

Supplemental information

Formaldehyde Dehydrogenase SzFaldDH: An Indispensable
Bridge for Relaying CO₂ Bioactivation and Conversion

Boxia Guo^{a,b}, *Xiuling Ji*^a, *Yaju Xue*^{a,b}, *Yuhong Huang*^{a,*}

^aBeijing Key Laboratory of Ionic Liquids Clean Process, CAS Key Laboratory of Green Process and Engineering, State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China

^bSchool of Chemical Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Experimental

1. Materials

All gene coding AqFaldDH, SzFaldDH, PpFaldDH, ChFaldDH and FDHPa were codon-optimized and synthesized by Shanghai Generay Biotechnology company. PaFDH was prepared following the study of the literature of Xue et al¹. Commercial enzymes used in this study alcohol dehydrogenase from *S. cerevisiae* (YADH) and formaldehyde dehydrogenase from *Pseudomonas aeruginosa* (PFaldDH) were purchased in Sigma-Aldrich (Shanghai, China).

2. PPR excavation and bioinformatics analysis

Based on the existing primary sequence features of formaldehyde dehydrogenase (FaldDH), the PPR was used to identify the peptide length (peptide length), number (limit) and the threshold of peptide number in each protein sequence (cut off), to establish a formaldehyde dehydrogenase FaldDH-specific protein sequence database and a short peptide module sequence library; and further using the short peptide module sequence In the process of mining FaldDH using PPR, 3020 FaldDH sequences were downloaded, and after PPR analysis, nine amino acid sequence groups were obtained from FaldDH, and the amino acid sequences in each group were scored, and four sequences with high scores were selected in the first group.

Protein sequences with the highest score in each PPR and characterized sequence groups of FaldDHs aligned by ClustalX, which were accompanied by sequence reported in other studied and commercial enzymes. After aligning, phylogenetic

analysis was then processed to build phylogenetic trees, using Maximum Likelihood method in MEGA11. Multiple Sequence Alignment was analyzed by ClustalX2. then the multiple alignment graph was constructed from the aligned sequence in ESPrint3.0² (<https://esprint.ibcp.fr>). Sequence similarity was completed by Blast analysis in NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Tertiary structure was predicted in an online tool I-TASSER³ (<https://zhanggroup.org/I-TASSER>) by a multiple threading approach and in SWISS-MODEL (<https://swissmodel.expasy.org>) by a homology modeling approach. Viewing editing, and analysis of protein structures were performed by YASARA (YASARA Structure, YASARA Biosciences, Vienna, Austria), ChimeraX⁴, and Pymol.

3. Molecular docking and simulation

Molecular docking was carried out by YASARA. Protein structure files were prepared by defining simulation cell according to the active site and environment, which were saved as receptor files. Then ligand files were prepared containing NAD(H) and substrate (formate, formaldehyde, methanol) by Chemical Draw. MD simulation was processed by forcefield AMBER14 in YASARA using a NIVIDA GPU, which was filled with water guided by a cell neutralization and pKa prediction. The system temperature and pH were set as the actual reaction conditions. The simulation box was periodic, energy minimization was performed after fixing the simulated molecule, and finally 80 ns of simulation was performed using the simulation program. After

simulation, parameters such as RMSD, RMSF, distance and angle were analyzed by program: MD analyze.

4. QM calculation of FaldDH

All the QM calculation was carried with program Guassian 16 using hybrid density functional theory. The initial model of the QM calculation was obtained from the energy minimization results of the MD simulation, where the QM region consists of 156 atoms, including residues Ser57, Cys57, His62, His71, Asp179, cofactor NADH, and the substrate sodium formate. B3LYP functional and 6-31G* basis set were used for all the QM atoms calculation containing geometry optimizations and transition state optimizations. Then the vibrational frequencies and IRC calculations was carried out to confirm the transition state at the same level of theory as for the geometry optimizations.

5. Cloning, expression, and purification of discovered FaldDHs

The AqFaldDH (WP_131357959), SzFaldDH (WP_043728606.1), PpFaldDH (WP_0587771131.1), ChFaldDH (WP_0587771131.1) gene were cloned into the pETDuet plasmid vector, and the restriction site was BamHI-SacI. The codon-optimized gene with C-terminal Histag (addition number AOAOVOPWGO) was inserted into the NcoI/SacI restriction site of the pETDuet plasmid to construct the plasmid petDuetor-AqFaldDH. The expression vector was transformed into E. coli BL21 (DE3). Finally, the AqFaldDH gene was replaced with other codon-optimized FaldDHs genes to construct six different recombinant strains.

Purification of the enzyme was obtained by fast protein liquid chromatography (FPLC) on AKTATM pure 25 using HisTrap HP coarse affinity column (5 mL, 17524802, Cytiva) and HiTrap desalting column (5 mL, 17140801, Cytiva). Firstly, high-pressure homogenizer was used to break bacteria and release proteins, the supernatant was taken after centrifugation at 12000 rpm for 30 minutes. The supernatant was purified in Binding buffer (pH 7.4) and Elution buffer (pH 7.4), then eluted using a flowing buffer (50 mM sodium phosphate, pH 7.4). The purified and desalted enzymes were further concentrated and collected in flow buffer at -80 °C in 1.5 mL centrifuge tubes.

6. Determination of protein concentration and enzyme activity

The protein was obtained by bicinchoninic acid (BCA) method using the Pierce® BCA Protein Assay Kit (Thermo Scientific, Rockford, USA) according to the manufacturer's instructions and using BSA as standard. The BSA standards were prepared following a concentration gradient from 0-2000 µg/mL. Then the mixture of 25 µL BSA standards or untested proteins with 200 µL BCA in 96 microplates were incubated in 37 °C oven for 30 minutes. The absorbance (595 nm) of the reaction mixture was measured with a spectrophotometer, and the Standard curve of protein concentration was obtained by BSA standards. The enzyme concentration was obtained by reading the absorbance at 595 nm and calculating by standard curve.

The activity of FaldDHs was determined by measuring the consumption of NADH at 340 nm with a BioTek Cytation 5 imaging reader⁵. The reaction system of FaldDHs in the reduction direction contains 10 mM sodium formate, 1 mM NADH, 0.2 mg/mL

FaldDHs and phosphate buffer (100 mM, pH 7.0). NADH consumption was measured continuously by mixing at 25 °C for 10 minutes. The amount of NADH consumed in one minute (μM) was defined as one unit of enzyme activity (U), and the specific enzyme activity was defined as the enzyme activity per unit mass of enzyme (U/mg). Oxidative activity was carried by a similar system with FaldDH containing 10 mM formaldehyde, 1 mM NAD^+ , 0.2 mg/mL FaldDH and phosphate buffer. The oxidation activity was determined mainly by measuring the absorption peak at 340nm, and the activity was characterized by an increase in NADH.

7. Characterizations of FaldDHs

The effect of pH was determined by measuring activity at 25 °C in various buffers, with a pH range of 3.0–6.0 (citric acid buffer), 6.0–8.0 (phosphate buffer), and Tris-HCl buffer (8.0–9.0) 10-11 (NaHCO_3 -NaOH buffer). The effect of temperature was measured at phosphate buffer (100 mM, pH 7.0) at temperatures ranging from 25 °C to 70 °C. The heat resistance is assessed by the following parameters $T^{1/2}_{15\text{min}}$ and half-life, obtained by measuring the remaining activity of enzymes, which were incubated at 35, 40, 45, 50, 55, 60 °C for 15 minutes, respectively. Half-life was measured by the same methods incubating at 40 °C remaining for 1 to 6 h, which was withdrawn by one hour. The measurement method was the same as above. The temperature and time at which 50% of the activity was lost after incubation are $T^{1/2}_{15\text{min}}$ and half-life.

Acid/alkali tolerance was also obtained by residual activity after incubation at different pH (5.0, 6.0, 8.0, 9.0) values for a certain amount of time (1 to 6 hours). To minimize the pH influence of acid and alkali in the reaction process, 20 μL buffer for

different pH was mixed and incubated with enzymes for 1-6 hours, then residual activity was measured in PB (pH 7.0, 25 °C). Antioxidant activity was characterized by measuring residual activity after incubation with hydrogen peroxide. The residual enzyme activity was measured after incubation with 20 μ L of hydrogen peroxide at different concentrations for 15 minutes. Then 5, 10, 20 mM H_2O_2 was chosen to evaluate the antioxidant, which were incubated with enzymes for 1-6 hours measured by one hour. The relative and residual activities at each condition were expressed relative to the highest activity value (100%). All experiments were performed in duplicate under standard assay conditions. Then all the residue activities at different temperatures and time points was fitted by a Boltzmann sigmoidal function in Origin software.

In this study, the kinetics of FaldDHs was measured at variable concentrations of one of the ligands (NADH or HCOONa), while the concentration of the second ligand was kept at a saturated fixed level. The reaction system contains different concentrations of NADH (0.1 mM to 1 mM), a saturating (10mM) concentration of HCOONa, 0.2 mg/mL enzymes and 100 mM PB buffer. As for oxidation activity, the NADH was replaced by NAD^+ , and the NADH production was measured at 340 nm. Then the data collected was imported into GraphPad Prism9 to obtain the parameters K_m , k_{cat} , V_{max} and k_{cat}/K_m by matching with a Michaelis-Menten model.

8. CO₂ bioactivation and conversion

The system to produce formaldehyde from sodium formate consisted of 10 mM sodium formate, 5 mM NADH and 0.2 mg/mL FaldDH for a reaction time of 1 hour.

The system to produce formaldehyde from CO₂ was consisted of 0.1mg/mL FDHPa

and 0.2 mg/mL FaldDH, 5mM NADH for a reaction time for 2.5 hours and time point was collected every 30 minutes. Formaldehyde was determined by the Nash's color development method ⁶ using a 1:1 mixture of Nash's reagent containing 0.05 M acetic acid, 0.02 M acetylacetone and 2 M ammonium acetate with the sample incubated at 40 °C for 0.5 hour. The concentration of formaldehyde was determined by HPLC with a C18 column and 0.4 mL flowrate acetonitrile/water eluent (20:80, v/v) in a 412 nm detection wavelength⁷. The product selectivity was carried out by measuring the methanol curve of FaldDH reducing formate system by Shimadzu Production gas chromatograph with a BID detector.

The double enzyme cascade reaction with FaldDH and ADH was carried out with 100 mM PB buffer, 10 mM sodium formate, 5 mM NADH, 0.4 mg/mL FaldDH and 0.04 mg/mL ADH as the reaction system. The methanol content was measured after 6 hours of reaction. Seven concentration gradients from 1 mM to 50 mM sodium formate were used in the above reaction system to determine the effect of the sodium formate substrate on the methanol yield and the methanol content was measured after 3 hours of reaction. The product was detected by Shimadzu Production gas chromatograph with a BID detector. The corresponding calibration curve was prepared by using the known concentrations of methanol ranging from 0.010 to 5.0 mM. Using an external standard method, the methanol concentration was calculated from the area corresponding to the characteristic peak of methanol observed.

The three-enzyme CO₂ conversion system consisted of 100 mM PB buffer saturated with CO₂ for 1 hour, NADH, FDH, FaldDH, and ADH, which had been s were added

to a 10mL glass long-necked tube to a volume of 1mL or 0.5 mL. After the reaction system was prepared, CO₂ was fed at a flow rate of 10 mL/minutes. The reaction system with sodium bicarbonate as substrate consists of 50 mM sodium bicarbonate, which replaced the CO₂ in the system. All the comparison of FaldDHs was performed in the same condition with 5mM NADH and 0.4 mg/mL FDH, 0.8mg/mL FaldDH and 0.1 mg/mL ADH for a three-hour reaction time. Yield-time curve was obtained by measuring the methanol production from 1 to 6 hours by withdrawing the sample at one hour interval. Afterward, a gradient of 0.2, 0.4, 0.6, 0.8, 1.2, 1.6, 2.0 mg/mL of FaldDH was added to the system and the methanol content was measured after 3 hours of reaction. The protein concentration was 0.8 mg/mL and the rest of the conditions were controlled. 5-50 mM NADH was added for 3 hours and the methanol content was measured. The NADH regeneration system contained 10mM NAD⁺ 10 mM, 0.4mg/ml PTDH for NADH regeneration ,0.4 mg/ml FDH, 0.8 mg/ml FaldDH. 0.1mg/ml ADH, 10mM Sodium phosphite and 10mM NaHCO₃.

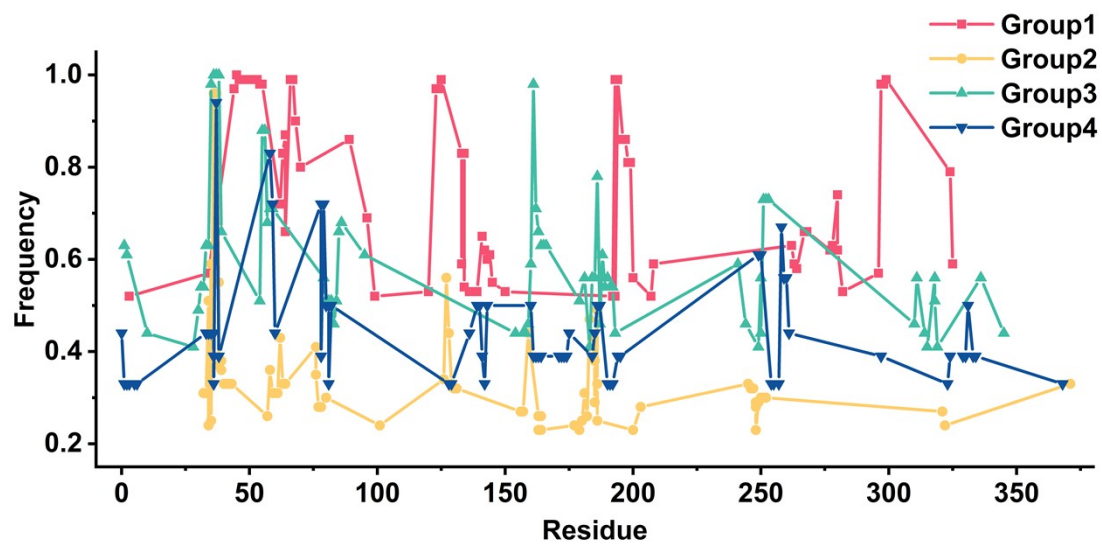


Fig. S1. Frequency of short peptides in different PPR groups (subgroup 1 to subgroup 4 was selected).

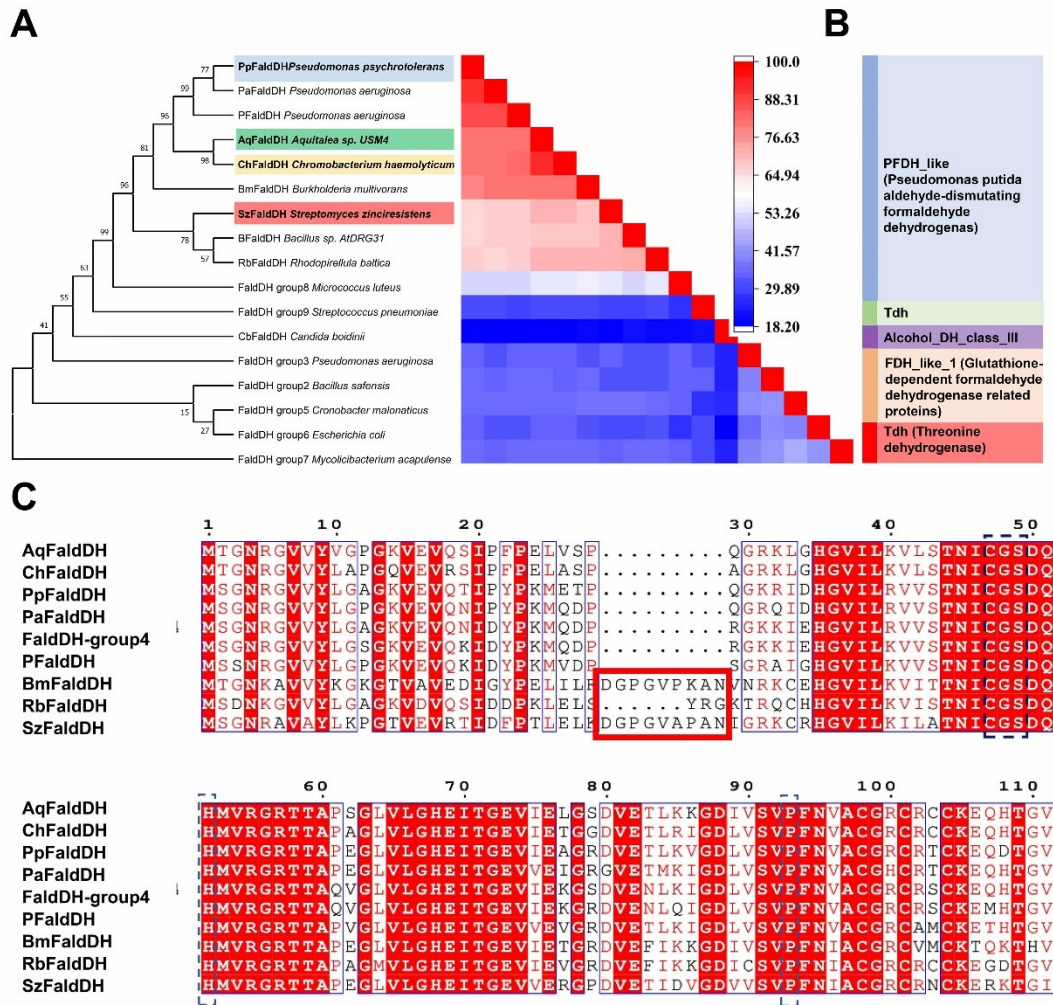


Fig. S2. Sequence analysis of FaldDHs. (A) Phylogenetic tree of FaldDHs. The colored bar represents the level of the percent identity of the enzymes. (B) Subfamily of FaldDHs. Different color stands for different subfamilies, data were collected from NCBI. (C) Multiple sequence alignments of FaldDHs in group 1, residue 1-110 were exhibited. The extra sequences (loop) were marked with red boxes and active site was marked with blue boxes.

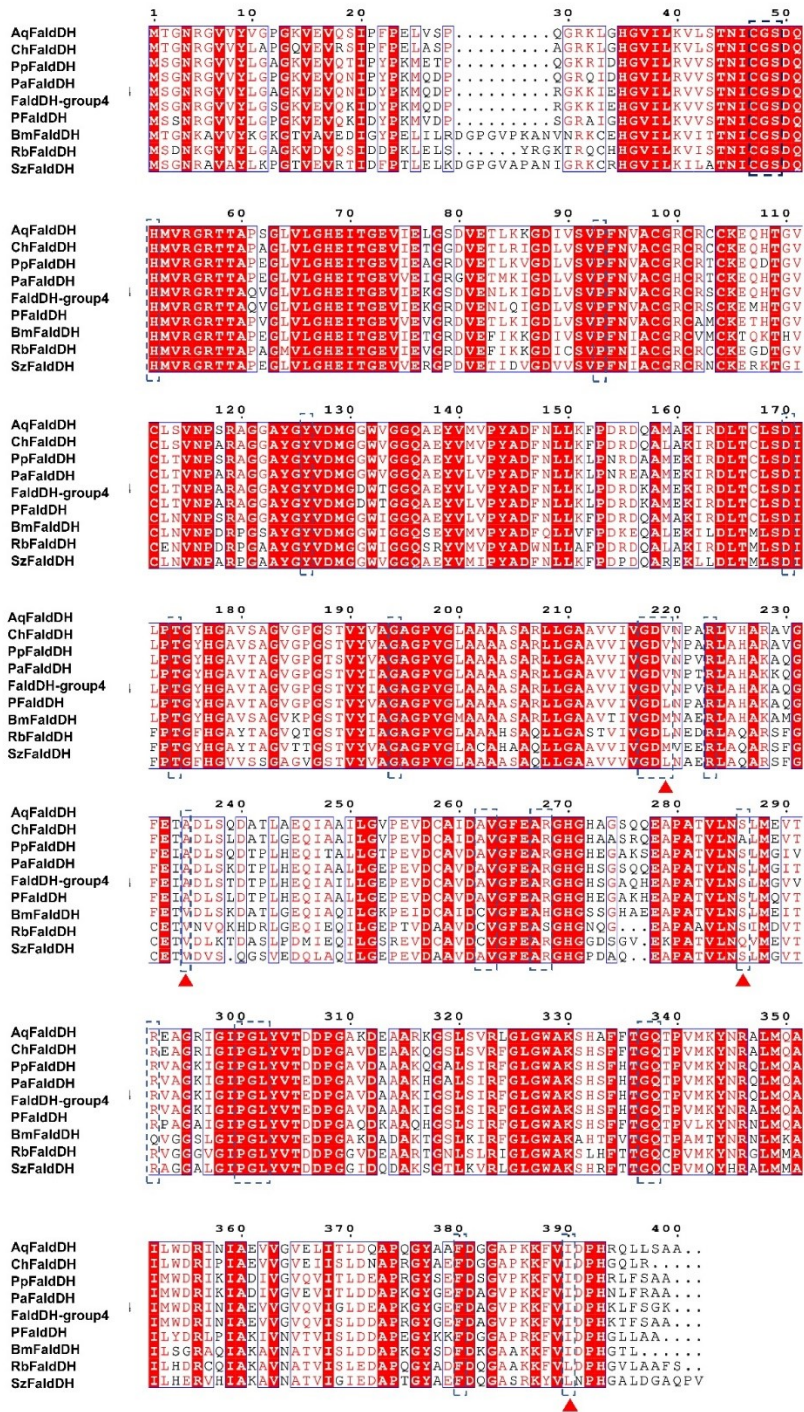


Fig. S3. Multiple Sequence Alignment of FaldDHs. The blue solid line frame is the highly conserved sequence, the red sequence is the completely conserved sequence, the dotted line frame is the active site, and the triangle label is the amino acid with differences in the active site.

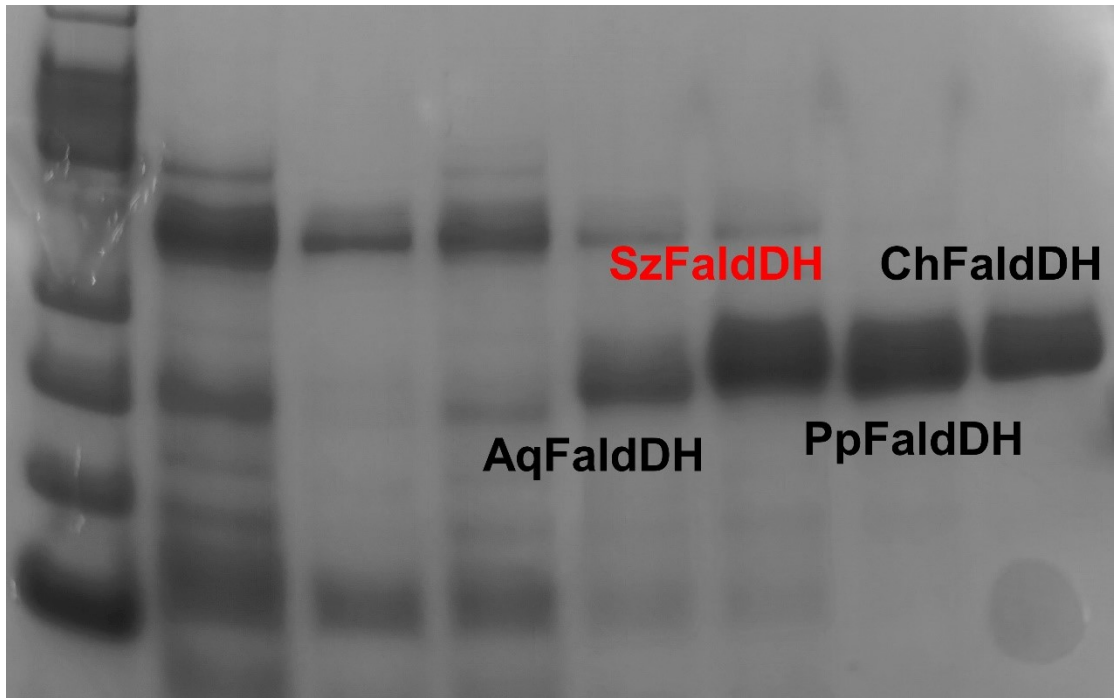


Fig. S4. SDS-page of discovered FaldDHs.

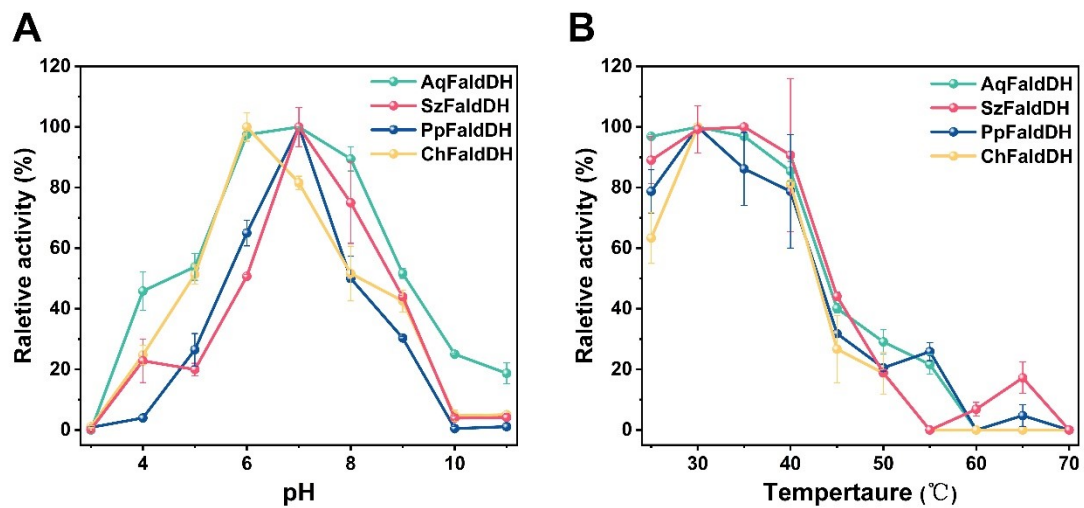


Fig. S5. Effect of temperature and pH on FaldDH activity. **(A)** Effect of temperature on FaldDHs. **(B)** Effect of pH on FaldDHs.

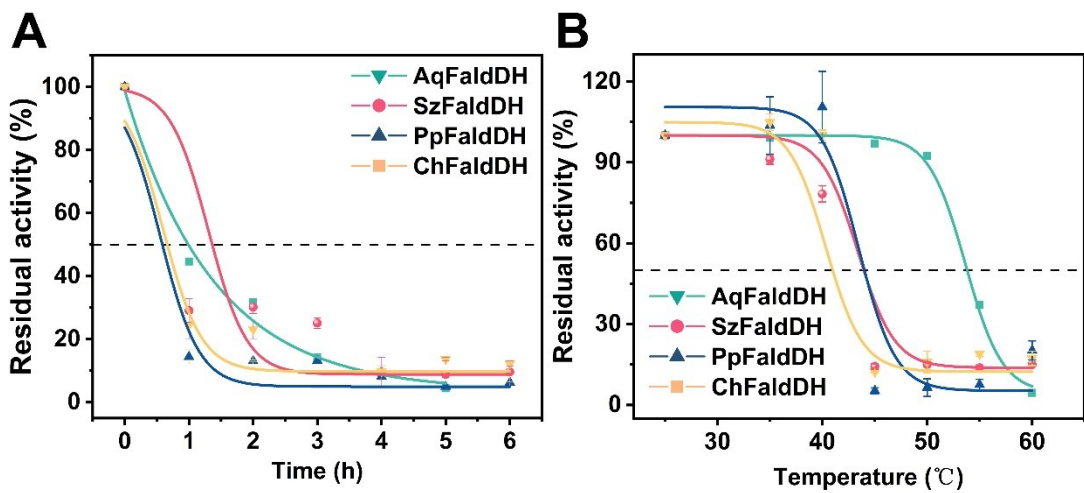


Fig. S6. Thermostability of FaldDH. **(A)** Change of residual activity with time. **(B)** Change of residual activity with incubating time.

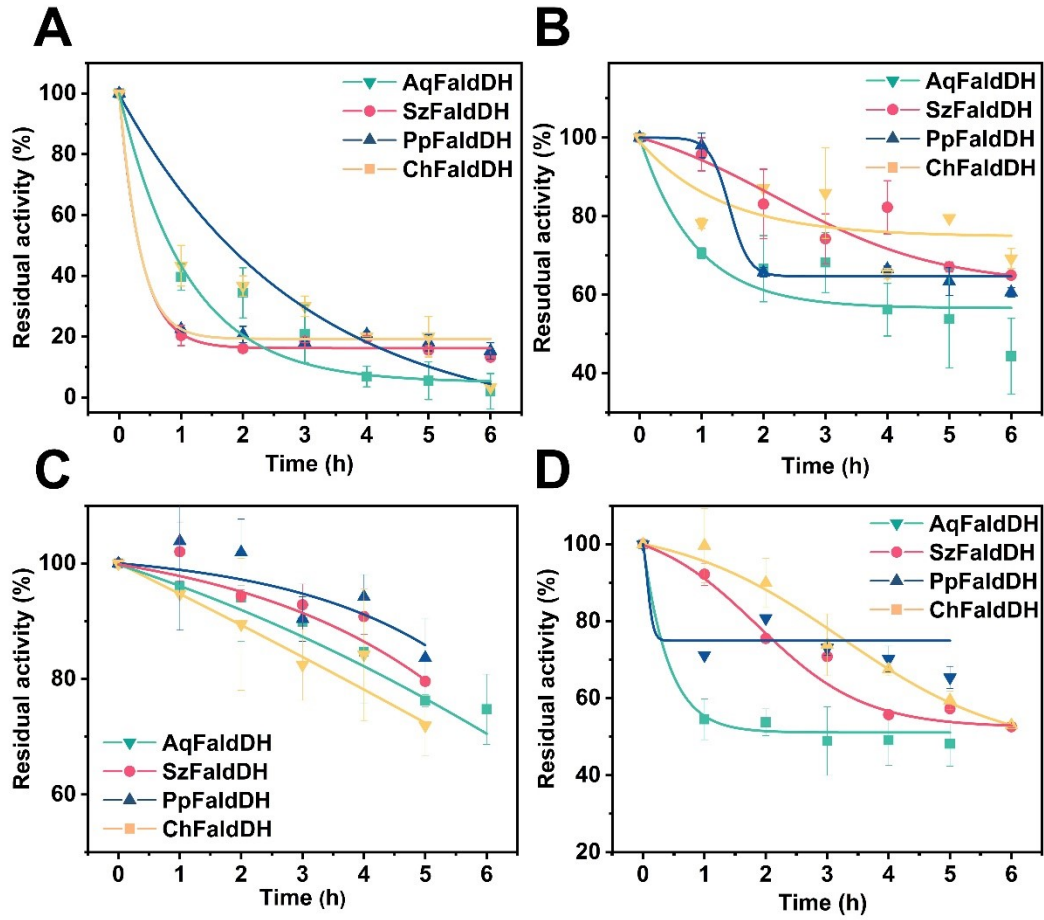


Fig. S7. Acid and alkali tolerance of FaldDHs. **(A)** Tolerance at pH 5. **(B)** Tolerance at pH 6. **(C)** Tolerance at pH 8. **(D)** Tolerance at pH 9.

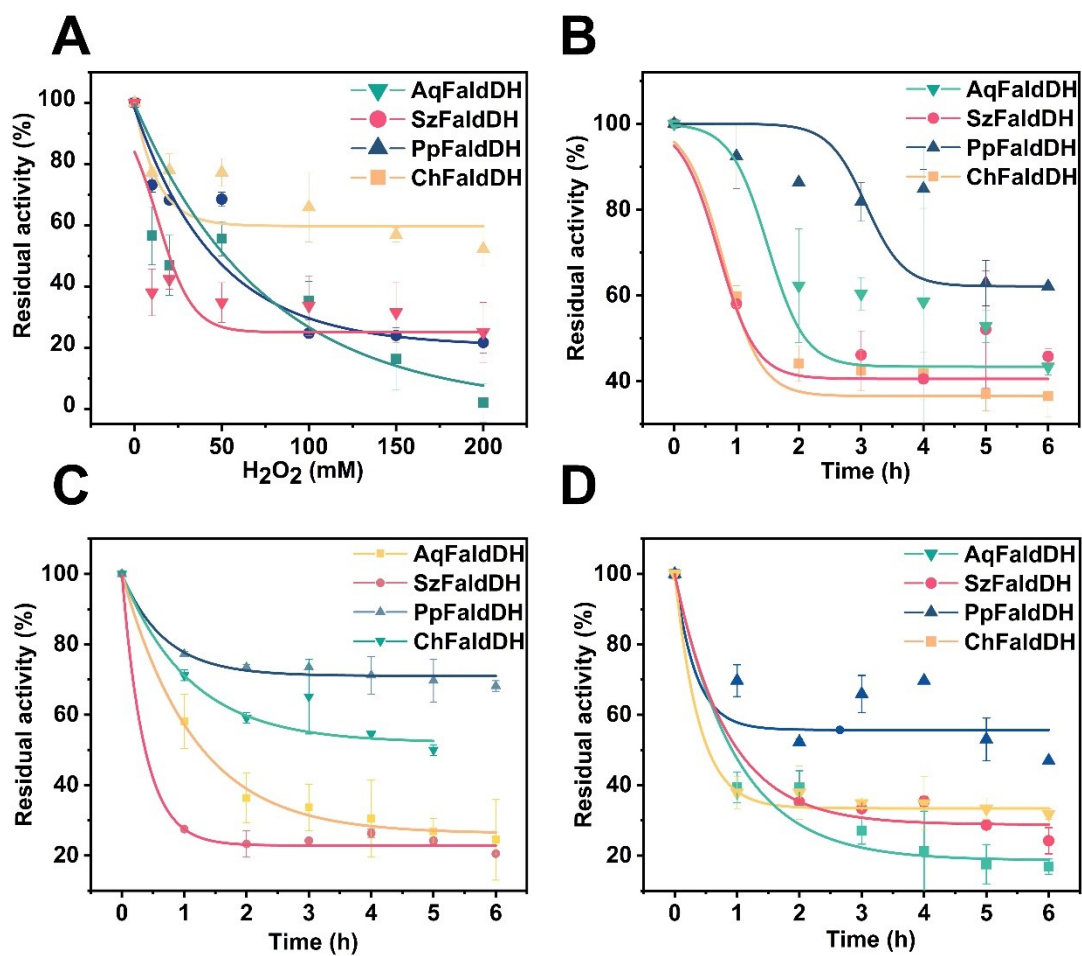


Fig. S8. Oxidative stability of FaldDHs. **(A)** Residual activity at different hydrogen peroxide gradients. **(B)** Residual activity at different time gradients 5mM H₂O₂ **(C)** Residual activity at different time gradients 10mM H₂O₂. **(D)** Residual activity at different time gradients 25mM H₂O₂.

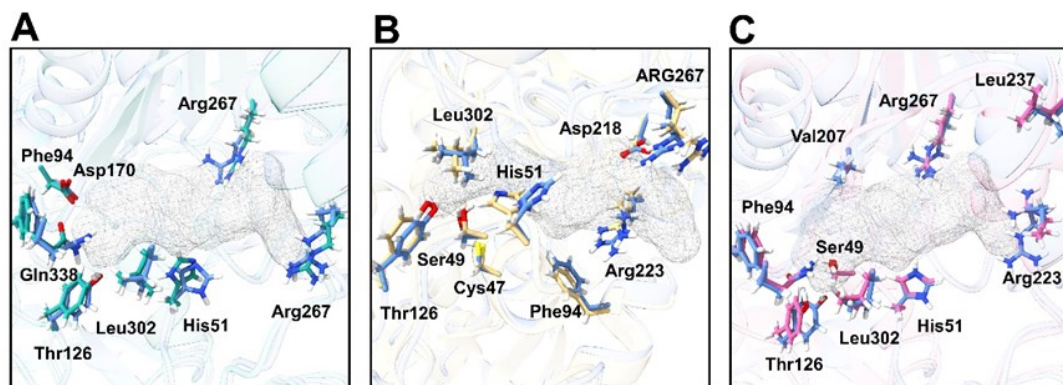


Fig. S9. Comparison of residue in FalddH tunnel, blue for SzFaldDH. **(A)** SzFaldDH with AqFaldDH. **(B)** SzFaldDH with PpFaldDH. **(C)** SzFaldDH with ChFaldDH.

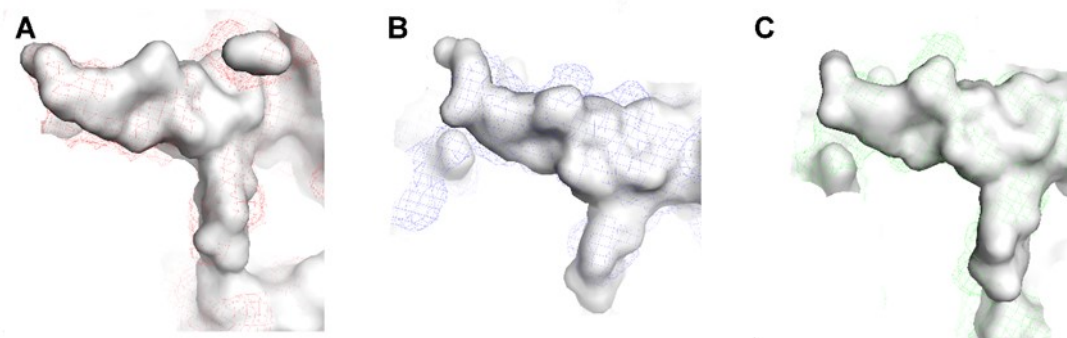


Fig. S10. Comparison of FalddHs tunnel shape. **(A)** SzFaldDH with AqFaldDH. **(B)** SzFaldDH and PpFaldDH. **(C)** SzFaldDH and PpFaldDH. **(D)** SzFaldDH and PpFaldDH.

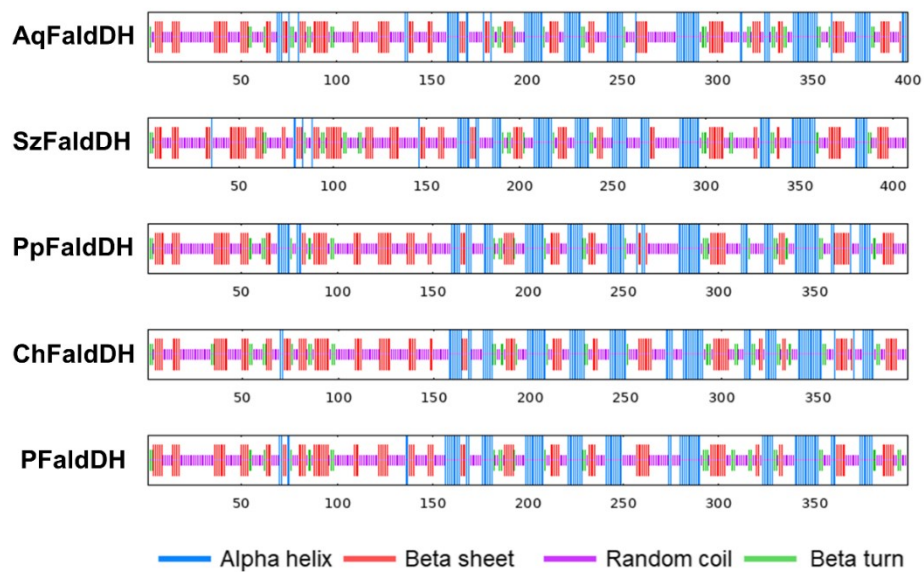


Fig. S11. Secondary structure distribution of FaldDHs.

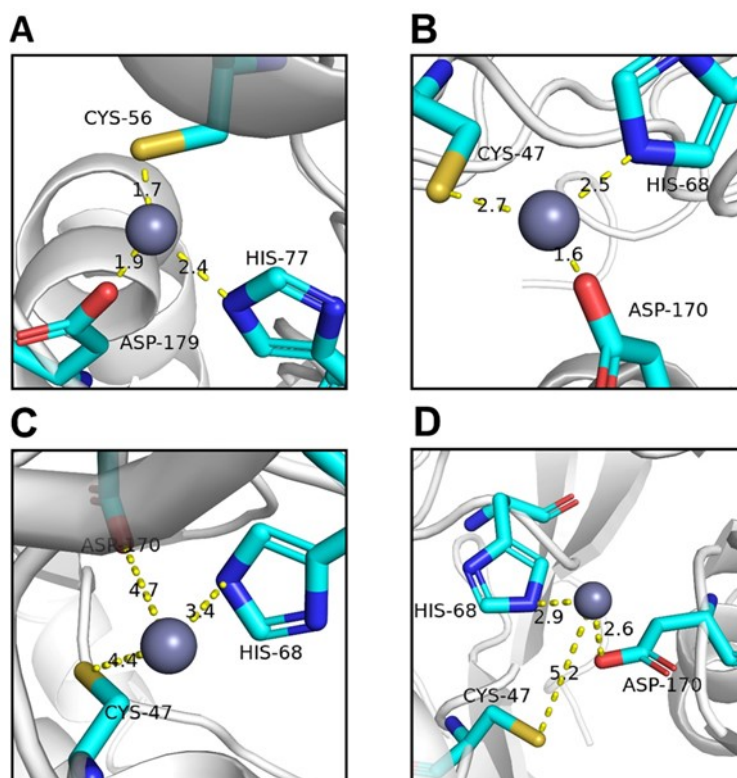


Fig. S12. Catalytic Zinc atom of FaldDHs. (A) AqFaldDH. (B) SzFaldDH. (C) PpFaldDH. (D) ChFaldDH.

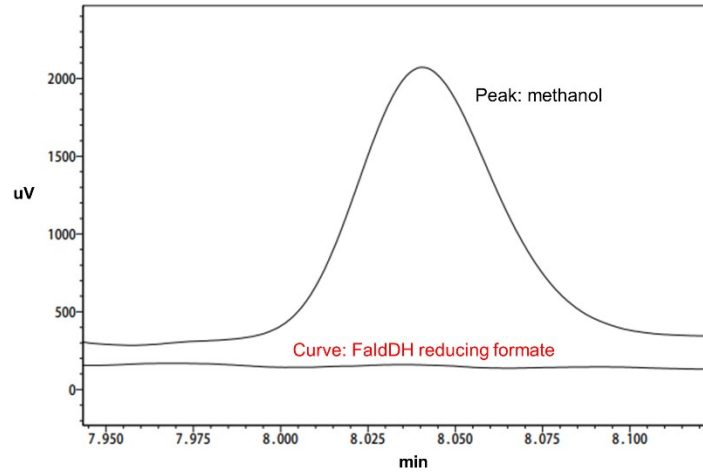


Fig S13. Characterization of FaldDH product selectivity.

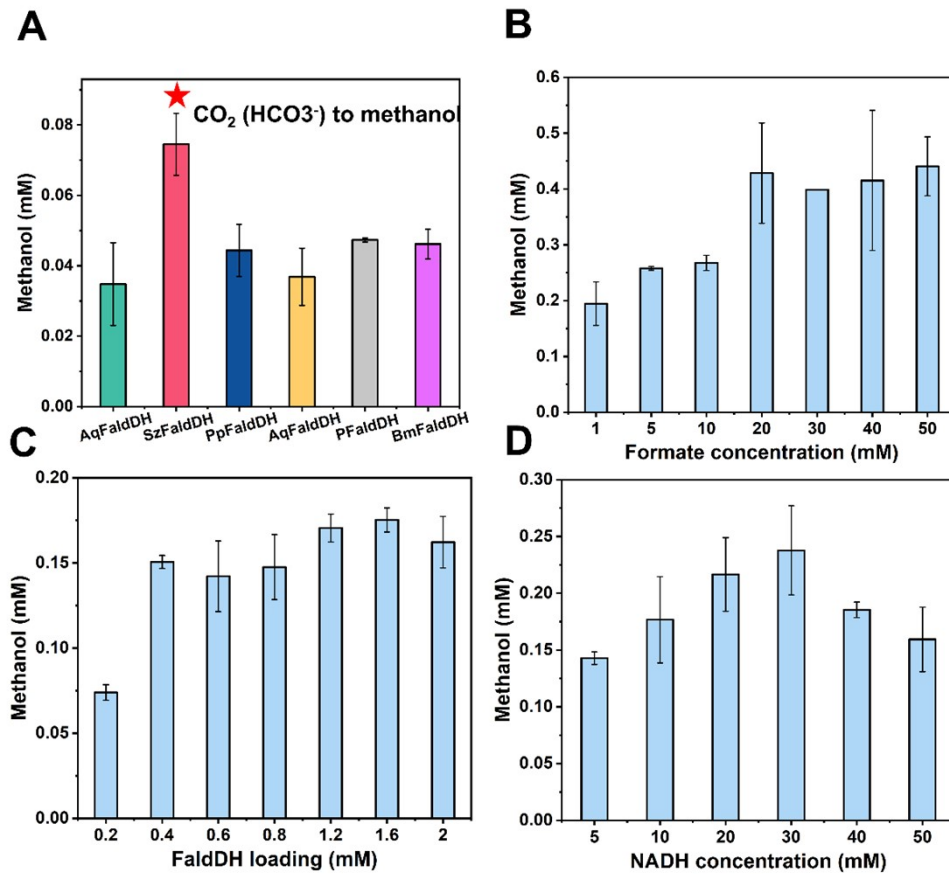


Fig. S14. (A) Methanol production from sodium HCO_3^- as substrate with FDH, FaldDH and ADH. (B) Effect of FaldDH loading on methanol yield. (C) Effect of NADH concentration on methanol yield. (D) Effect of sodium formate concentration on

methanol yield. All reactions were performed at 30 °C in phosphate buffer (100 mM, pH 7.0) with 10 mM NADH and 1-50 mM sodium formate.

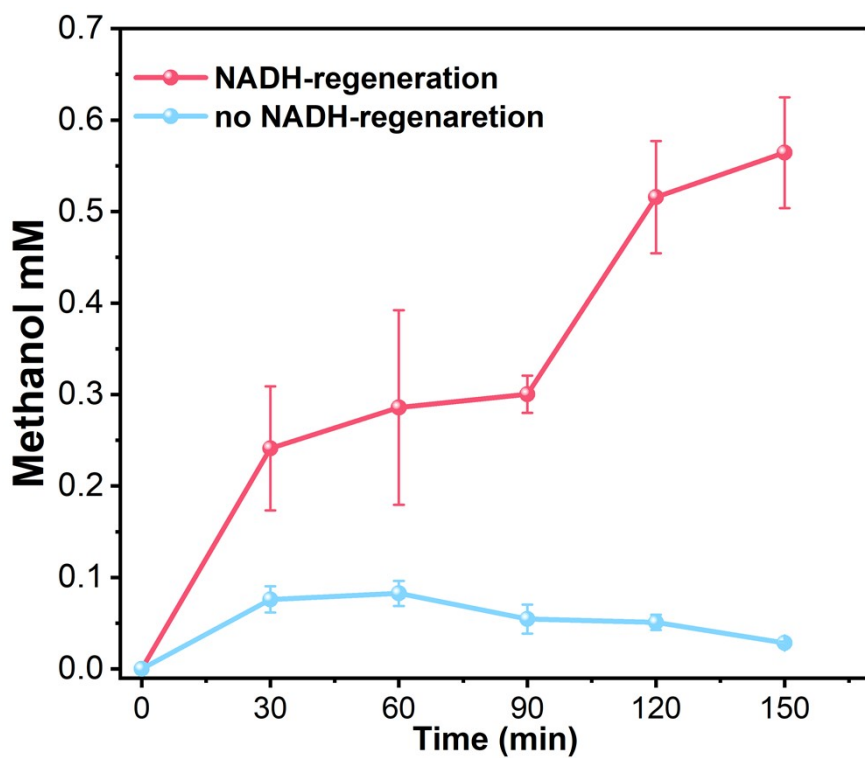


Fig. S15. NADH regeneration of a three enzymatic system. All reactions were performed at 30 °C in phosphate buffer (100 mM, pH 7.0). The NADH-regeneration system contained 10 mM NAD⁺, 0.4mg/ml PTDH, 10mM Sodium phosphite and 10mM NaHCO₃⁻.

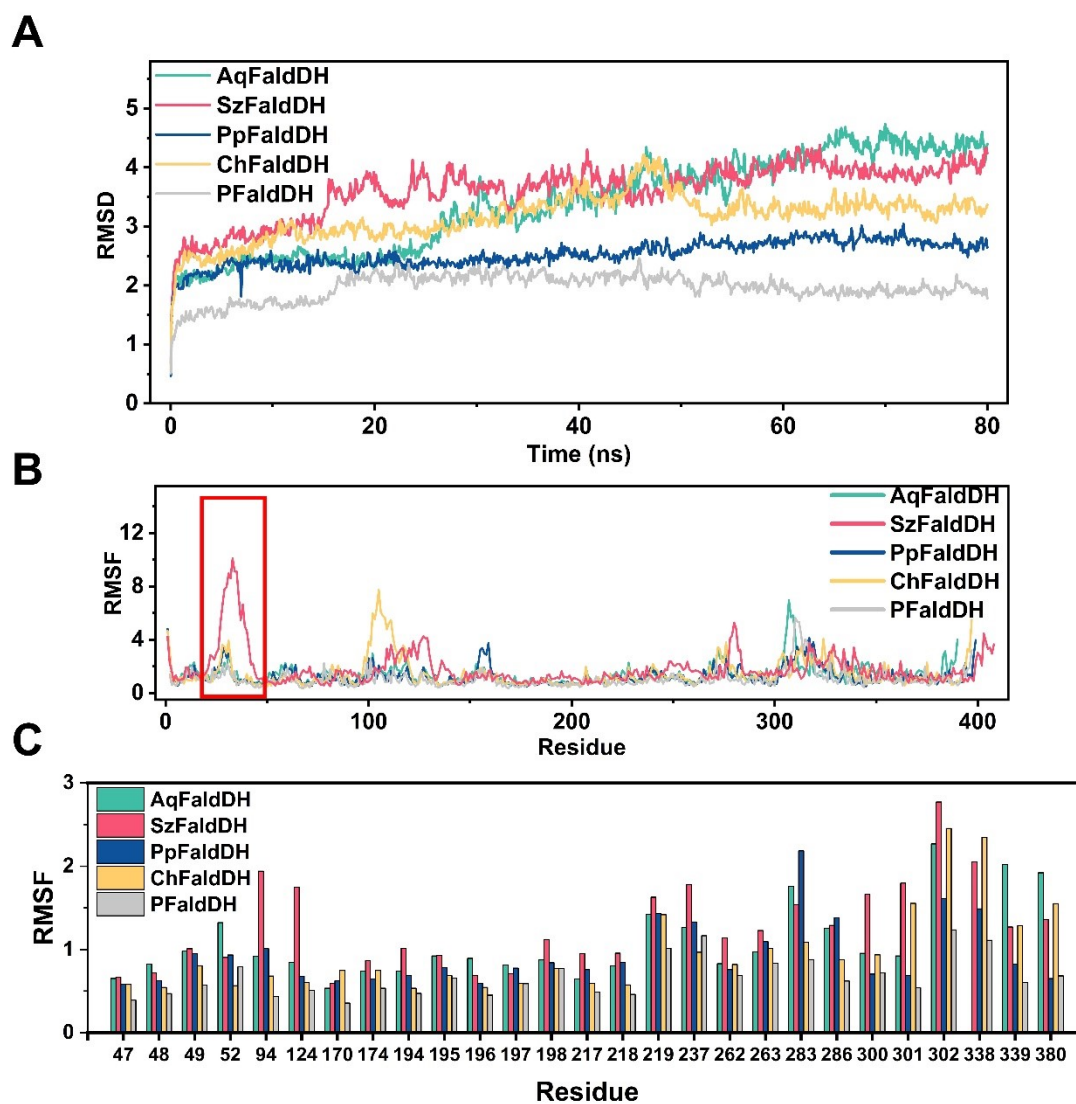


Fig. S16. RMSD and RMSF values for FaldDHs **(A)** RMSD of FaldDHs. **(B)** RMSF of FaldDH. **(C)** RSMF of residue relating to active site.

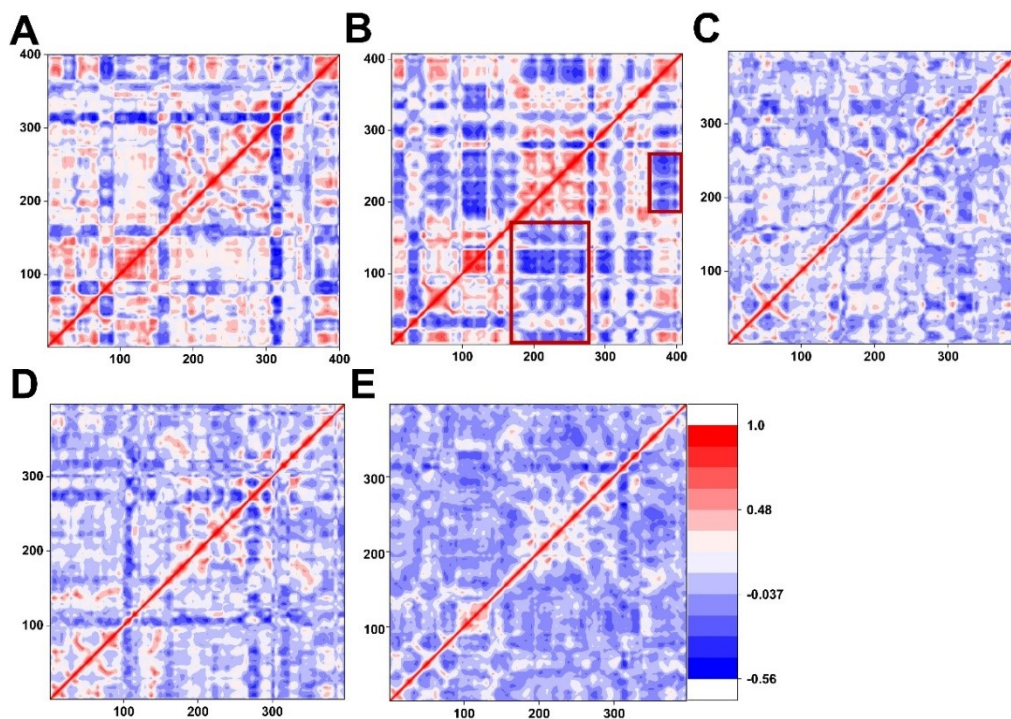


Fig. S17. Dynamic cross-correlation matrix (DCCM), with x-axes and y-axes corresponding to the amino acid residue number. Residues in the two domains are shown in the red boxer. **(A)** AqFaldDH **(B)** SzFaldDH **(C)** PpFaldDH **(D)** ChFaldDH **(E)** PFaldDH

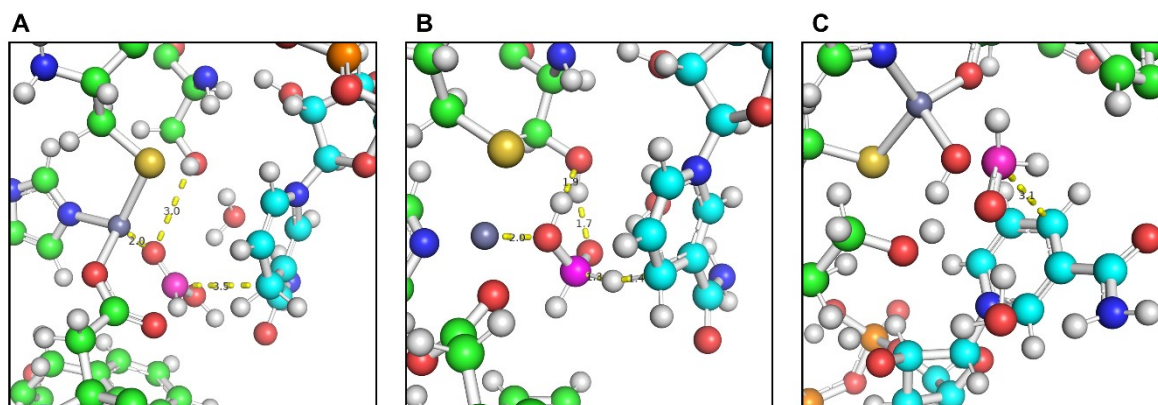


Fig. S18. Structure scheme of QM region of SzFaldDH. **A.** Reactant state. **B.** Transition state. **C.** Product state

Table S1. Comparison of different carbon dioxide conversion pathways to

Conversion method	Catalyst	Reaction condition	Advantage/disadvantage compared with this study	Ref.
	Homogeneous ruthenium catalyst	90°C, 90 bar H ₂ / 20 bar CO ₂		8
Chemical Reduction	Hydroborates	1 atm of CO ₂ , 25 to 80 °C in less than 3 h	Higher temperature and pressure in reaction condition. extra H ₂ needed,	9
	Polarized Hydroxyapatite	140 °C, 6 bar		10
Electrochemical Reduction	Cobalt Phthalocyanine	0.1 Mm KOH with pH 13	Low chemical selectivity	11
	Pt/Pt _{OX}	1 mol L ⁻¹ KHCO ₃		12
	Na ₂ Ti ₆ O ₁₃ (NTO) and K ₂ Ti ₆ O ₁₃ (KTO)	UV irradiation		13
Photochemical Reduction	rGO-grafted NiO-CeO ₂	300 W Xenon lamp	Extra input energy	14
	Biofunctionalized TiO ₂ film	300 W Xe lamp (L25, max = 500 nm)		15
Photoelectrochemical	Cu/rGO/PVP/Nafion	AM-1.5 solar simulated radiation	Lower selectivity for formaldehyde	16
Biocatalysis	FDHPa+SzFaldDH	PB buffer pH7, 30°C	This study	
Chemical Reduction+ Biocatalysis	ZnO+ZrO ₂ catalyst followed with AOX	Chemical :5.0 MPa, 320 to 315 °C, V(H ₂)/V(CO ₂)/V(Ar)=72:24:4. Biocatalysis: pH 7.5 30 °C	High temperature, hydrogen peroxide produced in enzymatic process	17

formaldehyde. AOX: alternative oxidase

Table S2. Enzyme discoveries with peptide pattern recognition (PPR) technology in

FaldDHs

Protein name	Organism	Hits	Frequency	Score
AqFaldDH	<i>Aquitalea sp</i>	51.8	68	68
SzFaldDH	<i>Streptomyces zinciresisten</i>	51.1	67	67

PpFaldDH *Pseudomonas psychrotolerans* 50.7 66 66

Enzyme	V_{\max} ($\mu\text{mol}/\text{min}/\text{mg}$)	k_{cat} (s^{-1})	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ ($\text{s}^{-1}\text{mM}^{-1}$)
AqFaldDH	33.70	0.591	0.344	1.718
SzFaldDH	104.5	1.856	0.201	6.512
PpFaldDH	71.32	1.245	2.657	0.468
ChFaldDH	11.68	0.201	0.122	1.647
PFaldDH	231.4	4.063	23.381	0.173
BmFaldDH	42.14	0.739	0.775	0.9539
ChFaldDH	<i>Chromobacterium haemolyticum</i>	50.6	66	66

Table S3. Comparison of kinetic parameters of FaldDHs in reduction

Table S4. Comparison of kinetic parameters of FaldDHs in oxidation.

Enzyme	V_{\max} ($\mu\text{mol}/\text{min}/\text{mg}$)	k_{cat} (s^{-1})	K_{m} (mM)	$k_{\text{cat}}/K_{\text{m}}$ ($\text{s}^{-1}\text{mM}^{-1}$)
AqFaldDH	185.3	3.250	0.826	3.934
SzFaldDH	12.77	0.227	0.219	1.036
PpFaldDH	382.6	6.71	1.087	6.172
ChFaldDH	451.4	7.919	3.424	1.647
PFaldDH	165.9	2.91	0.39	7.461
BmFaldDH	208.6	3.65	0.481	7.88

Table S5. Secondary structure of FaldDHs

Protein name	α -helix	β -sheet	β -turn	Radom coil
AsFaldDH	22.25%	25.75%	8.25%	43.75%
SzFaldDH	21.57%	24.51%	7.11%	46.81%
PpFaldDH	24.06%	23.81%	8.02%	44.11%
ChFaldDH	24.43%	28.21%	11.34%	36.02%
PFaldDH	22.81%	24.56%	8.27%	44.36%

Enzymes	NADH	NAD ⁺	HCOO ⁻	HCHO
AqFaldDH	12.227	12.604	1.66	1.57
SzFaldDH	10.838	10.376	1.93	1.38
PpFaldDH	10.488	10.492	2.086	1.230
ChFaldDH	10.064	10.388	0.72	0.41
PFaldDH	10.75	10.86	0.287	0.356

Table S6. Binding energy for molecular docking with different substrate.

Reference

1. X. Ji, Y. Xue, Z. Li, Y. Liu, L. Liu, P. K. Busk, L. Lange, Y. Huang and S. Zhang, *Green Chemistry*, 2021, **23**, 6990-7000.
2. X. Robert and P. Gouet, *Nucleic Acids Research*, 2014, **42**, W320-W324.
3. X. Zhou, W. Zheng, Y. Li, R. Pearce, C. Zhang, E. W. Bell, G. Zhang and Y. Zhang, *Nat Protoc*, 2022, **17**, 2326-2353.
4. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin, *Protein Sci*, 2021, **30**, 70-82.
5. S. K. Kuk, R. K. Singh, D. H. Nam, R. Singh, J. K. Lee and C. B. Park, *Angew Chem Int Ed Engl*, 2017, **56**, 3827-3832.
6. T. Nash, *Biochemical journal*, 1953, **55**, 416.
7. S. B. Jones, C. M. Terry, A. S. Lister and D. C. Johnson, *Analytical Chemistry*, 1999, **71**, 4030-4033.
8. M. Siebert, M. Seibicke, A. F. Siegle, S. Krah and O. Trapp, *J Am Chem Soc*, 2019, **141**, 334-341.
9. D. Zhang, C. Jarava-Barrera and S. Bontemps, *ACS Catalysis*, 2021, **11**, 4568-4575.
10. J. Sans, V. Sanz, P. Turon and C. Alemán, *ChemCatChem*, 2021, **13**, 5025-5033.
11. E. Boutin, M. Wang, J. C. Lin, M. Mesnage, D. Mendoza, B. Lassalle-Kaiser, C. Hahn, T. F. Jaramillo and M. Robert, *Angew Chem Int Ed Engl*, 2019, **58**, 16172-16176.
12. J. Baessler, T. Oliveira, R. Keller and M. Wessling, *ACS Sustainable Chemistry & Engineering*, 2023, **11**, 6822-6828.
13. L. F. Garay-Rodríguez, L. M. Torres-Martínez and E. Moctezuma, *Journal of Photochemistry and Photobiology A: Chemistry*, 2018, **361**, 25-33.
14. H. R. Park, A. U. Pawar, U. Pal, T. Zhang and Y. S. Kang, *Nano Energy*, 2021, **79**.
15. G. Qin, Y. Zhang, X. Ke, X. Tong, Z. Sun, M. Liang and S. Xue, *Applied Catalysis B: Environmental*, 2013, **129**, 599-605.
16. A. U. Pawar, U. Pal, J. Y. Zheng, C. W. Kim and Y. S. Kang, *Applied Catalysis B: Environmental*, 2022, **303**.
17. T. Cai, H. Sun, J. Qiao, L. Zhu, F. Zhang, J. Zhang, Z. Tang, X. Wei, J. Yang, Q. Yuan, W. Wang, X. Yang, H. Chu, Q. Wang, C. You, H. Ma, Y. Sun, Y. Li, C. Li, H. Jiang, Q. Wang and Y. Ma, *Science*, 2021, **373**, 1523-1527.