STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Crystal structure of a novel homodimeric D-allulose 3-epimerase from a Clostridia bacterium

Xiaofang Xie, Yixiong Tian, Xiaofeng Ban, Caiming Li, Hongshun Yang and Zhaofeng Li

Acta Cryst. (2022). D**78**, 1180–1191

STRUCTURAL BIOLOGY

ISSN 2059-7983

Crystal structure of a novel homodimeric D-allulose 3-epimerase from a Clostridia bacterium

Xiaofang Xie,[b,c]‡ Yixiong Tian,[a]‡ Xiaofeng Ban,[a] Caiming Li,[a] Hongshun Yang[b,c]* and Zhaofeng Li[a]*

[a]School of Food Science and Technology, Jiangnan University, Wuxi 214122, People's Republic of China, [b]Department of Food Science and Technology, National University of Singapore, Singapore 117542, Singapore, and [c]National University of Singapore (Suzhou) Research Institute, 377 Lin Quan Street, Suzhou 215123, People's Republic of China. *Correspondence e-mail: fstynghs@nus.edu.sg, zfli@jiangnan.edu.cn

Received 5 April 2022
Accepted 30 July 2022

Edited by Z. S. Derewenda, University of Virginia, USA

‡ These authors contributed equally to this work.

Keywords: D-allulose; ketose 3-epimerases; crystal structure; hydrophobic pocket; Clostridia.

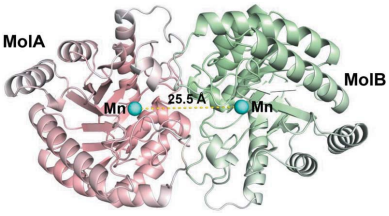
PDB reference: D-allulose 3-epimerase, 7x7w

Supporting information: this article has supporting information at journals.iucr.org/d

D-Allulose, a low-calorie rare sugar with various physiological functions, is mainly produced through the isomerization of D-fructose by ketose 3-epimerases (KEases), which exhibit various substrate specificities. A novel KEase from a Clostridia bacterium (CDAE) was identified to be a D-allulose 3-epimerase and was further characterized as thermostable and metal-dependent. In order to explore its structure–function relationship, the crystal structure of CDAE was determined using X-ray diffraction at 2.10 Å resolution, revealing a homodimeric D-allulose 3-epimerase structure with extensive interactions formed at the dimeric interface that contribute to structure stability. Structural analysis identified the structural features of CDAE, which displays a common $(\beta/\alpha)_8$-TIM barrel and an ordered Mn^{2+}-binding architecture at the active center, which may explain the positive effects of Mn^{2+} on the activity and stability of CDAE. Furthermore, comparison of CDAE and other KEase structures revealed several structural differences, highlighting the remarkable differences in enzyme–substrate binding at the O4, O5 and O6 sites of the bound substrate, which are mainly induced by distinct hydrophobic pockets in the active center. The shape and hydrophobicity of this pocket appear to produce the differences in specificity and affinity for substrates among KEase family enzymes. Exploration of the crystal structure of CDAE provides a better understanding of its structure–function relationship, which might provide a basis for molecular modification of CDAE and further provides a reference for other KEases.

1. Introduction

D-Allulose (also known as D-psicose) is a low-calorie functional rare sugar which is produced through the epimerization of D-fructose at the C-3 position. In recent years, D-allulose has gained increasing interest owing to its safe use as a low-calorie sweetener in food and dietary supplements and its modulation of a variety of physiological functions (Xia *et al.*, 2021; Hu *et al.*, 2021). Given the small quantities present in nature and its difficult chemical synthesis, biosynthesis is becoming a promising alternative strategy for the production of D-allulose. In the biological production of D-allulose, ketose 3-epimerases (KEases) play an irreplaceable role in the conversion of D-fructose to D-allulose. Based on their optimum substrates, KEases are mainly classified as D-tagatose 3-epimerases (DTEases; EC 5.1.3.31), D-allulose 3-epimerases (DAEases; EC 5.1.3.30) and L-ribulose 3-epimerases (LREases; EC 5.1.3.–) (Zhang *et al.*, 2021; Chen, Chen, Ke *et al.*, 2021). Although these three types of KEase have similar characteristics of catalyzing the conversion of D-fructose to D-allulose, DAEases are identified to have the highest activity towards D-allulose and usually exhibit the highest catalytic



efficiency in D-allulose production (Patel *et al.*, 2021). Currently, KEases have been identified and characterized from 30 microbial origins, most of which have limited industrial applications due to poor thermostability and low catalytic efficiency, as shown in Supplementary Table S1 (Zhang *et al.*, 2021). For example, the first identified DAEase from *Agrobacterium tumefaciens* was found to be thermally unstable at 55°C, with a half-life of only 8.9 min (Kim, Hyun *et al.*, 2006). Hence, molecular modification based on the structure–function relationship is of great importance, and many attempts have been made to explore the crystal structures of KEases (Zhu *et al.*, 2021; Zhang *et al.*, 2018). To date, the crystal structures of several KEases have been determined, including those of DAEases from *A. tumefaciens* (Agtu-DAE; PDB entry 2hk1; Kim, Kim *et al.*, 2006) and *Clostridium cellulolyticum* H10 (Clce-DAE; PDB entry 3vnk; Chan *et al.*, 2012), a DTEase from *Pseudomonas cichorii* (Psci-DTE; PDB entry 2qun; Yoshida *et al.*, 2007) and LREases from *Arthrobacter globiformis* (Argl-LRE; PDB entry 5zfs; Yoshida *et al.*, 2018), *Mesorhizobium loti* (Melo-LRE; PDB entry 3vyl; Uechi, Sakuraba *et al.*, 2013) and *Methylomonas* sp. DH-1 (Mesp-LRE; PDB entry 7cj5; Yoshida *et al.*, 2021). Their amino-acid sequence identity varies from 22.7% to 60.9%, but the completely conserved metal-binding and catalytic residues in the active center imply that they belong to the same superfamily and share a similar catalytic mechanism, despite their differing substrate specificities (Zhu *et al.*, 2019). Proteins within the same superfamily are structurally or functionally related and may exhibit low sequence identity, but still share significant structural features in common. Although a catalytic mechanism has been proposed for KEases, to our best knowledge there is a lack of systematic research on the structural features that are responsible for substrate preference among KEases.

In this study, a novel DAEase from a Clostridia bacterium (CDAE) was heterologously expressed in food-grade *Bacillus subtilis* and further biochemically characterized. Here, we determined the crystal structure of CDAE at 2.10 Å resolution and further investigated its structural features. Unlike the homotetrameric associations of Agtu-DAE and Clce-DAE, a homodimeric DAEase structure was revealed by X-ray structural analysis of CDAE. The monomeric structure of CDAE displayed several shared structural features of KEases, including a common $(\beta/\alpha)_8$-TIM barrel, conserved catalytic residues and metal-binding sites. The Mn^{2+} binding in CDAE, which provided a stable catalytic architecture, was responsible for the enhancement of enzyme activity and stability on Mn^{2+} addition. Additionally, extensive interactions formed at the dimeric interface contributed to the structural stability, which might explain the relatively high temperature optimum and the strong thermostability of CDAE. Additionally, several structural differences were revealed by comparison of KEase structures, and the hydrogen-bonding interactions at the O4, O5 and O6 positions of the bound substrate were found to vary among KEases. More importantly, the hydrophobic pocket around the substrate exhibited remarkable differences, which might be involved in substrate recognition at these positions and result in differences in specificity and affinity for substrates between DAEases, DTEases and LREases. Overall, the exploration of the crystal structure of CDAE enabled us to better understand its structure–function relationship, which lays a solid foundation for molecular engineering of CDAE. Our study could also provide a reference for other KEases.

2. Materials and methods

2.1. Protein production and purification

The *dae* gene from a Clostridia bacterium isolate INTA.CYC.091 contig-100_981 (GenBank SABL01000019.1) that encodes a DAEase (CDAE) was used to prepare the expression vector pST/*dae*, which was further transformed into *B. subtilis* WB600 to construct a DAEase expression strain. The recombinant *B. subtilis* WB600 was grown in 50 ml TB broth containing 10 µg ml^{-1} kanamycin at 30°C for 24 h. After harvesting by centrifugation at 4°C and 10 000*g* for 15 min, the cells were resuspended in 50 m*M* Tris–HCl buffer pH 7.5 containing 500 m*M* NaCl and incubated with 1 mg ml^{-1} lysozyme at 37°C for 30 min, followed by ultrasonic disruption for 15 min. The lysate was centrifuged at 10 000*g* for 20 min at 4°C to remove cell debris and the obtained supernatant, referred to as the crude enzyme, was purified by nickel-affinity chromatography using an ÄKTApure protein purification system (GE, USA). The crude enzyme was loaded onto a 5 ml HiTrap HP column and the unbound and unwanted proteins were eliminated with wash buffer (20 m*M* imidazole, 50 m*M* Tris–HCl, 500 m*M* NaCl pH 7.5). Subsequently, the target enzyme was eluted from the column using elution buffer (300 m*M* imidazole, 50 m*M* Tris–HCl, 500 m*M* NaCl pH 7.5). The eluted fraction was dialyzed against 50 m*M* Tris–HCl buffer pH 7.5 containing 10 m*M* ethylenediaminetetraacetic acid (EDTA) at 4°C for 12 h and was then dialyzed three times against EDTA-free Tris–HCl buffer at 4°C. The purity of CDAE was assessed using SDS–PAGE and the molecular-mass distribution of CDAE was analyzed using gel-permeation chromatography. The native purified protein was analyzed by HPLC using a TSKgel G3000SWxl column (Waters, USA) and was eluted using elution buffer (0.1 *M* sodium phosphate, 0.1 *M* sodium sulfate, 0.05% sodium azide pH 6.7) at a flow rate of 0.5 ml min^{-1}. The column was calibrated with Protein Standard Mix (Sigma–Aldrich, USA) including bovine thyroglobulin (670 kDa), γ-globulin (150 kDa), chicken egg albumin grade VI (44.3 kDa) and ribonuclease A type I-A (13.7 kDa).

2.2. Enzyme assay

The enzymatic activity of CDAE was measured according to the method of Li *et al.* (2019) with some modifications. 1 ml of a reaction system consisting of 80 g l^{-1} D-fructose, 1 m*M* $CoCl_2$, 50 m*M* Tris–HCl buffer pH 7.5 and an amount of purified enzyme was incubated at 60°C for 10 min and inactivated by boiling for 10 min. The generated D-allulose was determined using high-performance anion-exchange chromatography with pulsed amperometric detection (HPEAC-PAD;

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

Data collection	
Wavelength (Å)	0.97915
Space group	*I*121
a, *b*, *c* (Å)	60.02, 70.66, 153.78
α, β, γ (°)	90.00, 99.32, 90.00
Resolution range (Å)	37.95–2.10 (2.16–2.10)
Total no. of reflections	245714 (18942)
No. of unique reflections	37135 (2994)
Multiplicity	6.6 (6.3)
Completeness (%)	99.4 (97.8)
R_{merge}	0.032 (0.124)
R_{meas}	0.036 (0.145)
$R_{p.i.m.}$	0.014 (0.056)
$CC_{1/2}$	1.000 (0.996)
$\langle I/\sigma(I)\rangle$	4.54 [at 2.10 Å]
Refinement	
Resolution range (Å)	37.94–2.10
Completeness (%)	98.4
R_{work}/R_{free}	0.200/0.254
Total no. of atoms	4705
Average *B* factor, overall (Å^2)	50.0
R.m.s.d.s	
Bond lengths (Å)	0.007
Bond angles (°)	0.923
Ramachandran plot	
Preferred region (%)	96.33
Allowed region (%)	3.32
Disallowed region (%)	0.35
PDB code	7x7w

Xie *et al.*, 2019). Mixed D-fructose and D-allulose standards (Sigma–Aldrich, USA) were used as a control for qualitative and quantitative analyses. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 µmol of D-allulose per minute.

The influence of temperature on the activity of the enzyme was analyzed at various temperatures in the range 40–85°C. The thermostability was investigated by incubating purified CDAE in 50 m*M* Tris–HCl buffer pH 7.5 at 55–65°C and samples were taken at regular intervals. To calculate the half-life ($t_{1/2}$), a first-order deactivation kinetic model was used (Tseng *et al.*, 2018). To examine the effects of metal ions on the activity of CDAE, the enzyme was treated with various metal ions at 1 m*M* ($CoCl_2$, $MnCl_2$, $NiCl_2$, $CuCl_2$, $MgCl_2$, $FeCl_3$, $ZnCl_2$ and $BaCl_2$) at 4°C for 1 h, followed by activity measurement. The melting temperature (T_m) of CDAE with or without 1 m*M* of the abovementioned metal ions was determined using a differential scanning calorimetry instrument (Nano DSC III, TA Instruments, USA; Chen, Chen, Ke *et al.*, 2021) to determine the effect of metal ions on the structural stability of CDAE. The substrate specificity of CDAE was measured using D-allulose, D-fructose and D-tagatose as substrates. Kinetic parameters were determined using D-allulose and D-fructose as substrates and were calculated by nonlinear regression (Michaelis–Menten model) using *GraphPad Prism*.

2.3. X-ray crystallography

Purified CDAE in ultrapure water was subjected to crystallization experiments using the hanging-drop vapor-diffusion method. A high-quality protein crystal was obtained at 20°C within a seven-day incubation period using equivalents of protein solution (30 mg ml^{-1}) and reservoir solution [0.1 *M* succinic acid pH 7.0, 0.1 *M* HEPES pH 7.0, 0.1%(*w*/*v*) PEG MME 2000]. Crystals were harvested into liquid nitrogen after soaking in cryoprotectant prepared using reservoir solution, 0.1 *M* $MnCl_2$ and 20% glycerol. The data set was collected on beamline BL19U1 at Shanghai Synchrotron Radiation Facility.

X-ray diffraction data were collected at a resolution of 2.10 Å, and the space group was determined to be *I*121. The diffraction data were processed using *HKL*-2000 (Otwinowski & Minor, 1997) and the *CCP*4 program suite (Winn *et al.*, 2011). The molecular-replacement method was used to determine the initial phases of CDAE using the structure of Clce-DAE (PDB entry 3vnk) as the search model. Model-building and structure-refinement procedures were performed with *Coot* (Emsley *et al.*, 2010) and *Phenix* (Liebschner *et al.*, 2019). Data-collection and refinement statistics are summarized in Table 1. Structural information has been uploaded to the wwPDB (http://www.wwpdb.org) with accession code 7x7w. The interfaces were analyzed by *PISA* in *CCP*4. Molecular graphics were created using *PyMOL* (DeLano, 2002).

2.4. Molecular docking

A molecular-docking simulation was performed using *AutoDock Vina* (Trott & Olson, 2010) to explore the interactions between CDAE and various ketohexoses, including D-allulose, D-fructose and D-tagatose. CDAE (PDB entry 7x7w) was taken as the receptor and the ketohexoses were taken as the ligands; their 3D structures were obtained using *eLBOW* in *Phenix*. Grid number points (*xyz* 40, 40, 40) were established as the activity site for docking, with the center at (*xyz* 36, 29, 88).

3. Results and discussion

3.1. Sequence analysis

The discovery of novel KEases is a key factor in improving D-allulose production. To identify the sugar phosphate epimerase from a Clostridia bacterium, phylogenetic analysis was performed with the reported KEase family (Fig. 1). The polygenetic tree of the KEase family obtained by the neighbor-joining method (Wang *et al.*, 2021) indicated a close relationship of CDAE to other DAEases; in particular, an evolutionary relationship between CDAE and the DAEases from *Clostridium cellulolyticum* H10 (Clce-DAE; GenBank ACL75304.1; Mu *et al.*, 2011) and *Clostridium* sp. BNL1100 (Clsp-DAE; GenBank WP_014314767.1; Mu *et al.*, 2013). Similarly, the amino-acid identity also showed that CDAE had a much higher amino-acid identity to most DAEases from other sources (>40%) than to DTEases and LREases, and exhibited the highest identity to Clce-DAE and Clsp-DAE, reaching 70%. In general, the same types of KEases were closely related to each other, while differing types of KEases manifested relatively low identities (Chen, Chen, Liu *et al.*,

2021). Notably, the DAEases from *Pirellula* sp. SH-Sr6A (Pisp-DAE; Li, Li *et al.*, 2021), *Sinorhizobium fredii* (Sifr-DAE; Li, Zhang *et al.*, 2021) and *Rhodopirellula baltica* (Rhba-DAE; Mao *et al.*, 2020), as well as the LREase from *Thermotoga maritima* (Thma-LRE; Shin *et al.*, 2017), displayed low sequence identity to all other KEases (<30%). Overall, the polygenetic relationship and amino-acid identity revealed that CDAE was closely related to the DAEases but was relatively distant from the DTEases and LREases, which to some extent implies differences in the structures and catalytic functions of the different types of KEases.

Despite the various types of KEases showing considerable differences in amino-acid sequence, a structure-based amino-acid sequence alignment of the KEase family revealed several conserved residues (Supplementary Fig. S1). In addition to the strictly conserved residues involved in metal binding (Glu150, Asp183, His209 and Glu244 in Clce-DAE; PDB entry 3vnk), the substrate-binding residues Glu156, His186 and Arg215 that anchor the O1, O2 and O3 sites of ketose substrates in Clce-DAE were also highly conserved in other KEases, except for Rhba-DAE, which shared very low identity with all other KEases (less than 23%), indicating a similar structure of the catalytic center among the KEase family (Zhang *et al.*, 2016). However, the residues that participate in binding the O4, O5 and O6 sites of ketose in Clce-DAE (residues 6, 14, 66, 107, 112 and 246; Li *et al.*, 2019) were entirely distinct in the various KEases. In addition, the sequence alignment also highlighted the longer length of the C-terminus of LREase from *Mesorhizobium loti* (Melo-LRE; Uechi, Takata*et al.*, 2013) and the shorter C-terminus of Mesp-LRE among the KEases, implying a size difference in the C-terminal helix (α8), which was proposed to influence structure stability. Overall, the above-mentioned analysis suggested that the recombinant KEase from a Clostridia bacterium belongs to the DAEases, which

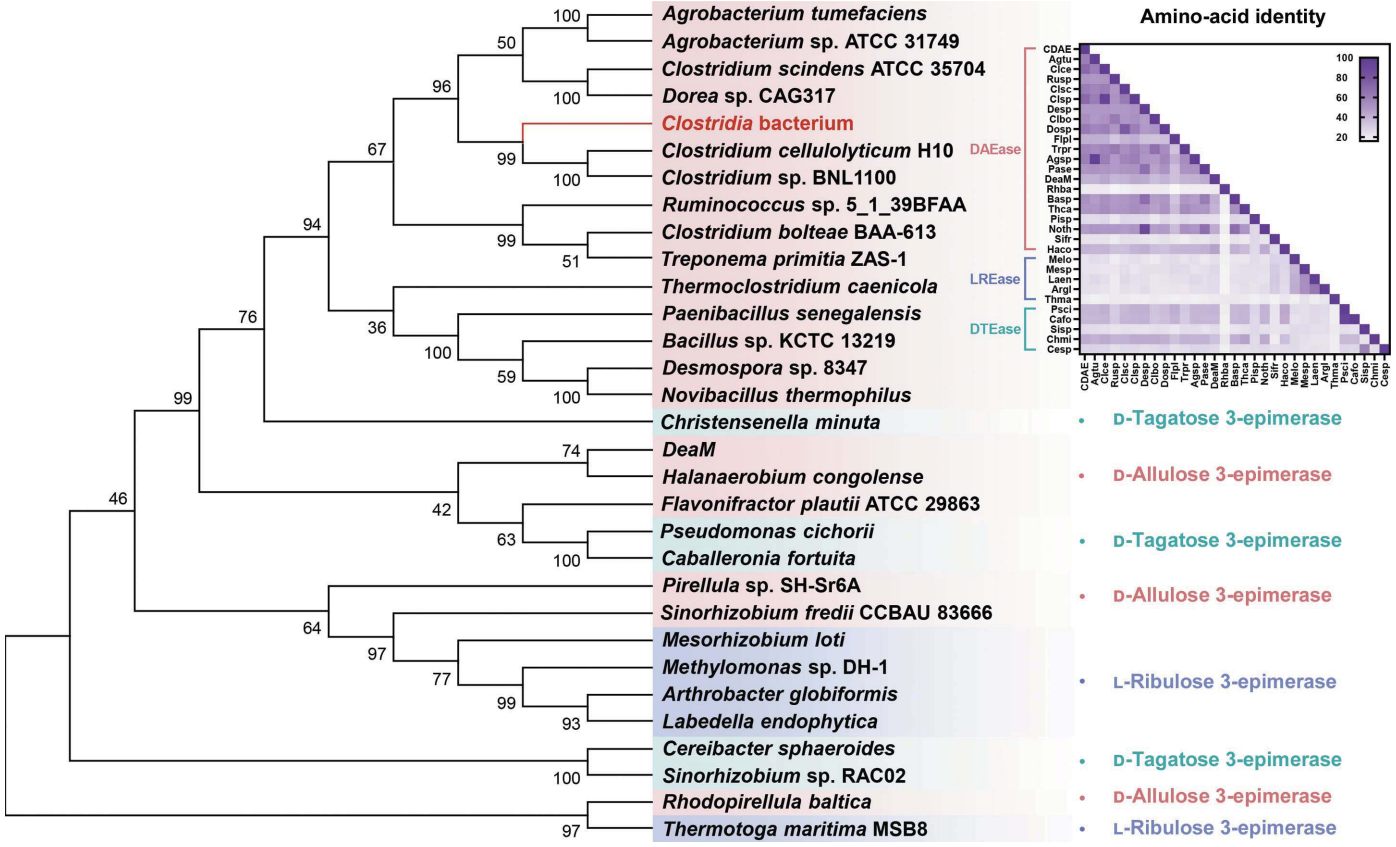


Figure 1
Phylogenetic relationships and amino-acid identities in the KEase family. The amino-acid sequence sources were *Agrobacterium tumefaciens* (Agtu; GenBank AAK88700.1), *Clostridium cellulolyticum* H10 (Clce; ACL75304.1), *Ruminococcus* sp. 5_1_39BFAA (Rusp; ZP_04858451), *Clostridium scindens* ATCC 35704 (Clsc; EDS06411.1), *Clostridium* sp. BNL1100 (Clsp; WP_014314767.1), *Desmospora* sp. 8347 (Desp; WP_009711885.1), *Clostridium bolteae* BAA-613 (Clbo; EDP19602.1), *Dorea* sp. CAG317 (Dosp; WP_022318236.1), *Flavonifractor plautii* ATCC 29863 (Flpl; EHM40452.1), *Treponema primitia* ZAS-1 (Trpr; ZP_09717154.1), *Agrobacterium* sp. ATCC 31749 (Agsp; EGL65884.1), *Paenibacillus senegalensis* (Pase; WP_010270828.1), DeaM from an uncultured bacterium (QHD25651.1), *Rhodopirellula baltica* (Rhba; WP_007330622.1), *Bacillus* sp. KCTC 13219 (Basp; KYG89858.1), *Thermoclostridium caenicola* (Thca; SHI77623.1), *Pirellula* sp. SH-Sr6A (Pisp; WP_146677337.1), *Novibacillus thermophilus* (Noth; WP_077721022.1), *Sinorhizobium fredii* CCBAU 83666 (Sifr; ASY72161.1), *Halanaerobium congolense* (Haco; WP_110301365.1), *Mesorhizobium loti* (Melo; BAB50456.1), *Methylomonas* sp. DH-1 (Mesp; WP_064020855.1), *Labedella endophytica* (Laen; WP_127049469.1), *Arthrobacter globiformis* (Argl; BAW27657.1), *Thermotoga maritima* MSB8 (Thma; AAD35501.1), *Pseudomonas cichorii* (Psci; BAA24429.1), *Caballeronia fortuita* (Cafo; WP_061137998.1), *Sinorhizobium* sp. RAC02 (Sifr; AOF93213.1), *Christensenella minuta* (Chmi; WP_066519968.1) and *Cereibacter sphaeroides* (Cesp; ACO59490.1).

was further supported by substrate-specificity analysis, as described later.

3.2. Overall structure of CDAE

Among DAEses that have been identified and characterized from 20 sources, the structures of Agtu-DAE and Clce-DAE have been determined, and they both assemble into tetramers (Kim, Kim *et al.*, 2006; Chan *et al.*, 2012). In contrast to the homotetrameric DAEases, *PISA* analysis of CDAE suggested a stable homodimeric structure that consists of molecule *A* (MolA) and molecule *B* (MolB) (Fig. 2*a*), which is consistent with the result of gel-permeation chromatography, in which a single peak corresponding to a dimer was observed (Fig. 3*a*). The recombinant CDAE was predicted to be 33.14 kDa in size and ran as a monomer on a reducing SDS–PAGE gel. The

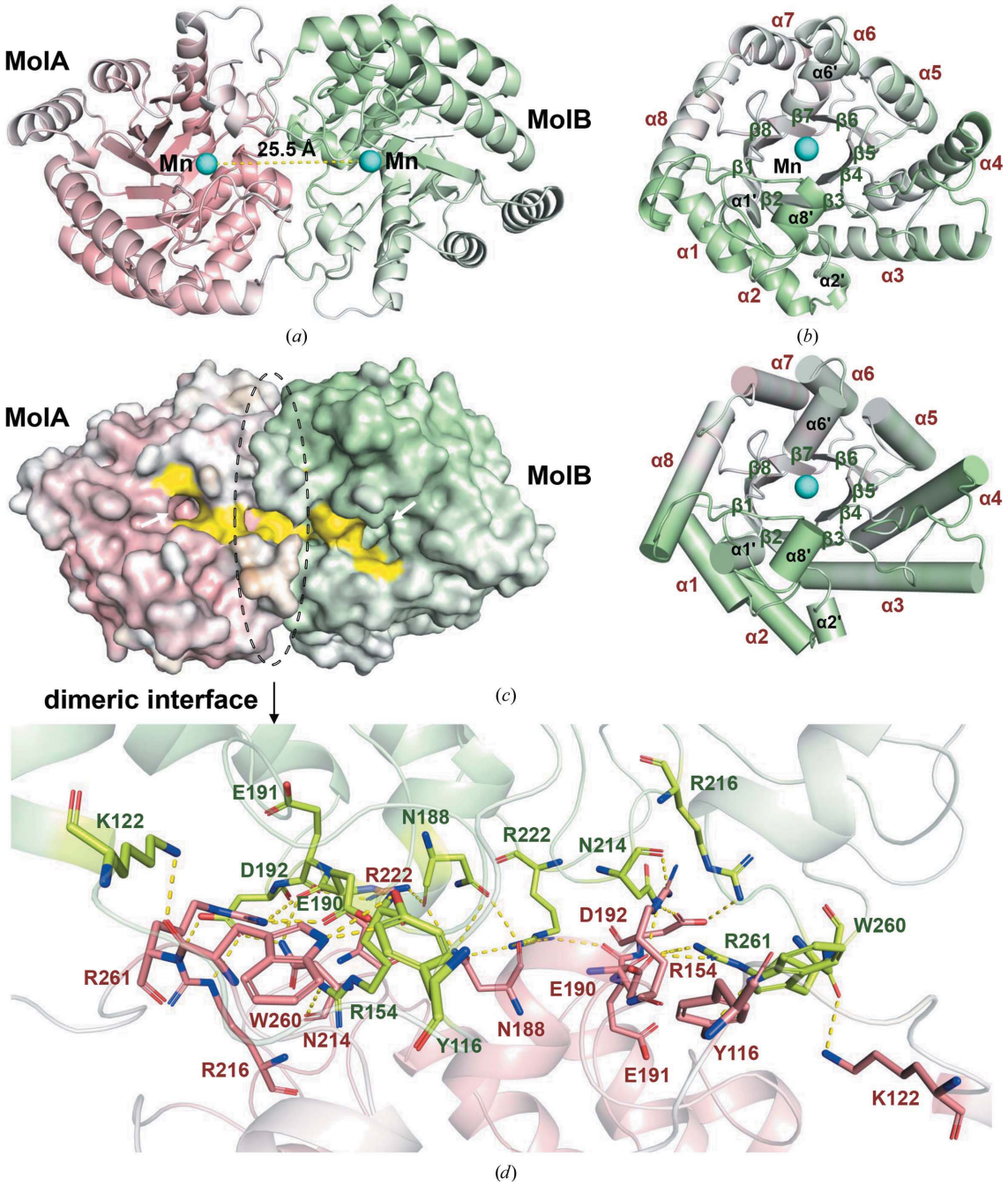


Figure 2
Overall structure of CDAE. (*a*) Biological assembly of the CDAE dimer. Subunits MolA (salmon) and MolB (green) are shown in cartoon representation, and the manganese ions in MolA and MolB are shown as spheres with their distance labeled. (*b*) The monomer structure of CDAE showing a TIM barrel with the α-helices and β-strands labeled. (*c*) The surface structure of dimeric CDAE. Arrows indicate the active center, and the hydrophobic groove to the active center is shown in yellow. (*d*) Interface interactions between MolA and MolB. Residues in MolA and MolB are shown as sticks colored salmon and green, respectively. Hydrogen bonds are shown as yellow dashed lines.

molecular mass of CDAE in solution was estimated to be 69.27 kDa by gel-permeation chromatography, suggesting that CDAE assembles as a dimer under nondenaturing conditions (Sim *et al.*, 2014). The subunit structure (MolA or MolB) of CDAE contains 12 α-helices and eight β-strands, adopting an $(\alpha/\beta)_8$-TIM barrel with four additional short helical segments before α1, α2, α6 and α8 (Fig. 2*b*). A manganese ion is bound at the center of the $(\alpha/\beta)_8$-barrel and is surrounded by two catalytic residues, Glu150 and Glu244, and plays a vital role in the catalytic process of CDAE. Since MolA and MolB are almost identical in structure, the structure description concentrates on MolA of CDAE.

MolA and MolB in CDAE are packed against each other by the closed sides of the barrel and expose the active center on the same side of the dimer. A hydrophobic groove leading to the active center is formed by Trp14, Gly67, Trp112, Pro113, Val114 and Phe155, providing a favorable accessible surface for substrate binding (Fig. 2*c*). The buried area of the dimer is 11.6% of the total surface area. MolA and MolB contact each other tightly and generate a multitude of interactions at the dimeric interface, with an interface solvent-accessible area of 1388.5 Å^2, covering 37 residues at the interface. 11 residues in the loop regions are involved in interface interactions by forming 32 direct hydrogen bonds between subunits (Fig. 2*d*). The salt-bridge network of the CDAE interface is created by four residues from MolA and MolB: Glu190, Arg261, Asp192 and Arg216. In addition, some hydrophobic interactions contribute to stabilizing the interface structure, including Phe155, Phe185, Ile189 and Trp260. It is worth mentioning that the side chains of Phe155 from each subunit face towards each other within a distance of 4 Å, and the resulting hydrophobic stacking helps to maintain the stability of the loop

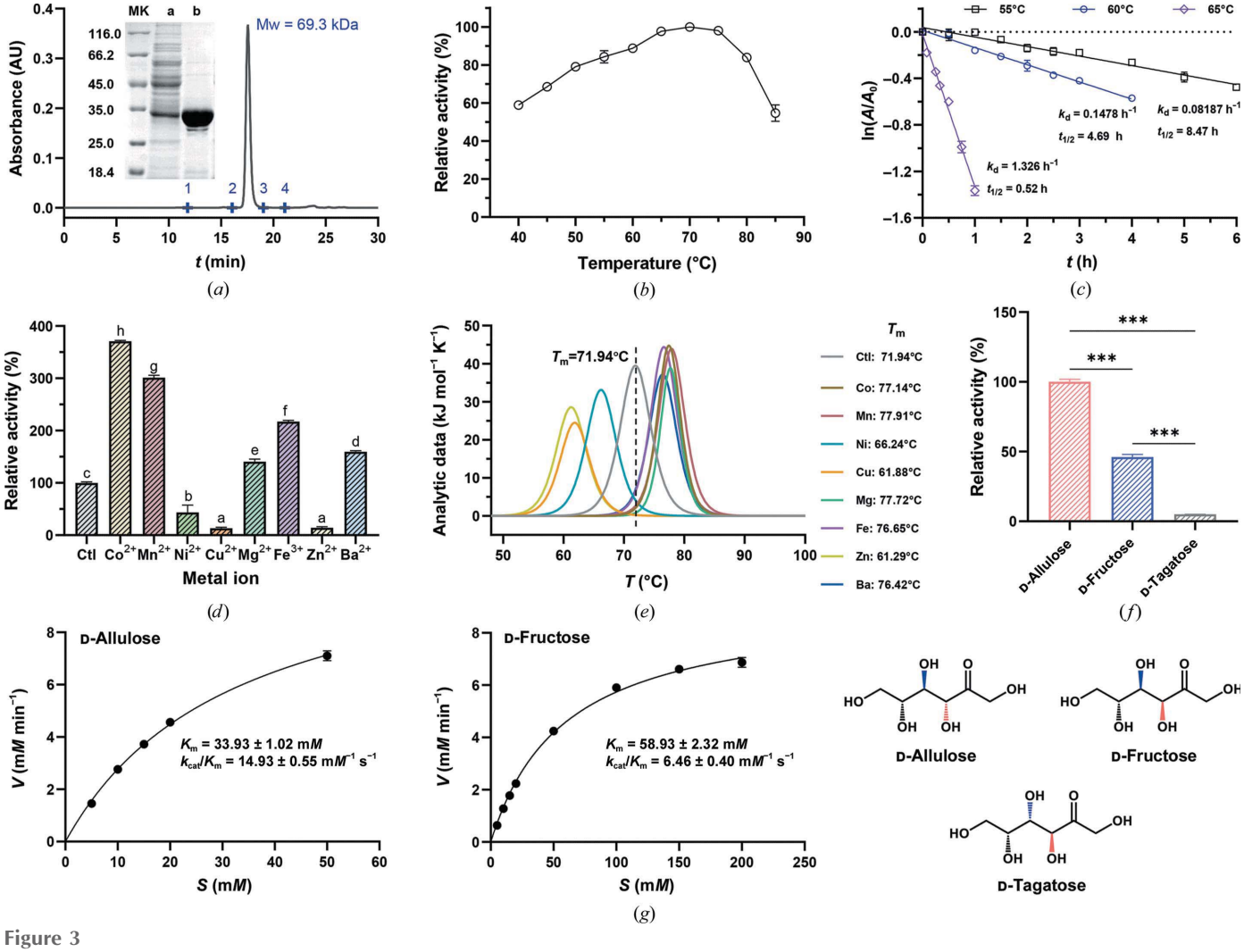


Figure 3
Biochemical characterization of CDAE. (*a*) Gel-permeation chromatography and SDS–PAGE gel (inset) of purified CDAE. Lane MK, molecular-weight marker; lane a, crude enzyme; lane b, purified CDAE. 1–4 represent the reference proteins. 1, bovine thyroglobulin (670 kDa); 2, γ-globulin (150 kDa); 3, chicken egg albumin grade VI (44.3 kDa); 4, ribonuclease A type I-A (13.7 kDa). The single peak that appeared at 17.6 min corresponds to a molecular weight of 69 269 Da. (*b*) Effect of temperature on the enzymatic activity of recombinant CDAE. (*c*) Thermostability of CDAE at 55, 60 and 65°C. A/A_0, residual activity; k_d, inactivation rate constant; $t_{1/2}$, half-life. (*d*) Effect of metal ions on the activity of CDAE. Ctl, CDAE without metal-ion addition; the relative activity was taken as 100%. (*e*) Effect of metal ions on the structural stability of CDAE. (*f*) Substrate specificity of CDAE. (*g*) Kinetic properties of CDAE.

structure between β5 and α5. Overall, the extensive interactions at the dimeric interface contribute to maintaining the relatively ordered loop regions and stabilizing the enzyme structure.

The homotetrameric structure of Agtu-DAE or Clce-DAE is assembled by two dimers (MolA–MolD and MolB–MolC), and tight contacts are formed at the dimeric interface. The monomer and dimer structures of these three DAEases were similar, and the calculated intersubunit contact area of CDAE (1388.5 Å^2) was similar to those of Clce-DAE (1407.4 Å^2) and Agtu-DAE (1383.1 Å^2). However, some differences were also observed at the dimeric interface. *PISA* analysis showed that 12 salt bridges were formed at the dimeric interface of CDAE, exceeding the eight found for Agtu-DAE or Clce-DAE (forming a total of 30 hydrogen bonds). Interestingly, previous studies have shown that the optimal temperatures of Agtu-DAE and Clce-DAE were 50°C and 55°C, respectively. Agtu-DAE lost half of its activity after incubation for only 8.9 min at 55°C, compared with a half-life of 24 min at 55°C for Clce-DAE. In contrast, CDAE exhibited its highest activity at 70°C (Fig. 3*b*). Investigation of its thermostability (Fig. 3*c*) indicated that CDAE exhibited strong thermostability at 55°C and 60°C, with half-lives of 8.47 and 4.69 h, respectively, despite its relatively weak thermostability at 65°C ($t_{1/2}$ = 0.52 h). The stronger interactions in the dimerization of CDAE may contribute to its increased thermostability and optimum temperature (Yoshida *et al.*, 2007).

3.3. Active-site structure of CDAE

The detailed active-site structure of CDAE (Fig. 4) shows the active sites located at the center of the $(\alpha/\beta)_8$-TIM barrel, involving metal-binding sites, catalytic residues and substrate-binding sites. The metal-binding sites and catalytic residues are conserved among DTEases, DAEases and LREases, implying that they belong to the same superfamily despite their different substrate specificities, which might be attributed to the differences in their substrate-binding networks. In the active site of CDAE, the predicted catalytic residues Glu150 and Glu244 are situated on two sides of the manganese ion.

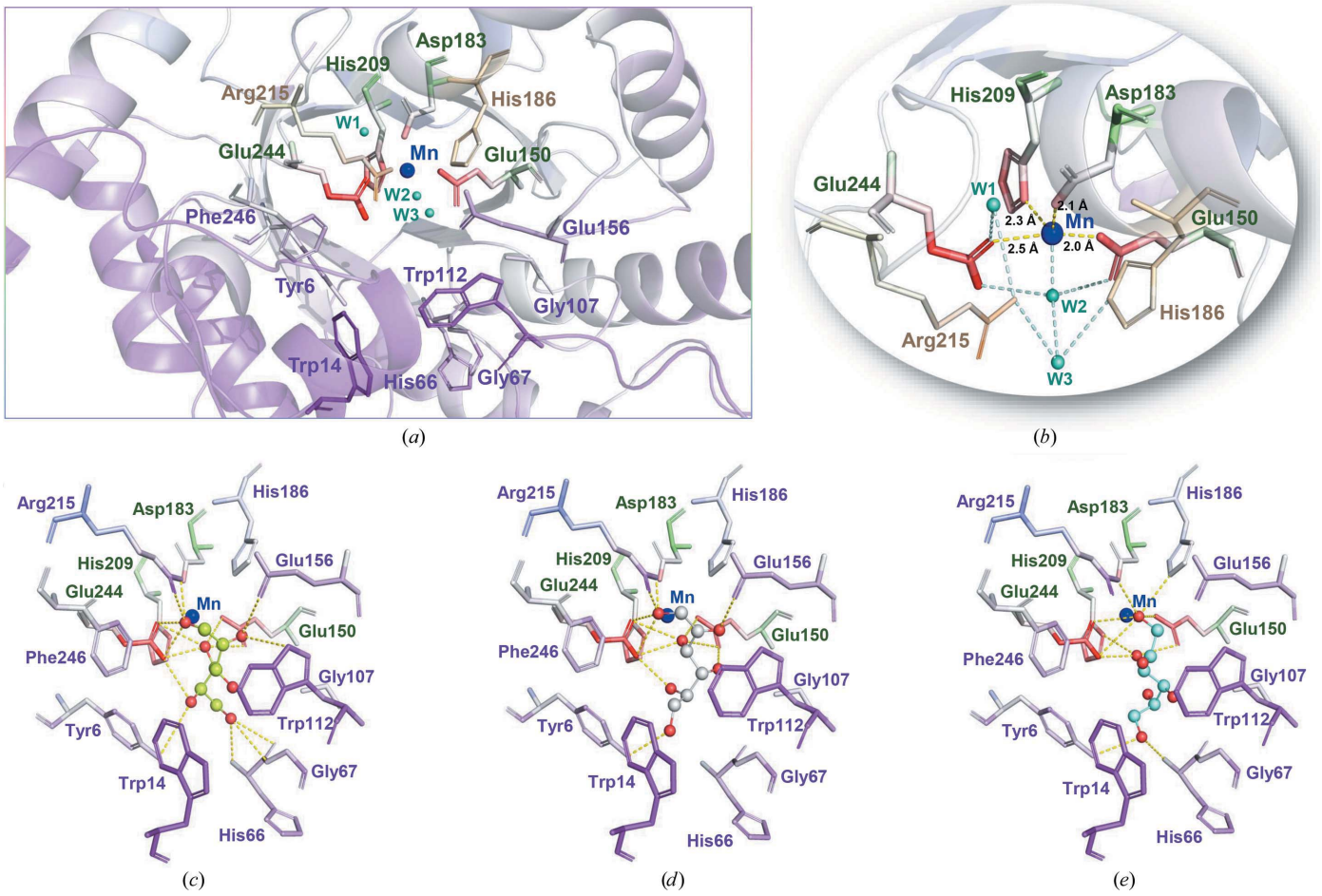


Figure 4
Active-site structure of CDAE. (*a*) Active center of CDAE. (*b*) Enlargement of the metal-binding structure. (*c*, *d*, *e*) The hydrogen-bonding networks of complex structures between CDAE and D-allulose (*c*), D-fructose (*d*) and D-tagatose (*e*). The bound substrates were generated by molecular docking with Glu156, His186 and Arg215 as the flexible residues. Active-site residues of CDAE are shown in stick representation. D-Allulose, D-fructose and D-tagatose are shown in stick and sphere representation, with their C atoms in green, gray and cyan, respectively. Mn^{2+} ions and water molecules (W1, W2 and W3) are shown as blue and cyan spheres, respectively.

The manganese ion makes coordination bonds to four residues, and is coordinated by Glu150 OE2 (2.0 Å), Asp183 OD2 (2.1 Å), His209 ND1 (2.3 Å) and Glu244 OE1 (2.5 Å). Moreover, three water molecules (W1, W2 and W3) are also observed in the Mn^{2+}-bound structure and W2 in the center is linked directly to Mn^{2+} (Fig. 4*b*). A water-mediated hydrogen-bonding network is formed through interactions with Glu150, His186, Arg215 and Glu244. Obviously, such an Mn^{2+} coordination network stabilizes the active site. In the KEase family metal ions could affect the hydrogen-bond network involved in the catalytic process (Zhang *et al.*, 2021), acting as cofactors to enhance the activity of KEase family enzymes. To explore the effect of metal ions on CDAE, the activity of CDAE was measured in the presence of various metal ions. The results showed that CDAE was active even without metal ions, but was strongly reinforced by the addition of Co^{2+} and Mn^{2+} (Fig. 3*d*), followed by Fe^{2+}, Ba^{2+} and Mg^{2+}. In contrast, Ni^{2+}, Cu^{2+} and Zn^{2+} decreased the activity of CDAE to a large extent. Similarly, Co^{2+}, Mn^{2+}, Mg^{2+}, Fe^{2+} and Ba^{2+} were found to enhance the structural stability of CDAE, with increased T_m values (Fig. 3*e*), while Ni^{2+}, Cu^{2+} and Zn^{2+} negatively impacted the stability. The enhancement of the activity and stability of CDAE by Mn^{2+} can be better understood through the Mn^{2+} coordination in the CDAE structure, which provides a more stable catalytic environment with a more ordered active site.

Together with the structural analysis, it was found that CDAE showed the highest activity towards D-allulose (Fig. 3*f*) followed by D-fructose (46.04% of that towards D-allulose), and had very low relative activity towards D-tagatose (5.03%). In addition, the affinity of CDAE for D-allulose was higher than that for D-fructose (Fig. 3*g*). To account for the sequential decrease in catalytic activity towards D-allulose, D-fructose and D-tagatose, as well as the higher affinity of CDAE for D-allulose than for D-fructose, molecular docking was performed to explore the interactions between CDAE and these ketohexoses (Jia *et al.*, 2021). The hydrogen-bonding

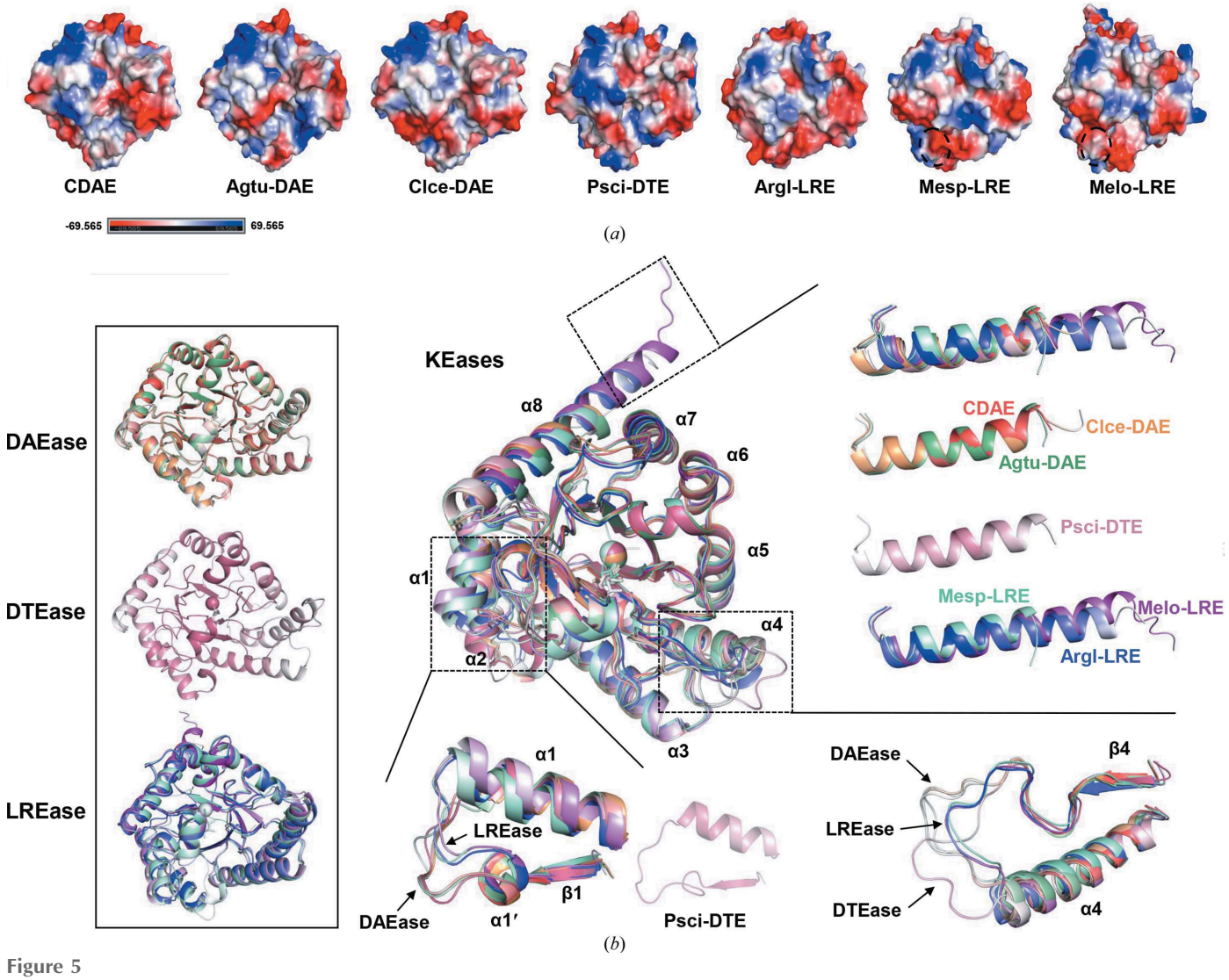


Figure 5
Overall structure comparison of the KEase family. (*a*) Surface models of KEases viewed from the same direction are shown with their charge distributions. (*b*) Superposition of the monomeric structures of KEases, including CDAE (red), Agtu-DAE (green), Clce-DAE (orange), Psci-DTE (pink), Argl-LRE (blue), Mesp-LRE (turquoise) and Melo-LRE (purple).

networks in these three complex structures (Figs. 4*c*–4*e*) elucidate the differences in enzyme–substrate interactions; more hydrogen bonds are formed between D-allulose and the enzyme, while the fewest are formed for D-tagatose. When binding D-allulose or D-fructose, both Glu150 and Glu244 each donate two direct hydrogens at the O3 position of the bound substrate, whereas only a total of two hydrogen bonds are formed between these two catalytic residues and D-tagatose at this position. In addition, the spatial conformation of the substrate was thought to be determined by residues encompassing the O4, O5 and O6 sites of the substrate in the KEase family. For this reason, emphasis was placed on the hydrogen bonds formed at the O4, O5 and O6 sites of each ketohexose (Qi *et al.*, 2017), which were found to gradually decrease in D-allulose, D-fructose and D-tagatose. The O4, O5 and O6 sites of D-allulose make five hydrogen bonds to nearby residues, while three are formed in D-fructose and only two in D-tagatose. More hydrogen bonds promote substrate orientation and limit its motion, and this might explain why a higher catalytic activity and affinity towards D-allulose was observed (Maskeri *et al.*, 2020). In addition, it was observed that the hydrophobic pocket formed by Trp14, Gly107, Trp112 and Phe246 around the substrate in CDAE was closer to bound D-allulose than D-tagatose; this also seems to participate in substrate recognition and affect the substrate preference.

3.4. Structural comparison of KEases

Among the KEases with determined structures, CDAE, Psci-DTE (PDB entry 2qun) and Mesp-LRE (PDB entry 7cj5) exhibit homodimeric structures, while others form a homotetramer with twofold symmetry, including Agtu-DAE (PDB entry 2hk1), Clce-DAE (PDB entry 3vnk), Melo-LRE (PDB entry 3vyl) and Argl-LRE (PDB entry 5zfs). However, all of them display a similar dimeric association in which two subunits interact in a tightly linked manner (Supplementary Fig. S2). A structural comparison of monomers of these KEases was conducted to understand the structural features and differences between DAEases, DTEases and LREases with different substrate preferences. The monomer surface model indicates their overall difference in structure. In addition to a common $(\alpha/\beta)_8$-TIM barrel and a conserved metal-binding structure, the superposition of KEase subunits also reveals some remarkable structural differences (Fig. 5). The loop region between $\beta4$ and $\alpha4$ of Agtu-DAE (107–120) that serves as a lid over the active site (Kim, Kim *et al.*, 2006) is

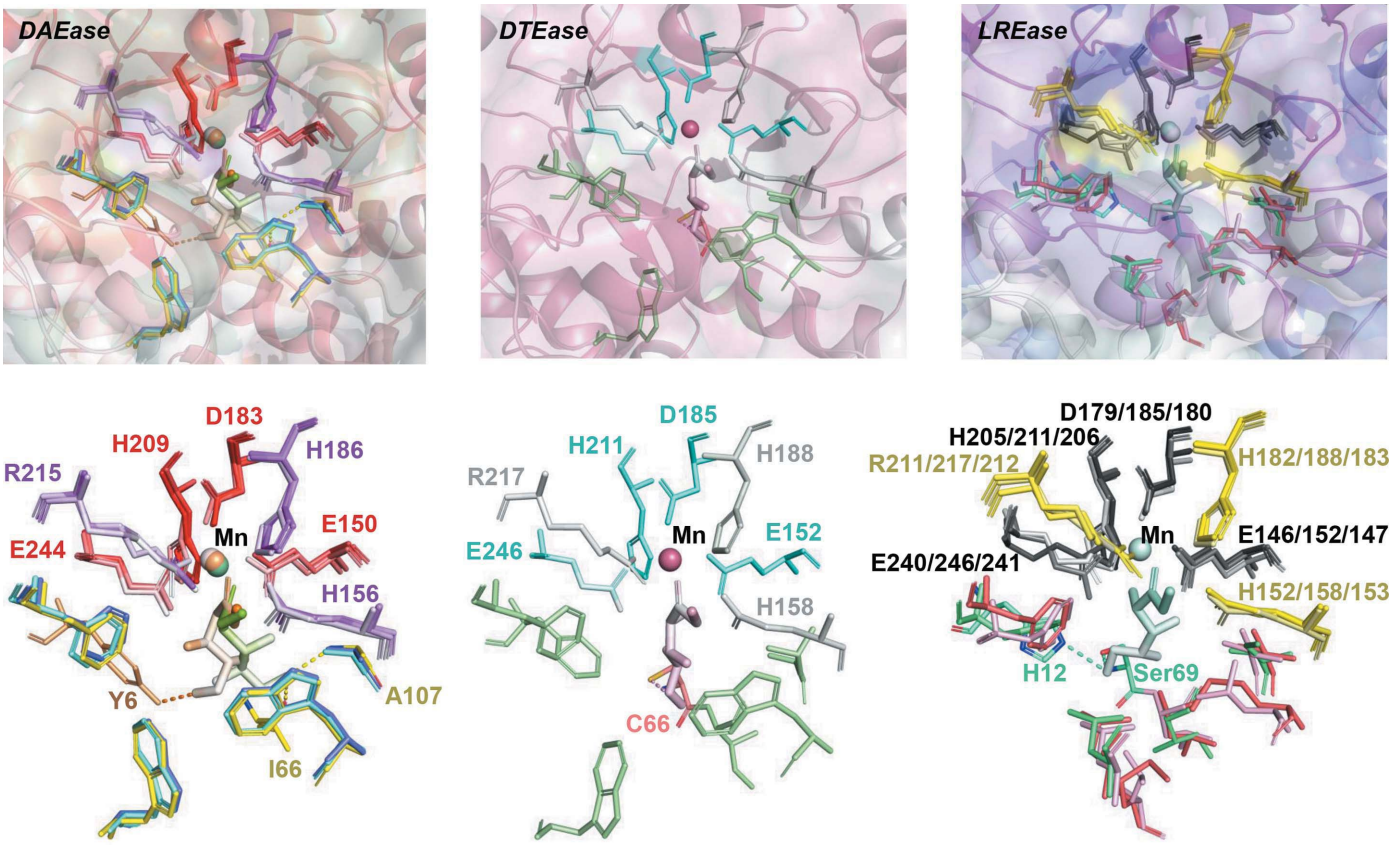


Figure 6
Active-site structure comparison of the KEase family. Metal-binding residues are shown as red sticks in DAEase, cyan sticks in DTEase and black sticks in LREase. Residues making hydrogen bonds to D-fructose at its O1, O2 and O3 sites are shown in purple in DAEase, gray in DTEase and yellow in LREase. Residues making hydrogen bonds to D-fructose at its O4, O5 and O6 sites involve Ile66 and Ala107 in Agtu-DAE, Tyr6 in Clce-DAE, Cys66 in Psci-DTE, His12 and Ser69 in Mesp-LRE and remain unknown in other KEases due to the lack of a complex structure with the substrate. Hydrophobic residues in the active center are also shown as sticks in various colors.

quite similar in DAEases, seems to be shorter in LREases and is completely different in DTEases. This region provides stronger hydrophobic interactions in the interface area of Psci-DTE, where Pro117 is sandwiched by Trp160 (MolA) and Trp262 (MolB) (Yoshida *et al.*, 2007). The longer loop in DAEases adopts a flexible structure that is favorable for conformational change, with Trp112 moving towards the bound substrate (Chan *et al.*, 2012). Besides, the loop region between $\alpha1'$ and $\alpha1$ in DAEase is closer to the active site, with the indolyl moiety of Trp14 oriented towards the bound substrate. Notably, these two segments of DAEase shift towards the entrance to the active center compared with the corresponding regions in DTEases and LREases. The structure comparison also highlights the difference in helix $\alpha8$, which is shortest in Mesp-LRE (Leu274–Tyr286) and longest in Melo-LRE (Asn266–Ser292), while DAEases and DTEases have $\alpha8$ helices of the same length. The longer C-terminal helix with an additional tail in Melo-LRE promotes the generation of additional inter-subunit interactions, thereby increasing the stability of the enzyme (Uechi, Sakuraba *et al.*, 2013).

The active-site structures of DAEase, DTEase and LREase bound to D-fructose (Fig. 6) display an Mn^{2+} ion in an octahedral coordination with four conserved residues (two glutamic acids, one histidine and one aspartic acid) and bound D-fructose (carbonyl oxygen O2 and hydroxyl group O3). Focusing on those residues proximal to the bound D-fructose, it was found that strong enzyme–substrate interactions were formed at the O1, O2 and O3 positions of the bound substrate. In addition to two catalytic residues, three strictly conserved residues located in the vicinity of the metal-binding sites participate in making extensive hydrogen bonds with D-fructose at the O1, O2 and O3 sites, including Glu156 with a negative charge and His186 and Arg215 with a positive charge in DAEase, and corresponding residues in DTEase and LREase. In contrast, a relatively weak interaction occurs at O4, O5 and O6 of the substrate, where the residues involved in hydrogen bonding vary considerably among KEases (Supplementary Table S2). The highly conserved active-site structure involving metal coordination and key residues anchoring the O1, O2 and O3 sites of the substrate imply that KEases share a similar catalytic mechanism, catalyzing C3

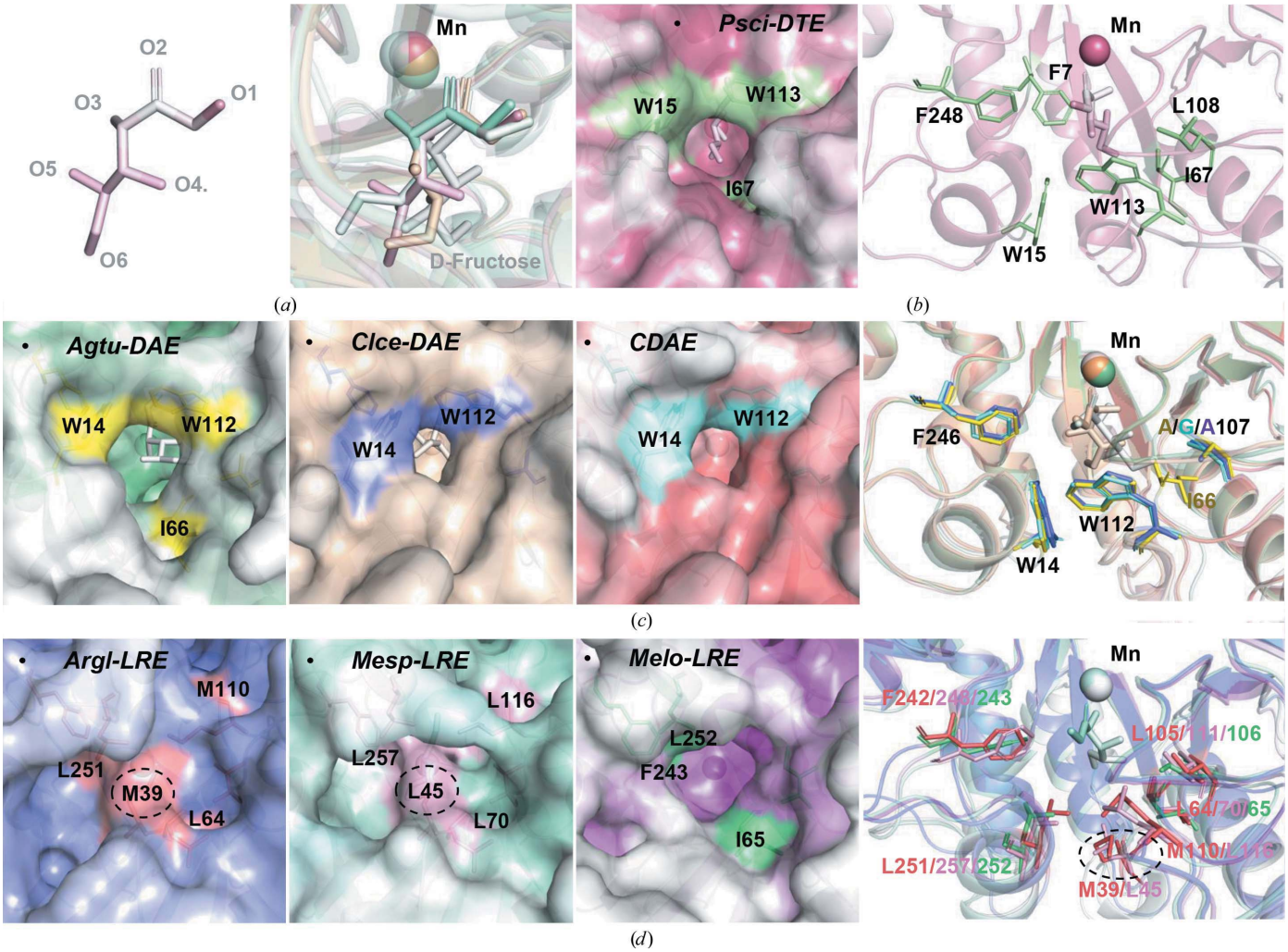


Figure 7
Hydrophobic pocket comparison in the KEase family. (*a*) Superposition of the bound D-fructose at the active center of various KEases. (*b*, *c*, *d*) A close-up view of the hydrophobic pocket at the active center and the channel to the catalytic site are illustrated for DTEase (*b*), DAEases (*c*) and LREases (*d*).

epimerization in a similar manner (Li *et al.*, 2019). Upon substrate binding, a proton is removed from C3 by a glutamic acid coordinated to Mn^{2+} (Glu150 or Glu244 in DAEase) to generate a *cis*-enediolate intermediate with an O2—C2—C3—O3 plane structure; the other catalytic residue subsequently protonates C3 on the opposite side (Kim, Kim *et al.*, 2006; Yoshida *et al.*, 2007; Chan *et al.*, 2012).

Although KEases catalyze epimerization in a similar manner, the significant difference in the hydrophobic environment around the catalytic core may result in the different substrate preferences of DAEases, DTEases and LREases. The structural superposition shows that D-fructose bound in various KEases in a linear form overlaps well at the O1, O2 and O3 positions, whereas it exhibits distinct conformational differences at the O4, O5 and O6 sites (Fig. 7*a*). The few hydrogen bonds at the O4, O5 and O6 sites provide a loose recognition of the substrate. Notably, the hydrophobic pockets surrounding the O4, O5 and O6 positions of D-fructose are markedly different in DAEases, DTEases and LREases, as suggested in the detailed hydrophobic environment architecture depicted in Fig. 7. DAEases have similar hydrophobic pockets formed by Trp14, Gly/Ala107, Trp112 and Phe246, apart from the additional Ile66 in the pocket of Agtu-DAE, whereas the corresponding residue His66 in CDAE and Clce-DAE creates steric hindrance in the channel to the catalytic site. D-Fructose enters the catalytic site through a narrow channel in a linear fashion, enabling the isomerization process in an open-chain state. Notably, Ile66 in Agtu-DAE provides more entry space for the substrate than the corresponding Ile67 in Psci-DTE due to a different rotation angle of the isoleucine side chain. Additionally, the side chain of Leu108 (Gly107 in CDAE) being directed towards the bound substrate, the additional hydrophobic residue Phe7 and the shorter distances of the indole moieties of Trp15 and Trp113 to the substrate (Trp14 and Trp112 in CDAE) generate a hydrophobic pocket with smaller volume. In comparison, the pocket of LREase differs substantially from those of DAEase and DTEase in terms of shape and hydrophobicity. In Mesp-LRE, Leu45, Leu70, Val111, Leu116, Phe248 and Leu257 constitute the hydrophobic pocket around the substrate, while the corresponding residues in Argl-LRE form another pocket with weaker hydrophobicity due to Met39, Met110 and Phe251 replacing the three leucines at positions 45, 116 and 257. Fewer hydrophobic residues participate in the pocket construction in Melo-LRE, including Ile65, Ile106, Phe243 and Leu252, producing a larger channel to the catalytic site with looser substrate recognition. It is worth mentioning that KEases usually prefer ketohexoses as substrates, despite having some activity towards ketopentoses. The relatively small hydrophobic pocket of LREase was proposed to be the main factor responsible for its marked specificity for ketopentoses (Uechi, Sakuraba *et al.*, 2013). In addition, it is observed that Met39 in Argl-LRE and Leu45 in Mesp-LRE severely block the channel to the catalytic site; this is absent in other KEases. Interestingly, the channel of the passageway to the catalytic site may shrink upon substrate binding due to a conformational change in which specific residues shift towards the bound substrate (Supplementary Fig. S3). The channel in Agtu-DAE shrinks from 6–7 Å to 4–5 Å on substrate binding, whereas the channel size is almost unchanged in Psci-DTE (Yoshida *et al.*, 2016). Overall, DAEases, DTEases and LREases exhibit strikingly different hydrophobic substrate-binding pockets, which may lead to a distinction in the specificity and affinity for substrates between DAEases, DTEAases and LREases.

Collectively, structural analysis identified the structural features of this novel DAEase and further provided a better understanding of the structure–function relationship of CDAE. Furthermore, KEase family enzymes were found to exhibit a similar catalytic environment, except for the hydrophobic pocket around the substrate, the size and hydrophobicity of which seem to provide the distinction in substrate preference among KEases.

Funding information

This work was financially supported by the National Natural Science Foundation of China (No. 31722040), China Agriculture Research System (CARS-02), the Distinguished Professor Project of Jiangsu Province, the Singapore Ministry of Education Academic Research Fund Tier 1 (R-160-000-A40-114) and the Applied Basic Research Project (Agricultural), Suzhou Science and Technology Planning Program (SNG2020061).

References

Chan, H.-C., Zhu, Y., Hu, Y., Ko, T.-P., Huang, C.-H., Ren, F., Chen, C.-C., Ma, Y., Guo, R.-T. & Sun, Y. (2012). *Protein Cell*, **3**, 123–131.
Chen, D., Chen, J., Liu, X., Guang, C., Zhang, W. & Mu, W. (2021). *Int. J. Biol. Macromol.* **189**, 214–222.
Chen, J., Chen, D., Ke, M., Ye, S., Wang, X., Zhang, W. & Mu, W. (2021). *Mol. Biotechnol.* **63**, 534–543.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Hu, M., Li, M., Jiang, B. & Zhang, T. (2021). *Compr. Rev. Food. Sci. Food. Saf.* **20**, 6012–6026.
Jia, D.-X., Sun, C.-Y., Jin, Y.-T., Liu, Z.-Q., Zheng, Y.-G., Li, M., Wang, H.-Y. & Chen, D.-S. (2021). *Enzyme Microb. Technol.* **148**, 109816.
Kim, H.-J., Hyun, E.-K., Kim, Y.-S., Lee, Y.-J. & Oh, D.-K. (2006). *Appl. Environ. Microbiol.* **72**, 981–985.
Kim, K., Kim, H.-J., Oh, D.-K., Cha, S.-S. & Rhee, S. (2006). *J. Mol. Biol.* **361**, 920–931.
Li, C., Li, L., Feng, Z., Guan, L., Lu, F. & Qin, H. (2021). *Food Chem.* **357**, 129746.
Li, C., Zhang, W., Wei, C., Gao, X., Mao, S., Lu, F. & Qin, H. (2021). *J. Agric. Food Chem.* **69**, 11637–11645.
Li, S., Chen, Z., Zhang, W., Guang, C. & Mu, W. (2019). *Int. J. Biol. Macromol.* **138**, 536–545.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Mao, S., Cheng, X., Zhu, Z., Chen, Y., Li, C., Zhu, M., Liu, X., Lu, F. & Qin, H. M. (2020). *Enzyme Microb. Technol.* **132**, 109441.

Maskeri, M. A., Brueckner, A. C., Feoktistova, T., O'Connor, M. J., Walden, D. M., Cheong, P. H.-Y. & Scheidt, K. A. (2020). *Chem. Sci.* **11**, 8736–8743.
Mu, W., Chu, F., Xing, Q., Yu, S., Zhou, L. & Jiang, B. (2011). *J. Agric. Food Chem.* **59**, 7785–7792.
Mu, W., Zhang, W., Fang, D., Zhou, L., Jiang, B. & Zhang, T. (2013). *Biotechnol. Lett.* **35**, 1481–1486.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Patel, S. N., Kaushal, G. & Singh, S. P. (2021). *Microb. Cell Fact.* **20**, 60.
Qi, Z., Zhu, Z., Wang, J. W., Li, S., Guo, Q., Xu, P., Lu, F. & Qin, H. M. (2017). *Microb. Cell Fact.* **16**, 193.
Shin, S.-M., Cao, T.-P., Choi, J.-M., Kim, S.-B., Lee, S.-J., Lee, S.-H. & Lee, D.-W. (2017). *Appl. Environ. Microbiol.* **83**, e03291-16.
Sim, L., Beeren, S. R., Findinier, J., Dauvillée, D., Ball, S. G., Henriksen, A. & Palcic, M. M. (2014). *J. Biol. Chem.* **289**, 22991–23003.
Trott, O. & Olson, A. J. (2010). *J. Comput. Chem.* **31**, 455–461.
Tseng, W.-C., Chen, C.-T., Hsu, C.-T., Lee, H.-C., Fang, H.-Y., Wang, M.-J., Wu, Y.-H. & Fang, T.-Y. (2018). *Int. J. Biol. Macromol.* **112**, 767–774.
Uechi, K., Sakuraba, H., Yoshihara, A., Morimoto, K. & Takata, G. (2013). *Acta Cryst.* D**69**, 2330–2339.
Uechi, K., Takata, G., Fukai, Y., Yoshihara, A. & Morimoto, K. (2013). *Biosci. Biotechnol. Biochem.* **77**, 511–515.
Wang, Y., Ravikumar, Y., Zhang, G., Yun, J., Zhang, Y., Parvez, A., Qi, X. & Sun, W. (2021). *Front. Chem.* **8**, 622325.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.
Xia, Y., Cheng, Q., Mu, W., Hu, X., Sun, Z., Qiu, Y., Liu, X. & Wang, Z. (2021). *Foods*, **10**, 2186.
Xie, X., Li, Y., Ban, X., Zhang, Z., Gu, Z., Li, C., Hong, Y., Cheng, L., Jin, T. & Li, Z. (2019). *Int. J. Biol. Macromol.* **138**, 394–402.
Yoshida, H., Yamada, M., Nishitani, T., Takada, G., Izumori, K. & Kamitori, S. (2007). *J. Mol. Biol.* **374**, 443–453.
Yoshida, H., Yoshihara, A., Gullapalli, P. K., Ohtani, K., Akimitsu, K., Izumori, K. & Kamitori, S. (2018). *Acta Cryst.* F**74**, 669–676.
Yoshida, H., Yoshihara, A., Ishii, T., Izumori, K. & Kamitori, S. (2016). *Appl. Microbiol. Biotechnol.* **100**, 10403–10415.
Yoshida, H., Yoshihara, A., Kato, S., Mochizuki, S., Akimitsu, K., Izumori, K. & Kamitori, S. (2021). *FEBS Open Bio*, **11**, 1621–1637.
Zhang, W., Chen, D., Chen, J., Xu, W., Chen, Q., Wu, H., Guang, C. & Mu, W. (2021). *Crit. Rev. Food Sci. Nutr.*, https://doi.org/10.1080/10408398.2021.2023091.
Zhang, W., Yu, S., Zhang, T., Jiang, B. & Mu, W. (2016). *Trends Food Sci. Technol.* **54**, 127–137.
Zhang, W., Zhang, Y., Huang, J., Chen, Z., Zhang, T., Guang, C. & Mu, W. (2018). *J. Agric. Food Chem.* **66**, 5593–5601.
Zhu, Z., Gao, D., Li, C., Chen, Y., Zhu, M., Liu, X., Tanokura, M., Qin, H. M. & Lu, F. (2019). *Microb. Cell Fact.* **18**, 59.
Zhu, Z., Li, L., Zhang, W., Li, C., Mao, S., Lu, F. & Qin, H. M. (2021). *Enzyme Microb. Technol.* **149**, 109850.