Supplementary Material

Whole-cell catalytic production of ethylene glycol from C1 compounds using engineered glycolaldehyde synthase

Xinyu Tian[1,2,#], Jianyu Long[1,2,#], Jia Yin[1,2], Mingfang Zhou[1,2], Chaofeng Shao[1,2], Biqiang Chen[1,2], Tianwei Tan[1,2,3*]*

1. Beijing Key Lab of Bioprocess, National Energy R&D Center for Biorefinery, Beijing 100029, PR China.

2. College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, PR China.

3. Lead contact

[#]These authors contributed equally to this work.
[*]Co-corresponding author.
E-mail address: Biqiang Chen: chenbq@mail.buct.edu.cn;
Tianwei Tan: twtan@mail.buct.edu.cn

Supporting methods

Molecular dynamics simulation

Molecular dynamics simulations were performed using GROMACS-2022 software[1] and GPU acceleration was used. The initial all-atom coordinates were derived from the crystal structure (PDB ID: 6a50)[2]. The amber99sb-ildn[3] all-atom force field was used to model GALS-M3 and WT. After ThDP binds to the enzyme backbone, there will be proton transfer and conformational changes (Figure S1b), so the charge distribution is significantly different from that of ThDP in the AP state. When Mg^{2+} ions cooperate with the coenzyme ThDP, the charge distribution of each atom of ThDP will be affected. Gaussian 16 combined with Ambertools 23 was used to fit the electrostatic potential charge of the Mg^{2+}-ThDP complex and the single ThDP. The functional used b3lyp and the basis set used 6-311+g(d,p). In the modeling process of Mg^{2+}-ThDP complex and single ThDP, ThDP adopts the enamine-carbon anion conformation (Figure S12). The ThDP conformation in this form is the near reaction conformation of the aldehyde condensation reaction[4].

Periodic boundary conditions were applied and the explicit water model TIP3P[5] was used to solvate the entire system. TIP3P water filled the entire dodecahedral water box (the protein was located in the center of the box and the minimum distance from the boundary was 10Å). Na+ and Cl- were used to balance the system charge to make the net charge of the system 0. The cutoff distance for short-range non-bonded interactions (short-range van der Waals forces and short-range Coulomb forces) was set to 14Å, and the long-range electrostatic interactions were treated using the PME method[6]. The grid spacing of the FFT was set to 1.6Å.

The steepest descent algorithm and the conjugate gradient method were used in combination to fully relax the entire system, with a maximum force of 1 kJ/mol/nm. The temperature was increased and the pressure was equilibrated by canonical ensemble (NVT) simulations and isothermal isobaric ensemble (NPT) simulations. During the NVT, the system temperature was uniformly increased to 310 K by simulated annealing, and the temperature was controlled using a V-rescale thermostat[7]. During the NPT, the temperature was controlled at 310 K and the system pressure was

equilibrated at 1 bar using a Berendsen barostat[8] . The equilibrium simulation time was 100 ps. After the NVT and NPT, a formal MD simulation was performed for 20 ns, at which the system temperature was controlled at 310 K using a Nose-Hoover thermostat[9] and the pressure was controlled at 1 bar using a Parrinello-Rahman barostat[10] . The formal MD simulation was performed three times in parallel, each with a different random initial velocity. The Newtonian equations of motion were integrated using the leap-frog algorithm with an integration step of 2 fs, and the bonds were constrained using the LINCS algorithm[11]. During the simulation, both the protein and the ligand were restrained by a force constant of 1000 kJ/mol·nm^2.

After the simulation, the rms and rmsf modules of GROMACS were used to extract the RMSD of the protein skeleton and ThDP and the RMSF of each amino acid of the protein; the cluster module was used to perform protein conformation clustering analysis and extract the stable state protein conformation. The clustering method was gromos[12] , and the RMSD cut-off was set to 0.1nm. Caver2.0[13] was used to analyze the tunneling situation during the simulation, and the stable protein conformation (15ns-20ns) was taken for tunneling scanning calculation. Starting from the 15ns (including the 15ns), the trajectory file was sampled every 10ps, and finally a protein conformation sample containing 501 frames of continuous frames was obtained for tunneling analysis.

Supporting Tables

Table S1. Primers used in this study.

Primers	Sequence
W463X-F	ctgcgtnnkttcgctggtgttctggaagc
W463X-R	cagcgaamnnacgcagcataccgtaggtacc
T87X-F	gctcgtnnktctcactctccgctgatcgt
T87X-R	gagtgagamnnacgagcgttagacagagcaccc
G401X-F	ctgctnnkggtctgggtttcgctctgcc
G401X-R	ccagaccmnnagcagcgcagaagtagtaagaacc
C398X-F	ctacttcnnkgctgctggtggtctgggt
C398X-R	caccagcagcmnngaagtagtaagaacccgg
F397X-F	ctacnnktgcgctgctggtggtctgggtttc
F397X-R	cagcgcamnngtagtaagaacccgggttacgc

3. Supporting Figures

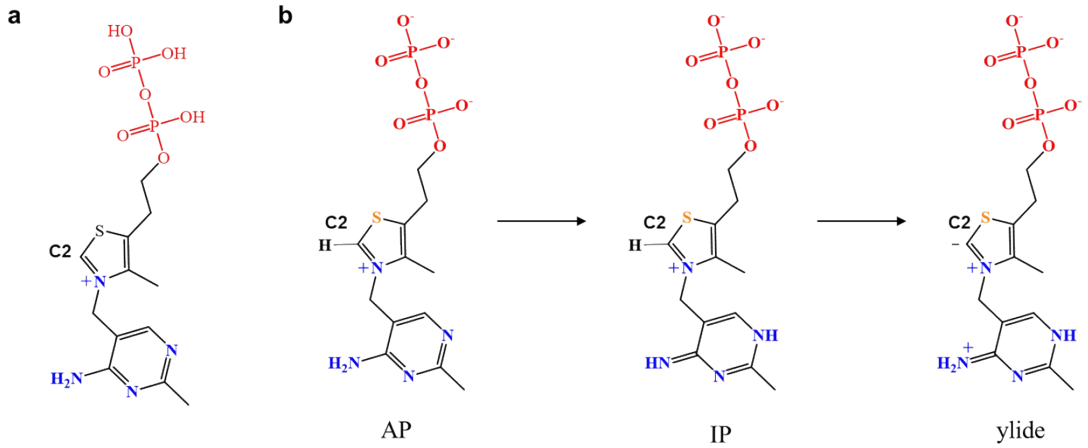


Figure S1. Structure of thiamine pyrophosphate (ThDP) and thiamine (VB1). a. ThDP. b. Structural changes of ThDP during the reaction. ThDP is mostly in the form of AP when not bound to the enzyme. It is activated to IP by proton transfer when bound to the enzyme, and then forms a ylide in the near-reaction state and an enamine-carbanion form by proton transfer.

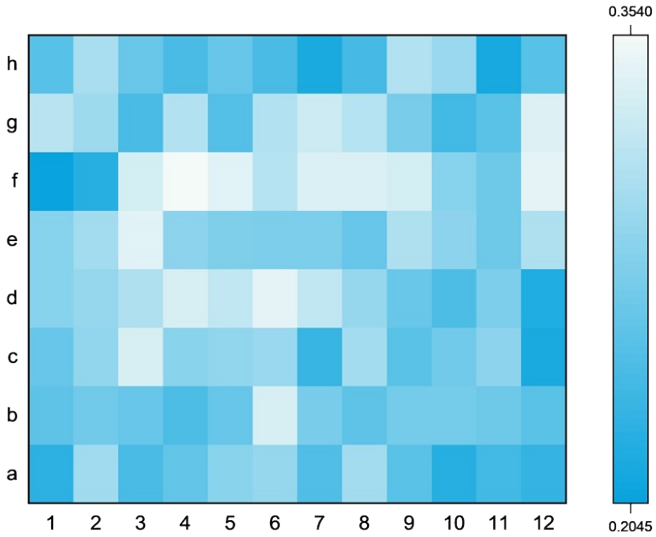


Figure S2. Coefficient of variation (CV) determination of TTC method. WT were cultured in 96-well plates and the response of all WT at 510 nm was measured when derivatized with TTC.

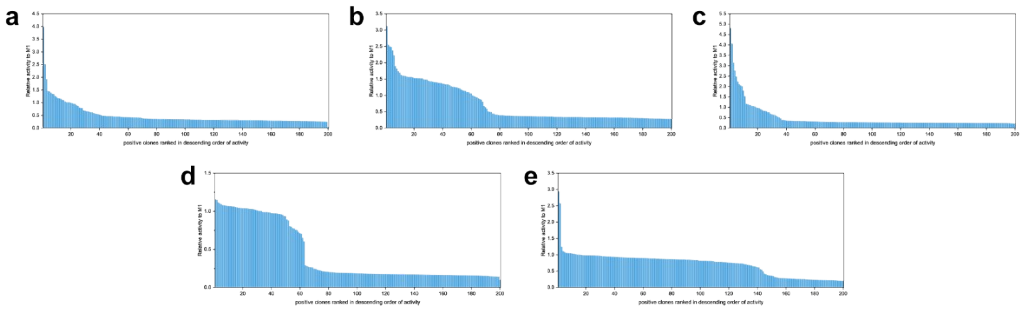


Figure S3. Catalytic activity screening of active clones in the saturation mutant library. The Y-axis corresponds to the relative activity of the variants compared to the WT. The X-axis corresponds to the number of clones in descending order. The clones with an activity 1.18 times higher than that of the WT were selected for sequencing analysis. a. W463X library; b. T87X library; c. G401X library; d. C398X library; e. F397X library.

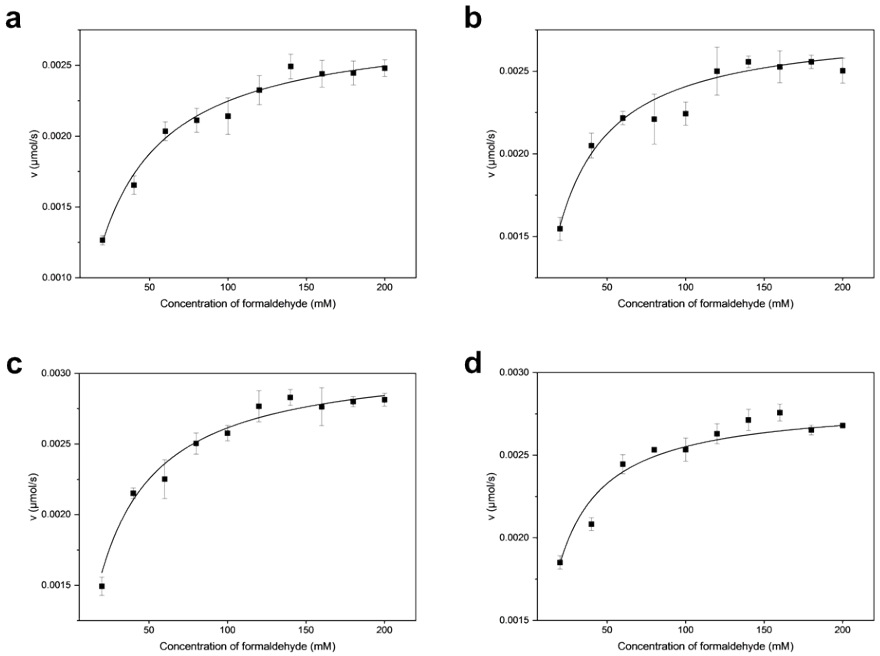


Figure S4. Michaelis-Menten plots of variants catalyzing FALD condensation to form GALD. a. Variant A416T; b. Variant T87A/A416T; c. Variant A416T/W463I; d. Variant T87A/A416T/W463I. The enzyme addition amount was 0.1 mg/mL and the reaction temperature was 37°C. The reactions were performed independently at pH 8. The data in a-d represent means ± s.d., as determined from n = 3 independent experiments.

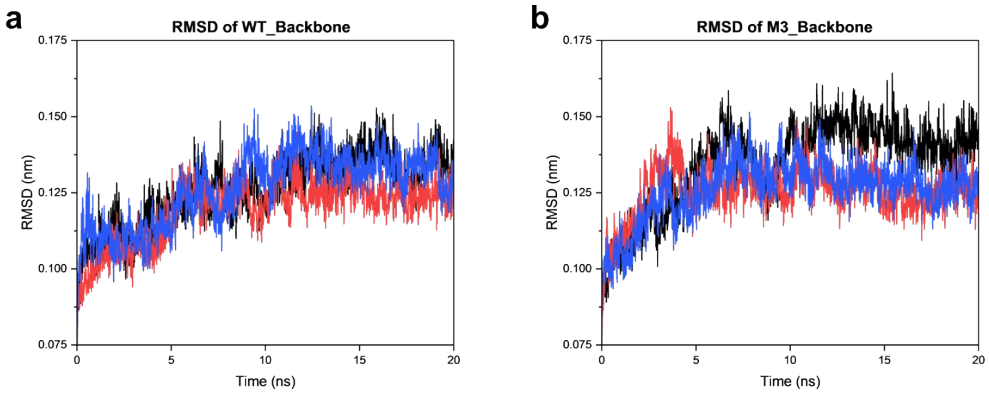


Figure S5. Changes of RMSD of WT and M3 protein backbone over time during three parallel 20 ns MD simulations. a. Changes of RMSD of WT protein backbone over time during three parallel 20 ns MD simulations. b. Changes of RMSD of M3 protein backbone over time during three parallel 20 ns MD simulations.

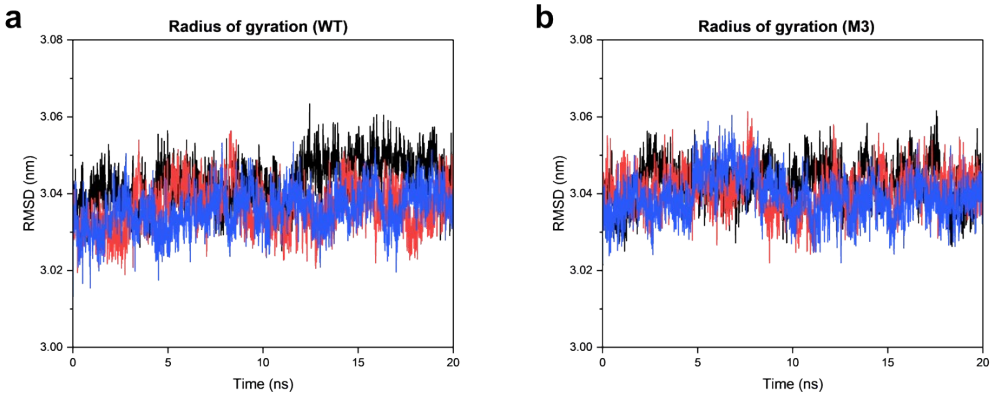


Figure S6. The change of gyration radius of WT and M3 proteins over time during three parallel 20 ns MD simulations. a. The change of gyration radius of WT protein over time during three parallel 20 ns MD simulations. b. The change of gyration radius of M3 protein over time during three parallel 20 ns MD simulations.

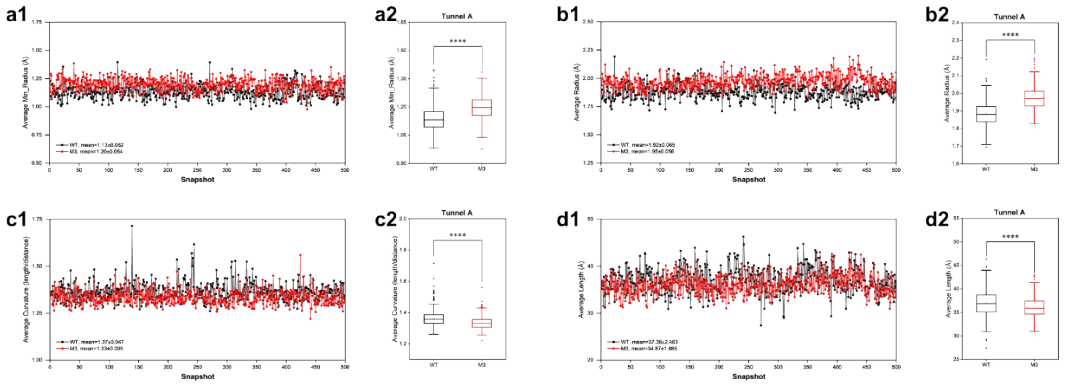


Figure S7. Changes and distributions of Curvature, Length, Min_radius, and Radius (Average_Radius) of Tunnel A in 15 ns-20 ns WT and M3 proteins during the simulation. a. Changes and distributions of Min_radius of Tunnel A in 15 ns-20 ns WT and M3 proteins during the simulation. b. Changes and distributions of Radius (Average_Radius) of Tunnel A in 15 ns-20 ns WT and M3 proteins during the simulation. c. Changes and distributions of Curvature of Tunnel A in 15 ns-20 ns WT and M3 proteins during the simulation. d. Changes and distributions of Length of Tunnel A in 15 ns-20 ns WT and M3 proteins during the simulation.

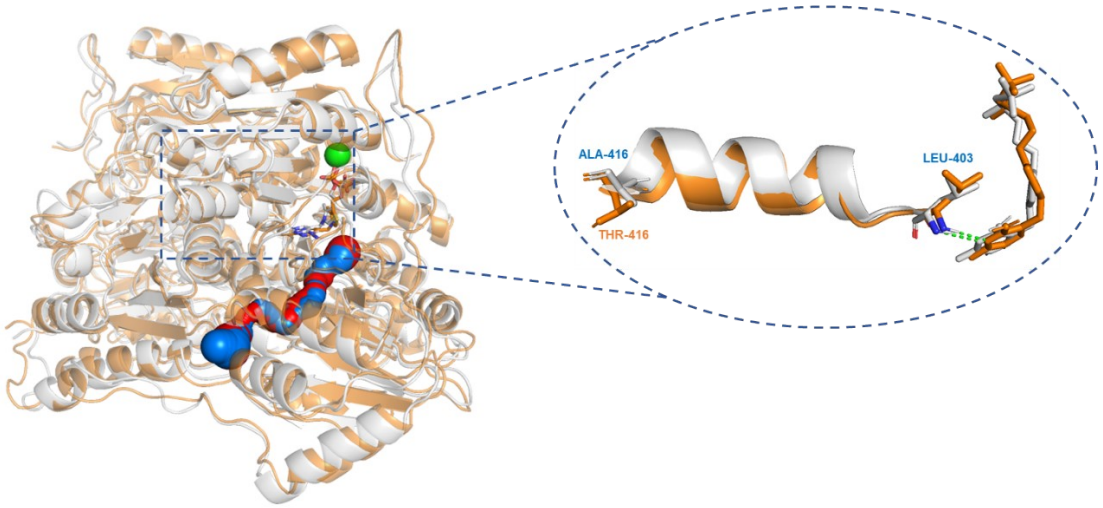
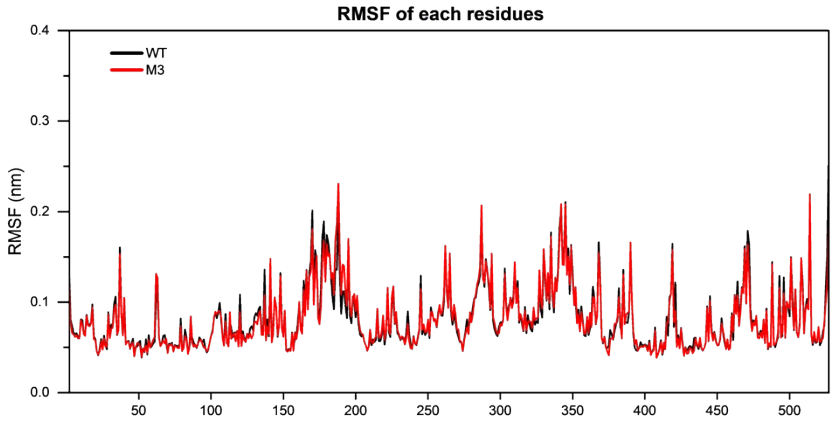


Figure S8. Relative positions and bonding diagrams of residues 416 and L403 in WT and M3. Residues 416, L403, and ThDP are represented as stick models. Gray represents WT; orange represents M3; and green lines represent hydrogen bonds.



129

130 **Figure S9. RMSF of all amino acids for WT and M3 during three parallel 20 ns**

131 **MD simulations.**

132

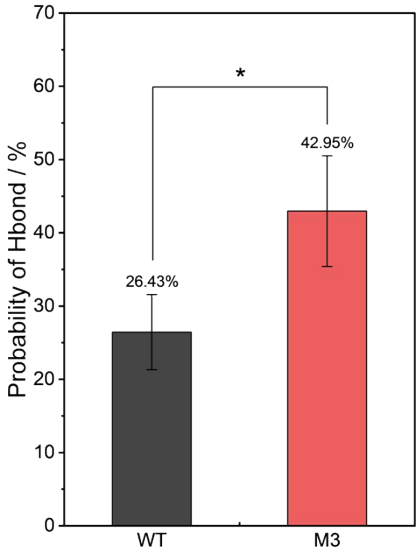
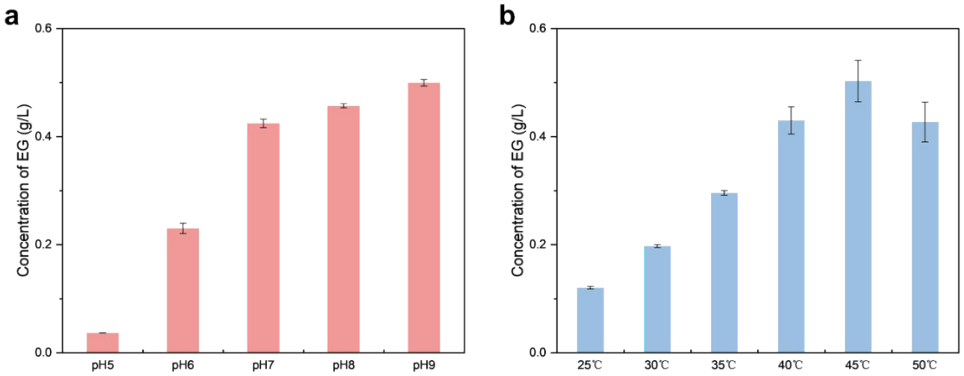


Figure S10. Probability of hydrogen bonding between residue L403 and ThDP in WT and M3.



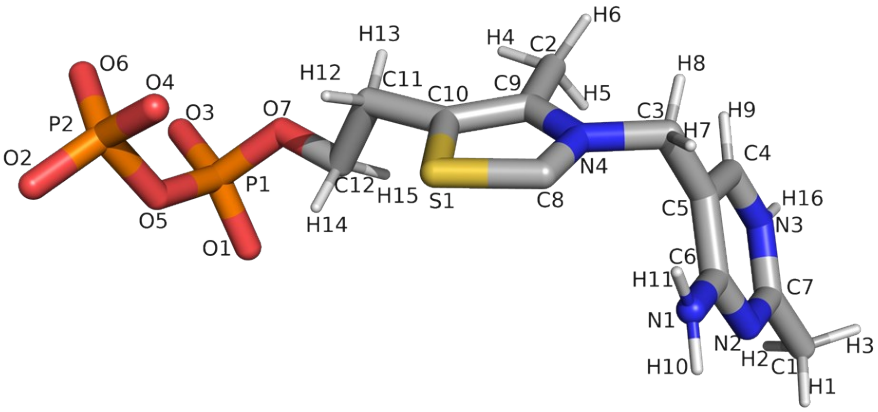
137

138 **Figure S11. Optimization of whole-cell catalytic reaction conditions.** a. Optimizing

139 reaction pH. b. Optimizing reaction temperature.

140

141



142

143 **Figure S12. Initial ThDP modeling conformation and its atomic nomenclature.** The

144 molecule is shown as a stick model, with C atoms in gray, S atoms in yellow, N atoms

145 in dark blue, P atoms in orange, and O atoms in red.

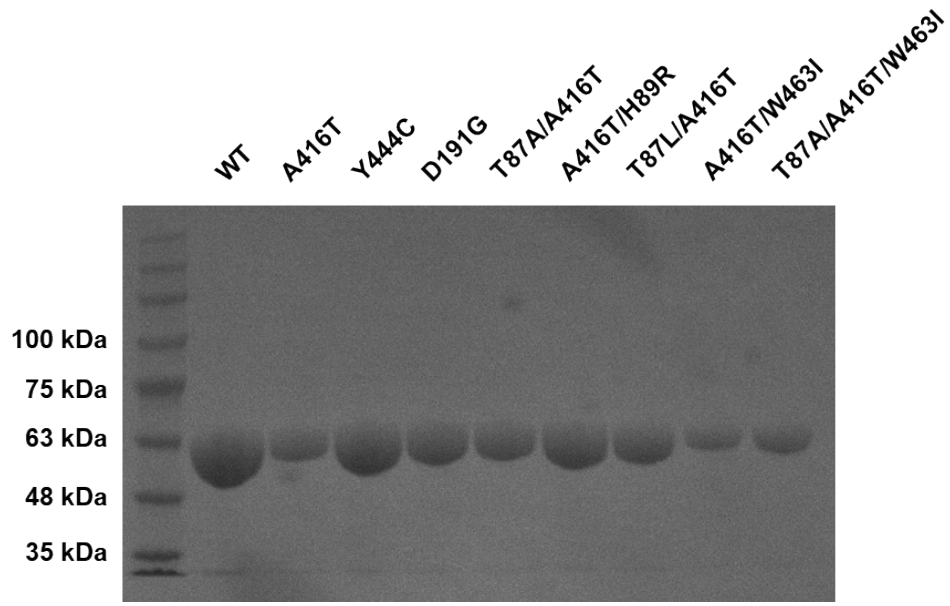


Figure S13. SDS-PAGE of WT and activity-enhanced variants. All sample volumes were kept the same to facilitate observation of the effect of mutations on enzyme expression.

Reference

1. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, SoftwareX, 2015, 1-2, Medium: ED; Size: p. 19-25.
2. X. Lu, Y. Liu, Y. Yang, S. Wang, Q. Wang, X. Wang, Z. Yan, J. Cheng, C. Liu, X. Yang, H. Luo, S. Yang, J. Gou, L. Ye, L. Lu, Z. Zhang, Y. Guo, Y. Nie, J. Lin, S. Li, C. Tian, T. Cai, B. Zhuo, H. Ma, W. Wang, Y. Ma, Y. Liu, Y. Li and H. Jiang, Nat Commun, 2019, 10, 1378.
3. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, Proteins: Structure, Function, and Bioinformatics, 2010, 78, 1950-1958.
4. F. Planas, M. J. McLeish and F. Himo, ACS Catalysis, 2019, 9, 5657-5667.
5. F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist and S. C. L. Kamerlin, The Journal of Physical Chemistry B, 2014, 118, 4351-4362.
6. T. A. Darden, D. M. York and L. G. Pedersen, Journal of Chemical Physics, 1993, 98, 10089-10092.
7. G. Bussi, D. Donadio and M. Parrinello, The Journal of Chemical Physics, 2007, 126.
8. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, The Journal of Chemical Physics, 1984, 81, 3684-3690.
9. D. J. Evans and B. L. Holian, The Journal of Chemical Physics, 1985, 83, 4069-4074.
10. M. Parrinello and A. Rahman, Journal of Applied Physics, 1981, 52, 7182-7190.

11. B. Hess, Journal of Chemical Theory and Computation, 2008, 4, 116-122.

12. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, Angewandte Chemie International Edition, 1999, 38, 236-240.

13. A. Jurcik, D. Bednar, J. Byska, S. M. Marques, K. Furmanova, L. Daniel, P. Kokkonen, J. Brezovsky, O. Strnad, J. Stourac, A. Pavelka, M. Manak, J. Damborsky and B. Kozlikova, Bioinformatics, 2018, 34, 3586-3588.