Supporting Information

Structural insight into the catalytic mechanism of a *cis*-epoxysuccinate hydrolase producing enantiomerically pure D(-)-tartaric acid

Sheng Dong[1,2,3], Xi Liu[4,†], Gu-Zhen Cui[1,2,3,†], Qiu Cui[1,2,3,*], Xinquan Wang[4,*], Yingang Feng[1,2,3,*]

[1] Shandong Provincial Key Laboratory of Synthetic Biology, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[2] CAS Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[3] Qingdao Engineering Laboratory of Single Cell Oil, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101, China

[4] Ministry of Education Key Laboratory of Protein Science, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;

Corresponding Authors

* Yingang Feng, E-mail: fengyg@qibebt.ac.cn; Qiu Cui, E-mail: cuiqiu@qibebt.ac.cn; Xinquan Wang, E-mail: xinquanwang@mail.tsinghua.edu.cn.

Table of Contents

SI Experimental Procedures

Protein Expression and Purification

The procedure of the expression and purification of CESH[D] and its mutants were same as the previous report[1]. Briefly, the codon-optimized gene of CESH[D] was cloned into pET28a between *Nde*I and *Eco*RI restriction sites. Overlapping PCR was used to construct the site-directed mutations. The proteins with an N-terminal His_6-tag were expressed in *E. coli* strain BL21(DE3) and purified using a Ni^{2+} Sepharose HP column (GE healthcare). The proteins were further purified by anion exchange chromatography using a Q-sepharose column and gel filtration using a Superdex 75 column (GE Healthcare). The purity of target proteins was greater than 95% as judged by polyacrylamide gel electrophoresis. The concentrations of proteins were determined by BCA assay kit (Thermo). The protein was stored at -80ºC in the storage buffer comprised of 10 mM Tris-HCl, pH 7.0, and 100 mM NaCl.

Activity Measurement

The measurement of the activity of CESH[D] and its mutants followed the procedure in the previous report[1]. Briefly, 0.1 mL enzyme solution was added into 0.9 mL 60 mM disodium *cis*-epoxysuccinate in 50 mM sodium phosphate buffer, pH 7.5. The solutions were incubated at 37 ºC for 20 min, and the reactions were terminated by adding 0.4 mL 1.0 M H_2SO_4. Then, 1 mL of 1 % (w/v) ammonium metavanadate was added to the reaction solution, and the solution was diluted to 10 mL. After a lapse of 5 min, the absorbance at 480 nm was measured on a Synergy HT Multi-mode microplate reader (BioTek Instruments, Inc.). The concentration of tartaric acid in the reaction solution was then calculated according to the standard curve recorded at various concentrations of tartaric acid. One unit of enzyme activity was defined as the amount of the enzyme that generated 1 μmol of tartaric acid per minute. To detect the residual activity of several CESH[D] mutants, an assay with 500 mM CES and 16 h reaction time was performed and the total conversion ratio from CES to D-TA was calculated.

Crystallization, Data Collection, and Structure Determination

The wild-type CESH[D], inactive mutant CESH[D]-D115A, CESH[D]-D115N, CESH[D]-E190Q, and CESH[D]-D251N were used for crystallization. In order to obtain the structure complex with CES, the inactive mutants were diluted to 0.5 mM by mixing with 10 mM Tris-HCl, pH 8.0, supplemented with 10 mM CES prior to crystallization. Crystallization conditions were determined using commercial high throughput screening kits purchased from Hampton Research and a nanoliter drop-setting Mosquito robot operating with 96-well plates. Crystallization conditions generating crystals was further optimized in 24-well crystallization plates. The crystallization conditions generating crystals used for x-ray data collection were shown in Table S1. The data were collected at the Shanghai Synchrotron Radiation Facility, Beamline BL17U[2], in a 100 K nitrogen stream. Data indexing, integration, and scaling were conducted using HKL2000[3] (wild-type CESH[D]) and XDS[4] (mutants D115A). The wild-type CESH[D] structure was determined by molecular replacement using CCP4i program suit and the PDB file 2Y7D as a search model[5, 6]. The structures were built automatically by ARP/wARP in the CCP4i

program package[7]. Refinement of the structures was performed using the programs COOT and PHENIX[8, 9]. The structures of wild-type CESH[D] was used as a search model to solve its mutant structures. Data collection and refinement statistics are shown in Table S1. All molecular graphics were created with PyMOL (Schrödinger, LLC. http://www.pymol.org/). The atomic coordinates and structure factors have been deposited in the Protein Data Bank (http://www.wwpdb.org) under the PDB IDs 5ZMU and 5ZMY for the wild-type CESH[D] and the mutant D115A, respectively.

References

1. G. Cui, S. Wang, Y. Li, Y. J. Tian, Y. Feng and Q. Cui, *Protein J.*, 2012, **31**, 432-438.
2. Q. S. Wang, F. Yu, S. Huang, B. Sun, K. H. Zhang, K. Liu, Z. J. Wang, C. Y. Xu, S. S. Wang, L. F. Yang, Q. Y. Pan, L. Li, H. Zhou, Y. Cui, Q. Xu, T. Earnest and J. H. He, *Nucl. Sci. Tech.*, 2015, **26**, 12-17.
3. Z. Otwinowski and W. Minor, *Method Enzymol*, 1997, **276**, 307-326.
4. W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 125-132.
5. M. Bellinzoni, K. Bastard, A. Perret, A. Zaparucha, N. Perchat, C. Vergne, T. Wagner, R. C. de Melo-Minardi, F. Artiguenave, G. N. Cohen, J. Weissenbach, M. Salanoubat and P. M. Alzari, *J. Biol. Chem.*, 2011, **286**, 27399-27405.
6. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 235-242.
7. S. X. Cohen, M. Ben Jelloul, F. Long, A. Vagin, P. Knipscheer, J. Lebbink, T. K. Sixma, V. S. Lamzin, G. N. Murshudov and A. Perrakis, *Acta Crystallogr. D Biol. Crystallogr.*, 2008, **64**, 49-60.
8. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 213-221.
9. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 486-501.

SI Tables

Table S1. Data collection and refinement statistics.

	CESH[D]-WT	**CESH[D]-D115A**
PDB code	5ZMU	5ZMY
Crystallization conditions	0.1 M sodium chloride, 0.1 M HEPES pH7.0, 1.6 M ammonium sulfate	0.1 M Tris pH 8.5, 0.1 M magnesium formate dihydrate
Data collection		
No. unique reflections	216905	234863
Space group	P 1 2 1	P 21 21 2
Cell dimensions		
a,b,c (Å)	78.49, 94.25, 94.36	135.16, 158.35, 134.09
α, β, γ (°)	90, 95.90, 90	90, 90, 90
Wavelength (Å)	0.97915	0.97915
Resolution (Å) [a]	39.04-1.50 (1.55-1.50)	158.35-1.87 (1.92-1.87)
Completeness (%)	99.7 (98.4)	99.9 (100)
Multiplicity	3.7 (3.6)	7.3 (7.5)
$I/\sigma(I)$	6.7 (1.8)	9.9 (3.2)
R_{merge}	0.115 (0.664)	0.128 (0.608)
Refinement		
No. reflections	216879	234769
R_{work} / R_{free} (%)	14.89/17.02	16.78/19.47
No. protein molecules / ASU	4	8
No. non-hydrogen atoms	10515	
Protein	8987	17715
Ligands	4	88
Water	1524	2563
B-factors		
Average B-factor	21.91	24.68
Proteins	19.24	23.05
Ligands	15.66	25.95
Solvent	37.71	35.93
R.M.S.D.		
Bond length (Å)	0.010	0.012
Bond angles (°)	1.07	1.39
Ramachandran statistics		
Favored (%)	98.65	98.50
Allowed (%)	1.35	1.50
Outliers (%)	0.00	0.00

[a]Highest resolution shell is shown in parentheses.

[b]Rmerge= $\Sigma_{hkl}\Sigma_i|I(hkl)_i-\langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i\langle I(hkl)_i\rangle$.

Table S2. Distances (Å) between zinc and coordinative atoms.

Atoms	WT	D115A
Zn - Glu14 OE2	2.02 ± 0.03	2.11 ± 0.03
Zn – His47 NE2	2.13 ± 0.02	2.09 ± 0.03
Zn – His49 NE2	2.16 ± 0.02	2.08 ± 0.03
Zn – wt1/D-TA O1	2.21 ± 0.04	2.19 ± 0.10
Zn – wt2/D-TA O3	2.17 ± 0.02	2.24 ± 0.07
Zn – wt3/D-TA O2	2.20 ± 0.01	2.04 ± 0.08
Glu14 OE2 - His47 NE2	2.90 ± 0.01	2.93 ± 0.08
His47 NE2 - His49 NE2	3.28 ± 0.02	3.30 ± 0.09
His49 NE2 - Glu14 OE2	2.87 ± 0.03	2.80 ± 0.08
wt1/D-TA O1 – wt2/D-TA O3	2.90 ± 0.04	3.00 ± 0.07
wt2/D-TA O3 – wt3/D-TA O2	2.82 ± 0.04	2.59 ± 0.07
wt3/D-TA O2 – wt1/D-TA O1	2.83 ± 0.04	2.75 ± 0.06

Table S3. Kinetics parameters of CESH[D] and its mutants. The mutants reported in this paper and those in the reference 2 with relative activity less than 50% are shown in this table. The mutation site of the mutants with relative activity more than 50% reported in the reference 2 are all far away from the active site, which are not shown in this table.

CESH[D]	Relative activity (%)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})	Reference	Mutation Type[3]
Wild-type	100	24.4	20.5	0.84	1	
Wild-type	100.0±0.7	24.5	16.4	0.67	2	
Wild-type	100.0±4.7	25.8	20.3	0.79	this paper	
E9Q	Not detected				this paper	Catalytic-perturbation
R11A	Not detected				this paper	Catalytic
E14Q	Not detected				this paper	Zinc binding / CES binding
H47N	Not detected				2	Zinc binding
H49N	Not detected				2	Zinc binding
R51K	1.4±0.0	20.5	0.23	0.01	2	Zinc binding perturbation
Y64F	41.7±0.7	24.2	6.1	0.25	2	CES binding perturbation
H80F	4.5±1.1	44.4	0.18	0.004	this paper	CES binding perturbation
T82A	Not detected				2	CES binding
P109A	42.0±1.1	27.3	7.3	0.27	2	CES binding perturbation
D110N	31.4±0.2	25.4	6.3	0.25	2	CES binding perturbation
P113A	41.8±3.8	16.2	3.6	0.22	this paper	CES binding perturbation
D115A	Not detected				this paper	Catalytic
Y138F	Not detected				2	CES binding / Catalytic-perturbation
N140D	Not detected				2	Catalytic-perturbation
W164F	1.0±0.0	21.3	0.13	0.01	2	Catalytic-perturbation
E190Q	0.6±0.6	21.7	0.04	0.002	this paper	Catalytic-perturbation
K227A	Not detected				this paper	Catalytic-perturbation
D251N	1.5±0.0	23.3	0.29	0.01	2	Catalytic-perturbation
N262D	26.8±0.2	23.5	1.5	0.06	2	Zinc binding perturbation

[1] G. Cui, S. Wang, Y. Li, Y. J. Tian, Y. Feng and Q. Cui, *Protein J.*, 2012, **31**, 432-438.

[2] W. Bao, H. Pan, Z. Zhang, Y. Cheng, Z. Xie, J. Zhang and Y. Li, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 1641-1649.

[3] The definition of mutation type: Catalytic, the mutation site is a catalytic residue; Catalytic perturbation, the mutation site directly interacts with the catalytic residues; CES binding, the mutation site is a CES binding residue; CES binding perturbation, the mutation site directly interacts with a CES binding residue; Zinc binding, the mutation site is a zinc binding residue; Zinc binding perturbation, the mutation site directly interacts with zinc binding residues.

SI Figures

Figure S1. CESH[D] molecules in one asymmetric unit (ASU) of the crystal structures. (A) Cartoon representation of four CESH[D] molecules in one asymmetric unit (ASU) of the wild-type CESH[D] crystal structure. (B) The overlay of four molecules in one ASU of the wild-type CESH[D] crystal structure. The zinc ions and coordinative waters were shown as balls. The coordinative residues E14, H47, and H49 were shown as sticks. (C) Cartoon representation of eight molecules in one ASU of the CESH[D]-D115A crystal structure. (D) The overlay of eight molecules in one ASU of the CESH[D]-D115A crystal structure. The zinc ions were shown as balls. The D-TA molecules were shown as sticks.

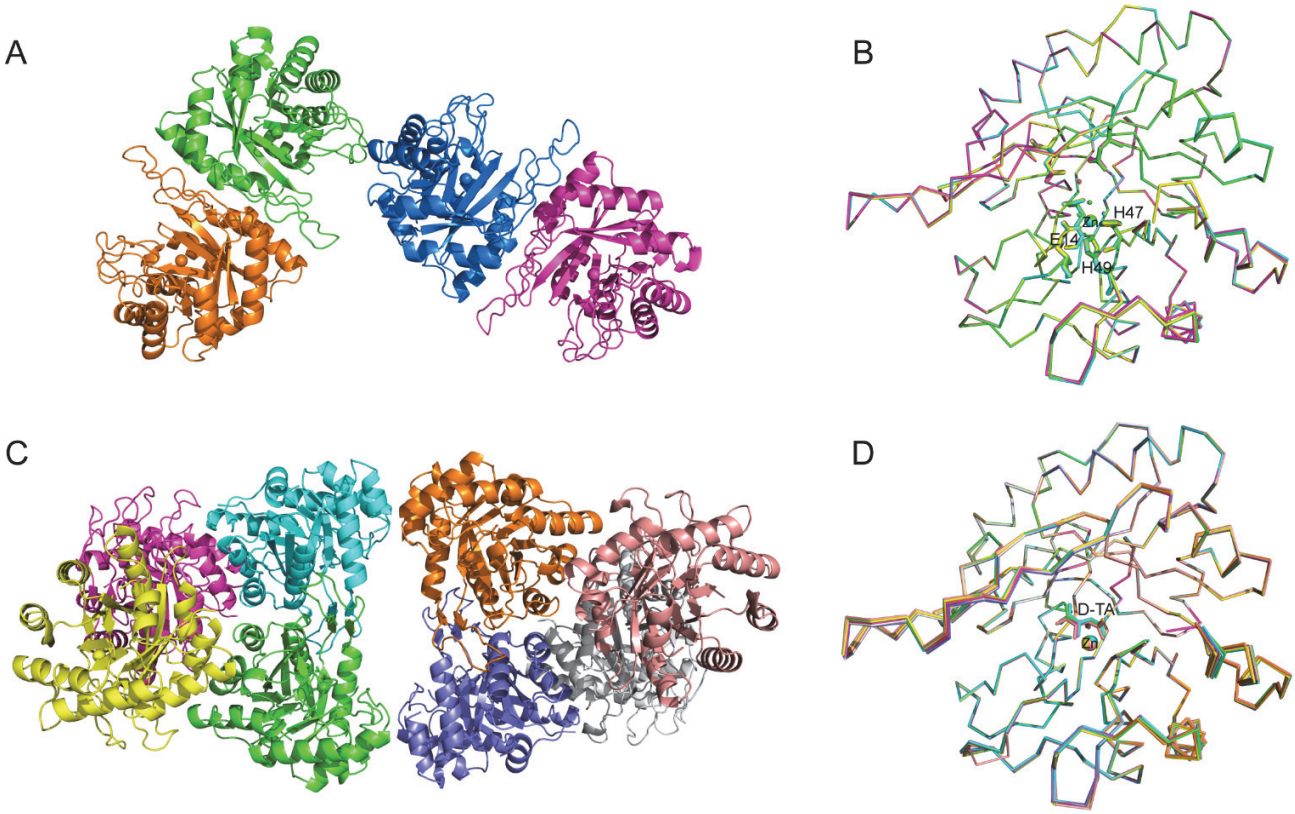

Figure S2. The overall structure of CESH[D]. (A) The surface representation of the CESH[D] dimer. The two monomers are colored in green and orange, respectively. The catalytic cavity is colored in red. Each monomer has ~2000 $Å^2$ buried surface (16% of the total monomer surface). (B) The dimer interface of CESH[D] mainly involve helices α5 and α6, as well as two long loops β4-α4 and β6-α6. The two monomers are represented as surface and cartoon separately. The zinc ion is shown as a cyan ball. The secondary elements of the interface are labeled and with the same color for both monomers. (C) Gel filtration analysis of CESH[D] shows that CESH[D] is a dimer in solution. The molecular mass of CESH[D] is 34.6 kDa. (D) The cartoon structural representation of one CESH[D] molecule. The α-helices, the β-sheets, and the zinc ion are colored in blue, orange, and cyan, respectively.