonds (with the bond dissociation energy of 96–101 kcal/mol and pK_a about 50)[1, 2]. Currently, a variety of chemical strategies for C–H functionalization have been developed, including transition metal catalysts[3], radical chain transfer[4], and photoredox catalysis[5]. However, these C–H activated reactions all depend on large and strongly coordinated oriented-groups, while the installation and removal of oriented-groups often lower the atomic economy of the reactions (**Scheme. 1**) [6]. Directly oxidative hydroxylation, as one of the most commodious and cost-efficient approaches used in C–H functionalization, has been broadly researched [7, 8]. Acquiring regioselectivity-hydroxylated products from a given substrate possessing multiple reaction sites still remain difficult.



Scheme 1. Overview of C–C bond formation by transition-metal-catalyzed C–H activation and two examples of how they can offer a conceptually orthogonal approach towards organic synthesis: A) by C(sp²)–H activation and B) by β -C(sp³)–H activation. Source: Advancing the Logic of Chemical Synthesis: C–H Activation as Strategic and Tactical Disconnections for C–C Bond Construction, *Angew Chem Int Ed*, 2020, 60(29): 15767–15790, DOI:10.1002/anie.202011901

Enzymes are promising alternatives to settle this synthetic challenge on account of their intrinsic chem-, regio- and stereoselectivity, and the capacity to be modified for sustainable activity and higher regioselectivity [9, 10]. Non-heme Fe²⁺/ α -ketoglutarate-dependent dioxygenases (KDDs) have exhibited outstanding ability in amino acids

hydroxylation[11] and reveal remarkable applications compared to NAD(P)H-dependent cytochrome P450 enzyme families because the cofactor (α -ketoglutarate, α -KG) of KDDs is much cheaper than NAD(P)H[12]. However, the regioselectivity of natural KDDs for amino acids with multiple hydroxylation sites still remains undesired. For instance, L-pipecolic acids (L-PA) are usually converted into regio-isomers mixture, *cis*-3 and *cis*-5-hydroxy-L-pipecolic acid (*cis*-5-HPA), catalyzed by L-proline *cis*-hydroxylase (*cis*-P4H) (**Fig. 1A**)[13, 14]. Only *cis*-5-HPA can be used to prepare avibactam, a β -lactamase inhibitor (**Fig. 1B**) [15].

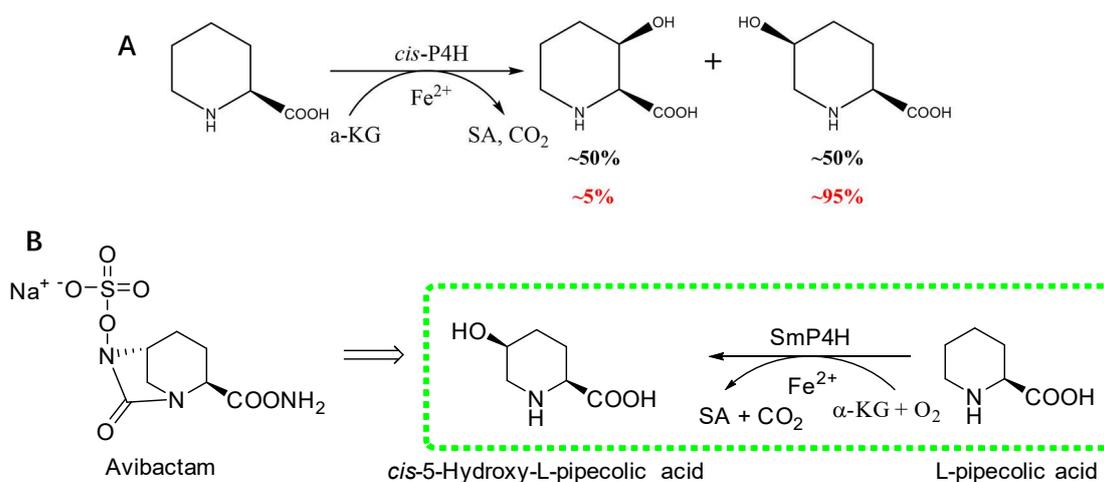


Fig. 1. A. The structure Avibactam and the enzymatic manufacture of *cis*-5-hydroxy-L-pipecolic acid from L-pipecolic acid. B. *cis*-P4H from *Sinorhizobium meliloti* (black) and *Kordia jejudonensis* (red) hydroxylates the C3 and C5 carbons of L-PA yielding two regioisomers, *cis*-5-HPA and *cis*-3-HPA. α -KG, α -ketoglutarate; SA, succinic acid.

Recently, great progresses have been made in enhancing regioselectivity and activity of KDDs via site-directed mutagenesis on specific amino acid residues in enzyme catalytic chamber so as to change the steric hindrance, hydrophilic/hydrophobic or electrostatic effects between the residues of chamber and the substrate. Wang and co-workers performed saturated mutations on 6 residues (Q161, T162, D176, F178, D257 and E260) of the catalytic chamber in L-lysine hydroxylase (NkLH4), obtaining a highly active mutant MT3 (Q161N/T162A/F178Y/E260D) with a 24.97-fold increase of k_{cat}/K_m compared to the wild-type enzyme, and it was suggested that the mutations around the active sites lead to the movement of the K156-F180 and A258-I262 rings and the expansion of the substrate binding cavity, which reduces the steric hindrance near the

C4 of the substrate, facilitates the enzymatic hydroxylation of C4 site easier, and produces the enantiomer pure (2S,4R) -4-hydroxylysine (**Fig. 2A**) [16]. Based on the X-ray crystal structure of L-proline *cis*-4-hydroxylase, Koketsu conducted three rounds of mutation on 10 sites of the catalytic chamber in SmP4H, and the obtained triple mutant (V97F/V95W/E114G) displayed significant increase in C5 hydroxylated products ratio (>95%) with marked decrease in activity (less than 10% of the wide type) [13]. Wu obtained a single-site variant (T244S) with enhanced regioselectivity by mutating 11 residues within 4 Å of the substrate on the catalytic chamber of L-isoleucine dioxygenase (IDO), in which the most frequent distance (MFD) difference value of Fe²⁺ to substrate C4 and C5 increased to 0.81 Å from 0.12 Å, resulting in the regioselectivity improvement of T244S and giving a maximum 4-hydroxyisoleucine purity of 96.6% (**Fig. 2B**)[17].

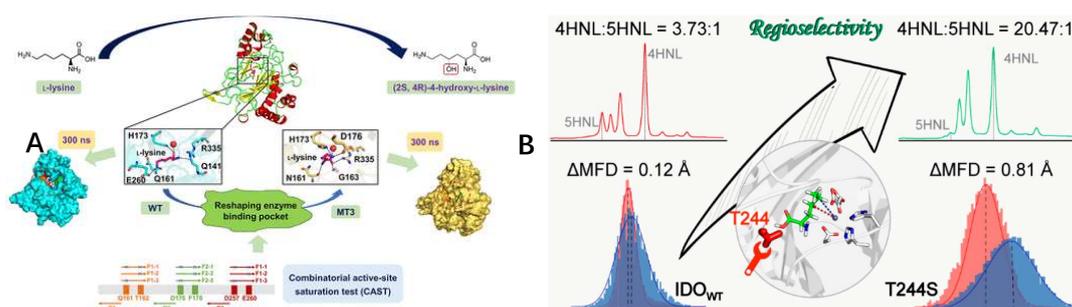


Fig. 2. A, Reshaping the binding pocket of lysine hydroxylase for enhanced activity. Source: *ACS Catal.* 2020, 10, 13946–13956, DOI: 10.1021/acscatal.0c03841. B, Molecular Insights into the Regioselectivity of the Fe(II)/2-Ketoglutarate-Dependent Dioxygenase-Catalyzed C–H Hydroxylation of Amino Acids, Source: *ACS Catal.* 2022, 12, 11586–11596, DOI:10.1021/acscatal.2c03106.

In our lab, a *cis*-P4H from *Sinorhizobium meliloti* (named SmP4H) has been used for producing *cis*-5-HPA with 93% conversion rate at 120 mM L-PA, while regioisomer (*cis*-3-HPA) was generated in almost equal proportions (*cis*-5/*cis*-3 about 1:1), along with the formation of numerous by-products leading to serious waste of substrate and the complicated *cis*-5-HPA purification procedure[14]. Herein, we also cloned another L-proline *cis*-4-hydroxylase from *Kordia jejudonensis* (named KjPH) and *cis*-P4H from *Mesorhizobium loti* (named MIP4H). *In vitro* hydroxylation of KjPH and MIP4H on L-PA showed the *cis*-5/*cis*-3 was 20:1 and 1:7, respectively. Together with sequence, structure and catalytic chamber comparisons of KjPH, SmP4H and MIP4H, the non-

conservative residues of K_jPH affecting regioselectivity were identified. To manufacture *cis*-5-HPA exclusively, site-directed mutagenesis was carried out on K_jPH to change the steric hindrance near C3 of L-PA. Subsequently, the molecular dynamics (MD) simulations were performed to confirm the mechanism of variant regioselectivity increasing.

2. Materials and methods

2.1 Strains, plasmids and reagents

E. coli BL21(DE3) strains and pET-28a (+) plasmids were stored in our laboratory. The *cis*-5-HPA standard was bought from Aladdin Biochemical Technology Co. (Shanghai, China). All other chemical reagents were purchased from Innochem Technology Co. (Beijing, China). Tool enzymes for molecular performances were ordered from TaKaRa Biotechnology Co. (Dalian, China) and Vazyme (Nanjing, China).

2.2 Structural bioinformatics analysis of the KjPH and homology modeling

We partially participated in this work and the complete work was mainly done by master graduate Fujun Huang.

The 3D structure of MIP4H (4P7X) was downloaded from the Protein Data Bank [2015]. Sequence alignments were conducted by ESPript (<https://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>). The homology model of KjPH and SmP4H was created through homology modeling based on 4P7W template using SWISS-MODEL [2018]. The model of KjPH was superimposed to the structures of SmP4H and MIP4H using Discovery Studio (DS) 2019 software, and then the RMSD values among these three structures were calculated.

2.3 Site-directed mutagenesis, expression and purification of recombinant KjPH

This work on site-directed mutagenesis, plasmid construction and gene sequencing were mainly done by GeneScript Technology and Tsingke Technology. We mainly complete the expression and purification of recombinant KjPH.

The genes of K_jPH (Accession number: WP_046758372) was synthesized by GeneScript Technology (Nanjing, China) and cloned into pET-28a (+) vector to construct the expression plasmid pET-28a-K_jPH. The *E. coli* BL21(DE3) strain harboring pET-28a-SmP4H is stored in our laboratory. *E. coli* BL21(DE3) introduced with pET-28a-K_jPH or pET-28a-SmP4H were cultured in Luria–Bertani (LB) media containing 30 mg/L kanamycin at 200 rpm and 37 °C to OD₆₀₀ 0.6 and induced by addition of 0.1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After overnight incubation at 20°C, wet cells were harvested by centrifugation in 8000×g at 4°C and disrupted by sonication (resuspended in 50 mM pH6.5 Sodium citrate buffer) or stored at -20°C for further use.

Primer pairs listed in **Table 1** were designed using the primerX website (<https://www.bioinformatics.org/primerx/>) and synthesized by GeneScript Technology (Nanjing, China). Polymerase chain reaction (PCR) experiments were performed according to the instructions of the kit (Vazyme). After amplification, the plasmid template was digested with *DpnI*, and the digested PCR mixture was introduced into *E. coli* BL21(DE3). The cells were spread onto LB plates pre-added 30 mg/L kanamycin. Single colonies from *E. coli* BL21(DE3) strains transformed pET-28a-K_jPH (mutants) were inoculated into 4 mL LB culture medium containing 30 mg/L kanamycin and grown at 37 °C overnight. The culture medium was sent to Tsingke Technology (Nanjing, China) for DNA sequencing the next day and the identified mutations were kept for later use. The expression, collection and sonication of mutants were performed as described in the previous section.

Table 1 Primer pairs used in this study

Name	Sequence
Y35F-F	GATAGCTATAGCGAGTTTGCCTCCGGCATTGG
Y35F-R	CCAAATGCCGGAGGCCAACTCGCTATAGCTATC
Y35A-F	GATAGCTATAGCGAGGCA GCCTCCGGCATTGG
Y35A-R	CCAAATGCCGGAGGCTGCTCGCTATAGCTATC
Y35W-F	GATAGCTATAGCGAGTGGGCCTCCGGCATTGG

Y35W-R	CCAAATGCCGGAGGC <u>CCA</u> CTCGCTATAGCTATC
S57L-F	GAAGGACGACCATTG <u>CTG</u> ATCGAACATGATACG
S57L-R	CGTATCATGTTTCGAT <u>CAG</u> CAAATGGTCGTCCTTC
S57A-F	CAGAAGGACGACCATTG <u>GCA</u> ATCGAACATGATACGTAC
S57A-R	GTACGTATCATGTTTCGAT <u>TGC</u> CAAATGGTCGTCCTTCTG
S57V-F	CAGAAGGACGACCATTG <u>GTT</u> ATCGAACATGATACGTAC
S57V-R	GTACGTATCATGTTTCGAT <u>AAC</u> CAAATGGTCGTCCTTCTG
F95A-F	CATATCAAACTGTACGCCTG <u>GCA</u> ATGTGCATCAACGGTTTGATC
F95A-R	GATCAAACCGTTGATGCACAT <u>TGC</u> CAGGCGTACAGTTTTGATATG
F95G-F	CATATCAAACTGTACGCCTG <u>GGT</u> ATGTGCATCAACGGTTTGATC
F95G-R	GATCAAACCGTTGATGCACAT <u>ACC</u> CAGGCGTACAGTTTTGATATG
F95V-F	CATATCAAACTGTACGCCTG <u>GTT</u> ATGTGCATCAACGGTTTGATC
F95V-R	GATCAAACCGTTGATGCACAT <u>AAC</u> CAGGCGTACAGTTTTGATATG
F95W-F	CAAACTGTACGCCTG <u>TGG</u> ATGTGCATCAACGGTTTG
F95W-R	CAAACCGTTGATGCACAT <u>CCA</u> CAGGCGTACAGTTTTG
F95Y-F	CAAACTGTACGCCTG <u>TAT</u> ATGTGCATCAACGGTTTG
F95Y-R	CAAACCGTTGATGCACAT <u>ATA</u> CAGGCGTACAGTTTTG
C97A-F	CTGTACGCCTGTTTCATG <u>GCA</u> ATCAACGGTTTGATCATC
C97A-R	GATGATCAAACCGTTGAT <u>TGC</u> CATGAACAGGCGTACAG
C97G-F	CTGTACGCCTGTTTCATG <u>GGT</u> ATCAACGGTTTGATCATC
C97G-R	GATGATCAAACCGTTGAT <u>ACC</u> CATGAACAGGCGTACAG
C97V-F	CTGTACGCCTGTTTCATG <u>GTT</u> ATCAACGGTTTGATCATC
C97V-R	GATGATCAAACCGTTGAT <u>AAC</u> CATGAACAGGCGTACAG
C97F-F	CTGTACGCCTGTTTCATG <u>TTT</u> ATCAACGGTTTGATCATC
C97F-R	GATGATCAAACCGTTGAT <u>AAA</u> CATGAACAGGCGTACAG
F95Y/C97V-F	GTACGCCTGTATATG <u>GTT</u> ATCAACGGTTTGATC
F95Y/C97V-R	GATCAAACCGTTGAT <u>AAC</u> CATATACAGGCGTAC

The harvested cells were resuspended in 20 mM Tris-HCl buffer (pH 7.0) containing

10 mM imidazole and 500 mM NaCl, and sonicated on ice bath for 25 min (sonication per 1s with 1 s interval) using an ultrasonic disrupter. After centrifugation at $8,000\times g$ for 40 min at 4 °C, the cell-free supernatant was collected. The proteins (tagged with 6 \times His) of wild-type KjPH and its mutants were purified with gradient concentration imidazole (10-500 mM) by His-Trap HP affinity chromatography (GeneScript). Imidazole and NaCl were eluted via PES10 ultrafiltration spin columns (Beyotime Biotech, Shanghai China) and the buffer was exchanged with 50 mM sodium citrate buffer (pH6.5). Bradford method was used for protein concentration determination (Sangon Biotech, Shanghai, China).

2.4 *In vitro* hydroxylation of L-PA

The *in vitro* hydroxylation experiments were performed in a reaction mixture which contains pH6.5 Sodium citrate buffer (50 mM), L-pipecolic acid (20 mM), α -ketoglutarate (20 mM), ascorbic acid (10 mM), $\text{FeSO}_4\cdot 7\text{H}_2\text{O}$ (1 mM) and suitable enzyme preparation, in a total volume of 10 mL. The reaction was started by the addition of enzyme solution. The amounts of hydroxylated products, *cis*-5-HPA and *cis*-3-HPA were detected by HPLC after the precolumn derivatization with Fluorene methoxycarbonyl chloride (Fmoc-Cl). The amount of enzyme that catalyzed the generation of 1 μmol *cis*-5-HPA plus *cis*-3-HPA per min was defined as one unit of the activity (U). Parallel experiments were performed for all.

2.5 Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed by Shiyanjia Lab using GROMACS 2021.5 package. SWISS-MODEL was used to conduct homology modeling for KjPH wild type (WT) of enzyme based on 4P7W template. The Co^{2+} was replaced by Fe^{2+} , and the co-substrate α -KG was added in the same position after aligned with template. Enzyme was parameterized by Amberff14sb force field. Fe^{2+} was parameterized by the HFE parameter set for TIP3P water model developed by Li/Merz. L-PA and α -KG were geometrically optimized by ORCA 5.0.3[2022] under

SMD (water) implicit solvation model with density functional theory B3LYP-D3(BJ)/def-TZVP level. Ambertools21[2021] and ACPYPE were employed to construct the general AMBER force field 2(GAFF2) parameters. Multiwfn was utilized to fit the restrained electrostatic potential 2(ESP) charge. The likely binding positions of L-PA on WT were explored by AutoDock Vina 1.2.3 and local docking was performed by setting grid to cover the entire active center. Phe95 was then mutated to Tyr to obtain the complex of the mutant (F95Y) and L-PA. Top 1 docking model was selected for further MD simulation and results were visualized by PyMOL (<http://pymol.sourceforge.net/>).

Cubic box was established by extending at least 1.2 nm outward along enzyme-PA complex ($9 \times 9 \times 9 \text{ nm}^3$). The system was solvated in TIP3P water and 0.15 M NaCl (72 Na^+ and 66 Cl^-) were added to keep electrically neutral. Energy minimization was performed using the steepest descent algorithm with a force tolerance of $500 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. In all the three directions, periodic boundary conditions were imposed. Then system was relaxed for 1 ns under NPT (constant pressure and temperature) MD simulations and position restraints with a constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ in three directions were performed on heavy atoms of complex. After completing the above steps, 100 ns NPT MD simulation was conducted on complex. Pressure was maintained at 1 bar by the Parrinello-Rahman barostat in an isotropic manner and temperature was kept at 298 K by the V-rescale thermostat. The LINCS algorithm was employed to constrain bond lengths of hydrogen atoms. Lennard-Jones interactions were calculated within a cutoff of 1.2 nm, and electrostatic interactions beyond 1.2 nm were treated with particle-mesh Ewald (PME) method with a grid spacing of 0.16 nm. Finally, the obtained trajectory was analyzed to get the snapshot of enzyme-PA complex at a specific time, and the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), the distance between C3 and C5 of substrate and Fe^{2+} was calculated. Additionally, the binding free energy (ΔG_{bind}) of enzyme-PA was analyzed according to MM/PBSA method by gmx_MMPBSA tool.

2.6 Derivatization and Detection of L-PA, *cis*-5-HPA and *cis*-3-HPA by HPLC

Both substrate(L-PA) and products (*cis*-5-HPA and *cis*-3-HPA) were analyzed by HPLC after precolumn derivatization. 0.25 mL of the reaction solution was mixed with isopycnic 50 mM sodium borate buffer (pH 9.4). The derivatization was conducted with 0.5 mL of 15 mM Fmoc-Cl (dissolved in acetone) and vigorously reacted for 1 min. Finally, the mixture was extracted twice with 1 mL of cyclohexane and the aqueous phase was retained for HPLC analysis.

The LC-20AD series HPLC system (Shimadzu, Shanghai, China), equipped with Supersil ODS2 C18 reverse-phase column (4.6×250 mm, 5 μm; Elitehplc Analytical Instrument Co., Dalian, China) was used for products determination. The flow rate was 1 mL/min, with the temperature of 40 °C and 254 nm for detection wave length. Gradient elution program was as follow: Buffer A (50 mM sodium acetate buffer of pH4.0), Buffer B (pure acetonitrile for HPLC); A:0-6min:70%, 6-9min:70%→50%, 9-12 min:50%→20%, 12-15 min:20%→0%, 15-18 min:0%, 18-20 min:0%→50%, 20-22 min:50% →70%, 22-25 min:70%. The method was described in our previous paper [2020].

3. Results and Discussion

3.1 Identification of regioselectivity difference for KjPH and Smp4H

Based on the strains preserved in our laboratory, the proline hydroxylase named Smp4H was used for generating *cis*-5-HPA, with the *cis*-5-HPA/*cis*-3-HPA ratio about 1:1 and tedious purification procedure of *cis*-5-HPA. We obtained another proline hydroxylase from *Kordia jejudonensis* (KjPH). *In vitro* catalytic reactions were conducted to characterize the difference between Smp4H and KjPH on hydroxylation of L-PA and products were analyzed by HPLC (Fig.3). Based on HPLC assay, we observed the formation of the targeted *cis*-5-HPA. The *cis*-5/*cis*-3 ratio reached 20:1 in the reaction mixture of KjPH (Fig.3c), much higher than that of Smp4H (about 1:1) (Fig.3b). Besides, Koketsu reported that the *cis*-5/*cis*-3 ratio in the reaction by *cis*-P4H from *Mesorhizobium loti* (MIP4H) was approximately 1:7[13]. Great differences in *cis*-5/*cis*-3 ratio revealed that specific residues in catalytic chamber regions are variable.

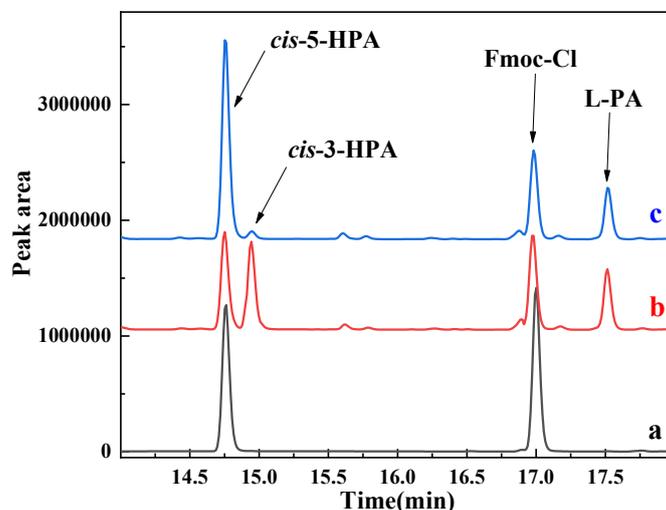


Fig. 3. HPLC chromatograms comparison of hydroxylated products between KjPH and Smp4H. Reactions were carried out at 25 °C for 3 h with vigorous shaking (200rpm) and the mixture contained 50 mM Sodium citrate buffer (pH6.5), 20 mM L-PA, 20 mM α -KG, 10 mM ascorbic acid, 1 mM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ and suitable enzyme supernatants, in a total volume of 10 mL. The reaction mixture including hydroxylated products were analyzed with HPLC after precolumn derivatization. The retention time of *cis*-5-HPA, *cis*-3-HPA, Fmoc-Cl and L-PA are 14.75, 14.95, 17.0 and 17.5 min, respectively. (a) standard sample of *cis*-5-HPA, (b) standard reaction mixture with Smp4H, (c) standard reaction mixture with KjPH.

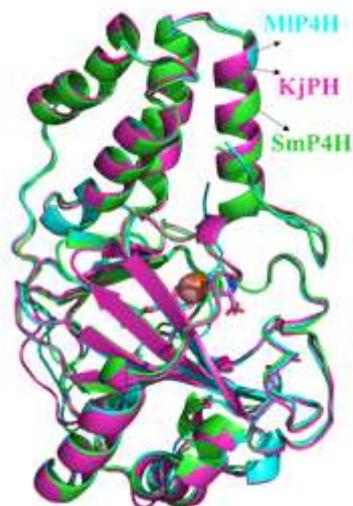


Fig 5. Structure alignment of KjPH, SmP4H and MIP4H. Sequence alignments were performed using ESPript (<https://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>).

The selected residues are listed in **Table 2**. Although C97* (in KjPH) and V97* (in SmP4H) were beyond the 4 Å range, mutations on V97* tremendously changed the *cis-5/cis-3* ratio, and V97*/C97* were also selected.

Table 2 Residues within 4Å in MIP4H, SmP4H and KjPH

Enzyme	entries	residues
MIP4H	L-PA/Fe ²⁺ /α-KG	Y32, F35, W40, V57, R93, I95, I97, I103, H106, D108, R118, H120, H154, A156, R164
	L-PA	Y32, F35, W40, V57, R93, I103, H106, D108, R118
	Fe ²⁺	H106, D108, H154
SmP4H	L-PA/Fe ²⁺ /α-KG	Y32, F35, W40, V57, R93, V95, V97*, I103, H106, D108, R118, H120, H154, A156, R164
	L-PA	Y32, F35, W40, V57, R93, I103, H106, D108, R118
	Fe ²⁺	H106, D108, H154
KjPH	L-PA/Fe ²⁺ /α-KG	Y32, Y35, W40, S57, R93, F95, C97*, I102, H105, D107, R117, H119, I147, H153, A155, R163, N165, V167

L-PA	Y32, Y35, W40, S57, R93, F95, I102, H105, D107, R117
Fe ²⁺	H105, D108, H153, R117
α -KG	W40, F95, C97*, I102, H105, D107, R117, H119, I147, H153, A155, R163, N165, V167

In addition, we noticed that residues I102/H105/D107/R117/H119/H153/A155/R163 in K_jPH are shifted relative to residues I103/H106/D108/R118/H120/H154/A156/R164 in Smp4H and MIP4H, while residues in sites 32/35/40/57/93/95/97 for three enzymes are all aligned. The reason of the mismatch of these residues may be the absence of a residue in the region that links two β -sheets (β 1 and β 2) in K_jPH (red dotted line box noticed in **Fig.6a** and **Fig.6b**). The absence of one residue in this region has little effect on the secondary structure. When inserted anyone residue in site 98 (K_jPH'), the mismatched residues in catalytic chamber are aligned and the sequence similarity between K_jPH' and Smp4H/MIP4H reached to 46% and 47%, respectively. Among the residues in catalytic chamber, strictly conserved residues (highlighted by a red background, **Fig.6c**) and I147/N165/V167 (located beneath α -KG) are not considered as mutation sites, so the variable key residues in K_jPH were mainly determined as Y35/S57/F95/C97 (pink sticks in **Fig.6d**). Y35 and F95 are located above and below C5 and C3 carbons of the substrate, respectively, while S57 is close to the carboxyl moiety of L-PA and C97 locates near α -KG (**Fig.6d**). strictly conserved residues and I147/N165/V167 are displayed as green lines in Fig.3d. Furthermore, the variation of catalytic chambers between Smp4H and MIP4H locates in 95/97 sites, while the regioselectivity showed differently in spite of 91% sequence similarity. Site 95 may push substrate C3 away from the Fe²⁺ core through steric hindrance. Hence, mutations of K_jPH focus on F95.

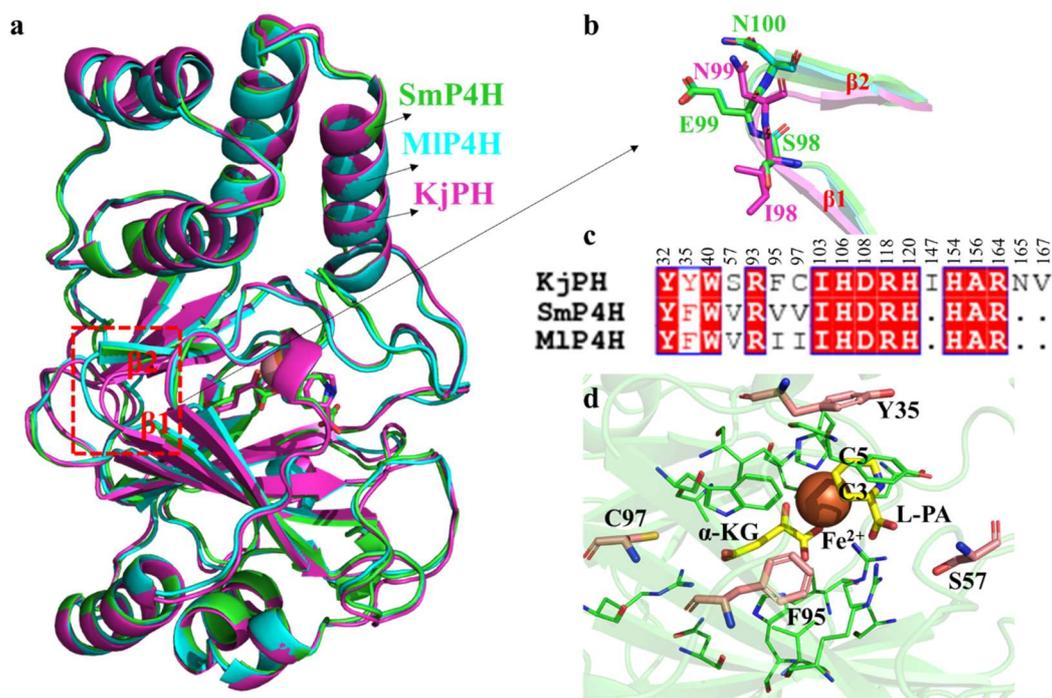


Fig. 6. Sequence and structure alignments between KjPH, SmP4H and MIP4H. (a) Structure alignment of KjPH, SmP4H and MIP4H. (b) Partial enlargement of the red dotted line box region in (a). I98/N99 (purple) are residues of KjPH and S98/E99/N100 (green) are residues of SmP4H or MIP4H. β1 and β2 are two β-sheets constituting the catalytic chamber. (c) Multiple sequence alignment of the residues in catalytic chambers of KjPH, SmP4H and MIP4H. (d) The variable key residues Y35/S57/F95/C97 (pink sticks) in KjPH. The rest of residues in catalytic chamber are exhibited as green lines. L-PA/α-KG are displayed as yellow sticks and the Fe²⁺ is showed as orange ball. Sequence alignments were performed using ESPript (<https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>).

3.3 Regioselectivity improvement via site-directed mutagenesis

To improve regioselectivity towards *cis*-5-HPA, site-directed mutagenesis was proceeded on the aforementioned sites of the KjPH. For Y35 site, the revertant Y35F (relative to SmP4H) displayed prominent descent in enzyme activity and regioselectivity with 653 mg/L of *cis*-5-HPA and *cis*-5/*cis*-3 ratio of 2:1 (1726 mg/L of *cis*-5-HPA produced by the wild-type with the *cis*-5/*cis*-3 ratio of 20:1) despite the lacking of only one hydroxyl group, so the declining of the other variant (Y35A and Y35W, with smaller and larger side chain respectively) in activity and regioselectivity was foreseeable (**Table. 3**).

Table 3. The production of HPAs by K_jPH and its mutants

Enzyme	<i>cis</i> -5-HPA (mg/L)	<i>cis</i> -3-HPA (mg/L)	<i>cis</i> -5/ <i>cis</i> -3
WT	1726±48.7	86±16.2	20
Y35F	653±26.5	323±20.6	2.0
Y35A	44±6.3	31±4.1	1.4
Y35W	744±33.6	262±19.7	2.8
S57L	650±31.3	398±24.9	1.6
S57A	430±13.7	111±12.6	3.9
S57V	618±24.6	219±17.7	2.8
F95A	182±21.8	47±11.2	3.9
F95G	156±19.7	39±14.8	4.0
F95V	390±13.2	73±14.6	5.3
F95Y	989±36.8	18±6.4	55.0
F95W	128±6.1	17±6.5	7.5
C97A	1823±54.7	69±15.8	26.4
C97G	109±16.8	18±12.2	6.1
C97F	755±26.3	72±11.8	10.5
C97V	2121±55.6	65±15.1	32.6
F95Y/C97V	107±9.1	17±4.3	6.3

For S57 site, the S57V mutant (revertant relative to SmP4H) yielded 618 mg/L of *cis*-5-HPA and the *cis*-5/*cis*-3 ratio significantly declined to 2.8:1 (**Table 3**). The detection result of S57L and S57A variant was similar to that of S57V, with 650mg/L of *cis*-5-HPA (*cis*-5/*cis*-3 = 1.6:1) for the former and 430mg/L of *cis*-5-HPA (*cis*-5/*cis*-3 = 3.9:1) for the latter (**Table 3**).

For revertant F95V (relative to SmP4H), the enzyme activity and regioselectivity descended significantly, yielding 390 mg/L of *cis*-5-HPA and 73 mg/L of *cis*-3-HPA respectively (*cis*-5/*cis*-3 ratio of around 5.3:1) (**Table 3**), which is greater than the *cis*-5/*cis*-3 ratio of approximate 1:1 for SmP4H[14]. Enlarging (F95W) or shrinking (F95A and F95G) the side chain of site 95 both decreased activity and regioselectivity. Variants

F95A (182 mg/L of *cis*-5-HPA, *cis*-5/*cis*-3 = 3.9:1) and F95G (156 mg/L of *cis*-5-HPA, *cis*-5/*cis*-3 = 4.0:1) shared almost parallel hydroxylation activity and regioselectivity (**Table 3**). It may be that the size difference of the side chain between Ala and Gly is not significant. The hydroxylation activity and regioselectivity of F95W mutant decreased tremendously, with only 128mg/L of *cis*-5-HPA generated and the *cis*-5/*cis*-3 ratio of 7.5:1, which was similar to the result of V95W variant for Smp4H. While the variant F95Y showed the highest regioselectivity (*cis*-5/*cis*-3=55:1) with 989 mg/L of *cis*-5-HPA formation in spite of retaining about only half the activity of the wild-type (**Table 3**).

The C97V and C97A variant displayed better productivity (2121 mg/L and 1823 mg/L of *cis*-5-HPA, respectively) and slightly worse regioselectivity (*cis*-5/*cis*-3=32.3 and 26.2, respectively) compared with that of the F95Y variant, but remarkable decreasing in activity and regioselectivity was observed for C97G (109 mg/L of *cis*-5-HPA, *cis*-5/*cis*-3=6.1) and C97F (755 mg/L of *cis*-5-HPA, *cis*-5/*cis*-3=10.5), which suggested that the amino acid residue in site 97 may alter activity and regioselectivity by affecting α -KG binding status. However, it was very strange that double mutant F95Y/C97V showed prominent declining in enzyme activity (107 mg/L of *cis*-5-HPA) and regioselectivity (*cis*-5/*cis*-3=6.3) (**Table 3**).

3.4 Molecular dynamics simulations

The MD simulations for WT K_jPH and variant F95Y were conducted over a 100-ns time scale respectively. In the presence of Fe²⁺/L-PA/ α -KG, the RMSD of the α -carbon atoms of the two enzymes reached equilibrium after 90 ns, and therefore, only the MD trajectories of the last 10 ns were used for further analysis (**Fig. 7A**). The analysis of RMSF values did not show remarkable changes at positions 90 to 100 between K_jPH and F95Y mutant (**Fig. 7B**). This was reasonable since F95Y has only one extra hydroxyl group in side chain than K_jPH, which has little effect on the backbone. However, a slight decrease in RMSF values was observed for the F95Y mutant over positions 100-110 and 190-200, indicating that the neighboring residues were more

flexible for K_jPH at the same position. The binding free energy between enzyme and substrate was calculated by MM/PBSA. The ΔG_{bind} of WT-PA and F95Y-PA were -17.57 and -37.92 kcal/mol, respectively. The primary cause is that the substrate carboxyl group interacts with different residues.

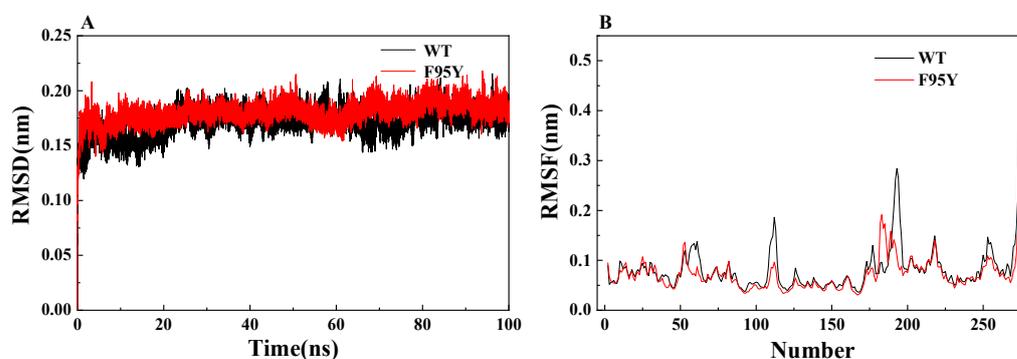


Fig 7. The RMSD (A) and RMSF (B) values of K_jPH and F95Y mutant during 100-ns MD simulations. The trajectories of the last 10 ns were used for data analysis.

Salt bridges (hydrogen bonds and electrostatic interactions) were generated between carboxyl of L-PA and R117/S57 in WT (**Fig. 8a**). While in F95Y, interactions of the phenol hydroxyl (Y95) and the guanidine (R93) with the substrate carboxyl were newly formed, despite the interactions from carboxyl to S57 lost (**Fig. 8c**). The C5 and C3 carbons of L-PA locate at distances of 4.7 and 5.3 Å, respectively, from the Fe²⁺ in WT, with the difference value of 0.6 Å (**Fig.8b**), while the C5 and C3 carbons of L-PA position at distances of 5.2 and 6.0 Å, respectively, from the Fe²⁺ in F95Y, with the difference value of 0.8 Å (**Fig.8d**). F95Y[(Fe²⁺-C3) - (Fe²⁺-C5)] is greater than WT[(Fe²⁺-C3) - (Fe²⁺-C5)], which explains the increasing of hydroxylation regioselectivity in F95Y variant. Based on the results and analysis, it is suggested that F95Y mutant pulls the whole substrate away from the Fe²⁺ core by the interaction formed between the substrate carboxyl and phenol hydroxyl (Y95) as well as guanidine (R93), leading to C3 carbon being too farther away from the metal core, and minimum C3 hydroxylated isomer was produced.

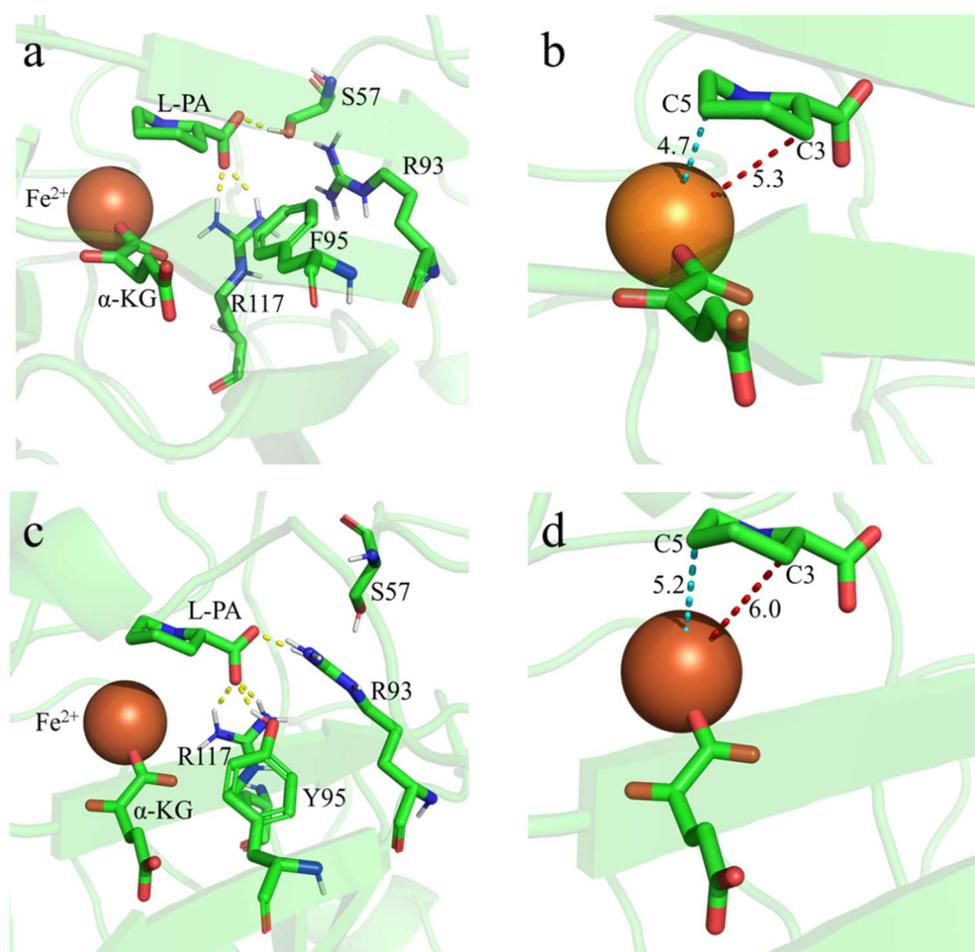


Fig 8. Interactions between substrate carboxyl and related residues and the distances between C5 and C3 carbons of L-PA and the Fe²⁺. (a) Interactions between substrate carboxyl and S57/R117 in WT. (b) The C5 and C3 carbons of L-PA locate at distances of 4.7 and 5.3 Å, respectively, from the Fe²⁺ in WT. (c) Interactions between substrate carboxyl and Y95/R93/R117 in F95Y. (d) The C5 and C3 carbons of L-PA position at distances of 5.2 and 6.0 Å, respectively, from the Fe²⁺ in F95Y.

4. Conclusion

High regioselectivity is one of the most attractive properties related to amino acid hydroxylases engineering perspective, however there exists challenge targeting key sites for site-directed mutagenesis to enhance regioselectivity performance of amino acid hydroxylases. Sequence and structural bioinformatics are beneficial in excavating structure–function relationships and offer guideline to protein rational design. In this work, we compared the differences in the sequence and 3D structure of the *cis*-P4H K_jPH from *Kordia jejudonensis* with the *cis*-P4H Smp4H from *Sinorhizobium meliloti* and the *cis*-P4H MIP4H from *Mesorhizobium loti* to confirm the structural determinants in the *cis*-P4H K_jPH that might be significant to increase the regioselectivity of this enzyme. Multiple sequence alignment suggested that the residues in the catalytic chamber were conservative among three *cis*-P4Hs to a large extent, in consistent with the functional significance of the active region. Oppositely, alterable residues in the catalytic chamber are generally relative to the functional multiformity of the homologous enzymes. Thus, site-directed mutagenesis was performed on the four variable residues (Y35, S57, F95, C97) of K_jPH to improve C5 hydroxylation regioselectivity. The variant F95Y demonstrated the highest regioselectivity (*cis*-5/*cis*-3=55:1) in spite of retaining about only half the activity of the wild-type. Then MD analysis supplied perspective into the enhanced regioselectivity performance of the mutant F95Y. Newly formed interactions between the phenol hydroxyl (Y95)/guanidine (R93) and the carboxyl of L-PA pulls the whole substrate away from the Fe²⁺ core, leading to C3 carbon being too farther away from the metal core, and minimum C3 hydroxylated product was generated. This work shows that utilizing the natural diversity of protein sequences and structures is a more effective tool to guide rational enzyme modification than direct evolution.

5. Reference

1. Abrams, D.J., P.A. Provencher, and E.J. Sorensen, *Recent applications of C-H functionalization in complex natural product synthesis*. Chemical Society Reviews, 2018. **47**(23): p. 8925-8967.
2. Luo, Y.R., *Handbook of Bond Dissociation Energies in Organic Compounds*, CRC Press LLC 2004.
3. Docherty, J.H., et al., *Transition-Metal-Catalyzed C-H Bond Activation for the Formation of C-C Bonds in Complex Molecules*. Chemical Reviews, 2023. **123**(12): p. 7692-7760.
4. Golden, D.L., S.-E. Suh, and S.S. Stahl, *Radical C(sp³)-H functionalization and cross-coupling reactions*. Nature Reviews Chemistry, 2022. **6**(6): p. 405-427.
5. Holmberg-Douglas, N. and D.A. Nicewicz, *Photoredox-Catalyzed C-H Functionalization Reactions*. Chemical Reviews, 2022. **122**(2): p. 1925-2016.
6. Lam, N.Y.S., K. Wu, and J.-Q. Yu, *Advancing the Logic of Chemical Synthesis: C-H Activation as Strategic and Tactical Disconnections for C-C Bond Construction*. Angewandte Chemie-International Edition, 2021. **60**(29): p. 15767-15790.
7. Huang, Y., et al., *H₂O₂-driven enzymatic oxyfunctionalization of tertiary C-H bonds*. Chemical Communications, 2023. **59**(60): p. 9219-9222.
8. Palone, A., et al., *C-H Bonds as Functional Groups: Simultaneous Generation of Multiple Stereocenters by Enantioselective Hydroxylation at Unactivated Tertiary C-H Bonds*. Journal of the American Chemical Society, 2023. **145**(29): p. 15742-15753.
9. Craven, E.J., et al., *Programmable late-stage C-H bond functionalization enabled by integration of enzymes with chemocatalysis*. Nature Catalysis, 2021. **4**(5): p. 385-394.
10. Wang, Z., et al., *Recent progress in enzymatic functionalization of carbon-hydrogen bonds for the green synthesis of chemicals*. Chinese Journal of Chemical Engineering, 2020. **28**(10): p. 2499-2506.
11. Peters, C. and R.M. Buller, *Industrial Application of 2-Oxoglutarate-Dependent Oxygenases*. Catalysts, 2019. **9**(3): p. 221-240.
12. Urlacher, V.B. and M. Girhard, *Cytochrome P450 monooxygenases: an update on perspectives for synthetic application*. Trends in Biotechnology, 2012. **30**(1): p. 26-36.
13. Koketsu, K., et al., *Refined Regio- and Stereoselective Hydroxylation of L-Pipecolic Acid by Protein Engineering of L-Proline cis-4-Hydroxylase Based on the X-ray Crystal Structure*. ACS Synthetic Biology, 2015. **4**(4): p. 383-392.
14. Lu, F., et al., *Enzymatic hydroxylation of L-pipecolic acid by L-prolinecis-4-hydroxylases and isomers separation*. Biotechnology Letters, 2020. **42**(12): p. 2607-2617.
15. 陆凡, et al., *合成 β -内酰胺酶抑制剂阿维巴坦的关键中间体的研究进展*. 中国抗生素杂志, 2019. **44**(10): p. 1125-1131.
16. Wang, F., et al., *Reshaping the Binding Pocket of Lysine Hydroxylase for Enhanced Activity*. ACS Catalysis, 2020. **10**(23): p. 13946-13956.
17. Wu, T., et al., *Reshaping Substrate-Binding Pocket of Leucine Dehydrogenase for Bidirectionally Accessing Structurally Diverse Substrates*. ACS Catalysis, 2023. **13**(1): p. 158-168.

6. Acknowledgements

课题源自我自己的看病经历。小时候体质比较弱，很容易感冒发烧甚至肺炎，然后就要去医院打点滴。在无论白天还是晚上都拥挤的输液室，看着吊瓶里的液体一滴滴地流入身体，感觉手臂有些凉。那时候并不知道输的什么液体，后来才知道大多是抗生素。长大了之后，身体强壮了，输液的次数也少了很多。但每每回想起来，总还有些感慨。

上了高中，生物课中关于抗生素知识点以及相应的实验“探究抗生素对细菌的选择作用”让我对这一小时候经常接触到的药物有了更深的理解。抗生素作为“20 世纪最伟大的发明之一”，在给病患带来福音的同时，也因为早期的滥用造成环境污染和加剧细菌耐药性，正成为目前全球最紧迫的公共卫生问题之一。由于抗生素药物发现和开发都缺乏创新，我们可能正在接近一个“后抗生素时代”。如果感染了耐药性的细菌而现有抗生素药物都不能起效时，我们是否会回到“最伟大的发明”之前的困境？

带着这一疑问，我来到了父亲的朋友，南京工业大学曹飞教授的实验室。他介绍了正在开发的一种 β -内酰胺酶抑制剂，阿维巴坦，与现有的头孢他啶组成复方抗生素来解决细菌耐药性的问题。其中，关键的步骤就是利用脯氨酸羟化酶来合成顺式-5-羟基哌啶甲酸中间体。在实验室，我又一次接触到：筛基因、挑克隆、表达蛋白、活性位点、定点突变、分子模拟、氨基酸残基、液相检测……这些事物都让我感到新奇，也让我对它产生了兴趣。

在征得父母的同意以及召集到合伙人后，我、殷岂墨和郁雅涵来到了南京工业大学生物与制药工程学院曹飞教授的实验室，开始了我们的酶改造探索之旅。一年来，PCR 和挑克隆的操作已经熟练，核酸胶和蛋白胶也跑了无数块，液相的

小瓶子也不知消耗了多少……

感谢曹飞教授对选题、理论授课、实验过程和论文书写的大力帮助。感谢武红丽副教授的理论授课和实验指导。感谢课题组硕士生黄富军、王君宇和时蕴林在蛋白结构、分子对接和定点突变上的大力帮助。感谢指南针团队在酶的铁中心与底物 3/5 位点距离计算上的协助。这部分蛋白结构工作我们已经尽量地去学习，但真正的理解还差得很远。

最后，感谢父母的支持。

7. Supervisor and Students Information

曹飞，南京工业大学教授。江苏省“青蓝工程”优秀青年骨干教师、江苏省“六大人才高峰”高层次人才、江苏省“双创计划”科技副总。主持国家科技部重点研发计划课题、国家“973”项目子课题、国家自然科学基金项目、江苏省自然科学基金项目、南京市科委面上项目，并完成多项企业横向项目。发表 SCI、EI 论文 30 余篇，获得授权美国专利 1 项、中国专利 12 项。

研究方向：

- (1) 基于结构生物信息学进行酶结构与功能分析和理性设计改造
- (2) 低劣生物物质的高值化转化与利用

学生简介

张鹭飞，南京外国语学校 2022 级学生。主要完成 KjPH, SmP4H 和 MIP4H 的生物信息学分析、KjPH 表达和酶活测定，F95, C97 位点的突变工作；

殷岂墨，南京外国语学校 2022 级学生。主要完成 MIP4H 表达和 SmP4H 表达、S57 位点的突变工作。

郁雅涵，南京外国语学校 2022 级学生。主要完成液相检测、KjPH 酶 Y35 位点的突变工作。