Catalysis Science & Technology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d2cy01542k

Structural and mechanistic insights into enantioselectivity toward near-symmetric esters of a novel carboxylesterase *Ro*CE†

Zhe Dou, [a] Peng Jia,[b] Xiaoyu Chen,[a] Zheng Wu,[a] Guochao Xu *[a] and Ye Ni *[a]

Received 1st September 2022,
Accepted 31st October 2022

DOI: 10.1039/d2cy01542k

rsc.li/catalysis

A novel carboxylesterase designated as *Ro*CE was identified from *Rhodococcus opacus* with high activity and enantioselectivity toward asymmetric esters such as ethyl 2,2-dimethylcyclopropane-1-carboxylate (DMCPE). Moreover, *Ro*CE could catalyze the enantioselective resolution of near-symmetric oxyheterocyclic esters such as ethyl tetrahydro-2*H*-pyran-2-carboxylate (THPCE), which are generally regarded as "hard-to-be-discriminated" by chemical and biological catalysts. The crystal structure of *Ro*CE was resolved at a resolution of 1.78 Å. Theozyme calculation, MD simulations and pre-reaction state analysis were performed to clarify the molecular basis for the enantioselectivity toward oxyheterocyclic carboxylic acid esters with a nearly symmetric structure. F166 plays an important role in manipulating the enantioselective recognition of (*S*)- and (*R*)-DMCPE through steric effect. The intrinsic symmetric structure of (*S*)- and (*R*)-THPCE is mainly responsible for the relatively lower enantioselectivity than DMCPE. By introducing hydrogen bond interactions, a mutant M144T was successfully obtained with an *E* value of 2.44-fold that of WT. MD simulations further prove the increased enantioselectivity of M144T in terms of pre-reaction state and binding free energy. This study provides a novel carboxylesterase and important molecular insights into the enantioselectivity of carboxylesterase toward heterocyclic carboxylic acid esters with a nearly symmetric structure, which will facilitate further engineering of the enantioselectivity of carboxylesterase.

Introduction

Optically active carboxylic acids and ester derivatives, referring to compounds with carboxylic acids and derivatives ligated to a stereogenic center, are important building blocks for the synthesis of extensive bioactive compounds, pharmaceuticals and agrochemicals, such as (*S*)-ibuprofen, (*S*)-naproxen, (*S*)-flurbiprofen, artemisinin, sacubitril, *etc.*[1–3] Carboxylic functionality could impact the water solubility of bioactive compounds and influence the pharmaceutical efficacy by manipulating the electrostatic interactions between drug and target.[4,5] Recent surveys showed that roughly 25% of all commercialized pharmaceuticals and 40% of all marked crop protection agents bear at least one carboxylic group.[2,6] Hence, the development of efficient and green synthetic methods targeting chiral carboxylic acids and derivatives has become increasingly important in synthetic chemistry. Traditional chemical methods, such as asymmetric hydrocarboxylation and hydroesterification of alkenes, suffer from high-pressure toxic CO and transition metals.[7,8] Enzyme-catalyzed asymmetric synthesis of chiral carboxylic acids and derivatives is of special interest in terms of green chemistry and has been practiced at a large scale.[9,10] Carboxylesterases are the most promising enzymes in the enantioselective preparation of chiral carboxylic acids and derivatives.[11,12]

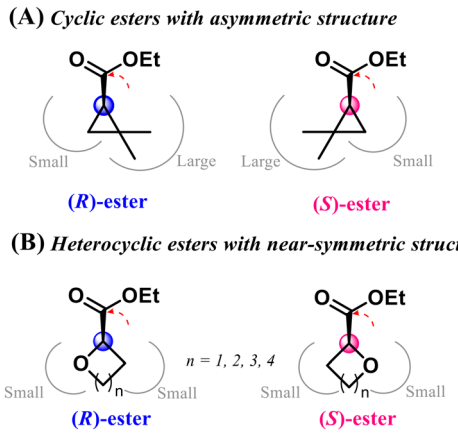
Carboxylesterases (CEs) or carboxylic ester hydrolases (EC 3.1.1.1) are members of the α/β hydrolase family and are widely distributed in microorganisms, animals and plants, with important physiological, industrial and medical roles in the synthesis and hydrolysis of stereospecific compounds.[13–15] The physiological roles of CEs are believed to protect cells from exogenous xenobiotics through cleavage of esters, amides, thioester groups, *etc.*[16,17] At present, CEs have been identified and manipulated to hydrolyze numerous structurally diverse compounds, such as cocaine, oseltamivir (Tamiflu), permethrin, irinotecan, *etc.*[15] Moreover, CEs usually display high resistance against harsh conditions such as high temperature and organic solvents.[18] Apart from that, in contrast to lipases which are also of industrial relevance for the synthesis of lipids with long side-chain length and

[a] *Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi 214122, Jiangsu, China. E-mail: guochaoxu@jiangnan.edu.cn, yni@jiangnan.edu.cn*
[b] *CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cy01542k

high hydrophobicity,[19,20] CEs prefer esters with short chain length and high hydrophilicity, or aromatic substituents, and exhibit high catalytic promiscuity.[21,22] Considering their application potential, many bacterial CEs have been identified from *Mycobacterium tuberculosis*, *Bacillus subtilis*, *Sphingobium faniae*, *etc.*[23–25] They share a consensus catalytic motif of Ser-Asp/Glu-His and conserved α/β fold surrounded by α-helices.[26,27] A number of studies have been committed to the enantioselective resolution of acyclic or cyclic carboxylic acid esters consisting of one 'large' and one 'small' substituent ligated to the stereogenic center (Scheme 1), such as ibuprofen, naproxen or cilastatin, which are widely prescribed analgesic and anti-infectious drugs.[28–30] Attributed to the large difference between the two substituents, stereo-discrimination and production of carboxylic acids with high enantioselectivity could be readily achieved in enzymatic reactions. In contrast to the acyclic or cyclic carboxylic acid esters, heterocyclic carboxylic acid esters refer to substrates with two 'small' substituents of little difference ligated to the stereogenic center (Scheme 1), which are also important chiral building blocks of pharmaceuticals such as danuglipron (PF-06882961) targeting type II diabetes.[31] Few enzymes have been reported with high enantioselectivity in discriminating enantiomers of heterocyclic carboxylic acid esters.[32] Moreover, the chemical resolution of heterocyclic carboxylic acid esters employing chiral auxiliaries is often challenging, attributed to the special symmetric structure.[33] Hence, heterocyclic carboxylic acid esters could be regarded as a 'hard-to-be-discriminated' substrates. Identification of novel carboxylesterase with high enantioselectivity toward heterocyclic carboxylic acid esters is of special interest for the efficient synthesis of chiral carboxylic acids and derivatives and also for elucidating the molecular mechanism of enantioselective carboxylesterases.

In this study, a novel carboxylesterase *Ro*CE was identified from *Rhodococcus opacus* by screening the hydrolase library preserved in our laboratory.[34] Catalytic performance revealed that *Ro*CE displayed relatively high enantioselectivity toward heterocyclic carboxylic acid esters such as ethyl 2,2-dimethylcyclopropane-1-carboxylate (DMCPE) and ethyl tetrahydro-2*H*-pyran-2-carboxylate (THPCE). We attempted to resolve the crystal structure of *Ro*CE and explore the molecular basis of enantioselectivity toward DMCPE and THPCE by theozyme calculation, molecular dynamic (MD) simulations and pre-reaction state analysis. Based on the clarified evidence, the enantioselectivity of *Ro*CE was rationally designed by introducing extra hydrogen bond interactions. This study provides significant molecular insights into the enantioselectivity of carboxylesterase toward esters with a nearly symmetric structure, which will facilitate further engineering of the enantioselectivity of carboxylesterase.



Scheme 1 Cyclic and heterocyclic carboxylic esters with asymmetric substituents (A) and near-symmetric substituents (B) ligated to the stereogenic center. Dashed red arrow denotes the direction of nucleophilic attack.

Experimental

Reagents and strains

Racemic ethyl 2,2-dimethylcyclopropane-1-carboxylate (DMCPE), ethyl oxirane-2-carboxylate (ORCE), ethyl oxetane-2-carboxylate (OTCE), ethyl tetrahydrofuran-2-carboxylate (THFCE) and ethyl tetrahydro-2*H*-pyran-2-carboxylate (THPCE) were obtained from Aladdin (Shanghai) Co. Ltd. Chemicals for crystallization of ultra-purity were purchased from Sigma Co. Ltd. Plasmid pET28a and strain *Escherichia coli* BL21(DE3) were stored in our laboratory.

Protein expression and purification

Cultivation and expression of recombinant *E. coli* BL21(DE3)/pET28-*Ro*CE were conducted as previously described.[34] After expression, the cells were harvested by centrifugation at 8000*g* and 4 °C for 10 min and washed with physiological saline at least two times. The resultant cells were resuspended in buffer A (20 mM PBS pH 7.4, 0.5 M NaCl) followed by disruption using a high-pressure homogenizer at 800 bar (ATS-BASIC I, ATS Co. Ltd). The crude enzyme extract was obtained by centrifugation at 10 000*g* for 20 min. The resultant lysates were filtered with a 0.2 μm filter and loaded onto a 5 mL HisTrap FF column (GE Healthcare, Co. Ltd), which was preequilibrated with buffer A. The recombinant *Ro*CE was gradiently eluted from the nickel column with imidazole concentrations increasing from 10 mM to 500 mM at a flow rate of 5 mL min^{-1}. Furthermore, gel filtration chromatography was conducted to remove impurities employing a Superdex 200 column (Cytiva Co. Ltd) and eluted with Tris-HCl (25 mM, pH 8.0) buffer supplemented with 150 mM NaCl. After evaluation by SDS-PAGE, the eluted proteins were concentrated to >40 mg mL^{-1} followed by freezing immediately in liquid nitrogen and used for crystallization.

Protein crystallization, data collection and structure determination

Purified *Ro*CEs were mixed with an equal volume of reservoir solution containing 0.1 mol L^{-1} HEPES (pH 7.5) and 40% (w/

v) PEG 400 and used for crystallization by the sitting-drop vapor diffusion method at 16 °C. Prior to data collection, crystals were soaked in a cryoprotectant solution composed of 20% (v/v) glycerol and cooled immediately in liquid nitrogen.

Crystallization of purified proteins was performed by using the sitting-drop vapor diffusion method. For each trial, 0.8 μL of the protein sample (10 mg mL^{-1}) and an equivalent volume of bath solution were added to each well of a crystal plate and equilibrated against 40 μL of the bath solution, then the plate was cultivated at 16 °C and checked frequently for crystal growth. For crystallization, the following kits (Hampton Research, USA) are frequently used: Index PEGRX-86, MBClass, Crystal Screen 1, Crystal Screen 2, GRAS Screen 3, GRAS Screen 4, GRAS Screen 5, GRAS Screen 6, *etc.* The best diffraction-producing crystal was produced in 0.1 M Bis-Tris buffer (pH 6.5) and 25% PEG 300 (w/v). All crystals were picked by using Crystal Cap SPINE HT (FAstal BioTech, Shanghai) and stored in liquid N_2. Data collection was conducted at the Shanghai Synchrotron Radiation Facility in BL17B and BL18U and data were processed by using the program HKL3000.[35] The crystal structure was solved by using the molecular replacement (MR) method with Phaser in the CCP4 program with *Rh*Est1 (PDB 4RNC), which shows 76% sequence identity to *Ro*CE, serving as a searching model.[36] Data reduction and structural refinement were carried out using WinCoot and Phenix software.[37] The Protein Data Bank (PDB) has the atomic coordinates of *Ro*CE under accession number 7YII. All graphics for the protein structures as well as electrostatic surface calculation were prepared by using the PyMOL program (**https://pymol.sourceforge.net/**). Data collection and refinement statistics are summarized in Table S1.†

Theozyme calculation

The enzyme:substrate complex, transition state and intermediate models were constructed according to the spatial structure of *Ro*CE. Density functional theory calculation was performed with the Gaussian16 package using the basis set B3LYP-D3/6-31+G(d,p) for conformation optimization and basis set B3LYP-D3/6-311+G(d,p) for single point energy calculation. Reaction coordinates are shown in the ESI.†

MD simulations

(*S*)- and (*R*)-DMCPE and THPCE were docked into the active center of *Ro*CE employing the AutoDock Vina.[38] Poses with relatively higher scores were examined and regarded as the initial structure for MD simulations. *Ro*CE was protonated using H^{++} with protonated S101 and H253 in HIP form. All the MD simulations were performed using the Amber20 packages with GPU acceleration on the hypercomputer at the School of Biotechnology.[39] The force fields for protein, substrates and water were ff14SB, gaff and tip3p, respectively.[40] Sodium and chloride were added to neutralize the systems, and then a TIP3P water box with a clearance distance of 12 Å around the protein was added. The final simulation systems consisted of about 20 000 atoms. First, energy minimization of 10 000 steps with the steepest descent method was performed, followed by 1 ns heating at NVP from 0 K to 300 K and 2 ns equilibration at NPT. MD production was conducted at 300 K in NPT ensembles for 100 ns with dt of 0.002 ps. All the simulations were performed at least five times. After simulation, root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) were analyzed using Amber20. The binding free energy with MM-PBSA was calculated with Amber20.[41] The sampled MD trajectories were de-solvated and used for analysis of the distance and angle using Chimera 1.6.[42] Interactions between enzymes and substrates were analyzed employing Discover studio 4.5. Characterized MD conformations were visualized with PyMOL.

Site-directed mutagenesis

Site-directed mutagenesis was conducted using KOD Plus Neo polymerase. The PCR reaction mixture consisted of 20 ng templated DNA, 0.5 U KOD Plus Neo, 0.1 μM of each primer, appropriate amount of KOD Plus Neo buffer, and sterilized deionized water up to 20 μL. The PCR program was set as follows: pre-denaturation at 95 °C for 5 min, 20 cycles of denaturation at 98 °C for 15 s, annealing at 55 °C for 30 s, elongation at 68 °C for 3.5 min, followed by further elongation at 68 °C for 10 min. The resultant PCR product was verified by electrophoresis and digested with 10 U *Dpn*I at 37 °C for 1 h to remove the template plasmids. Plasmids were chemically transformed into competent cells of *E. coli* BL21(DE3) followed by incubation at 37 °C for 12 h. Positive colonies were identified by colony PCR and sequencing. Expression of the mutants was performed as mentioned above.

Enantioselective resolution reaction

The general protocol for enantioselective hydrolysis reaction was as follows: 10 mL Tris-HCl buffer (pH 8.0, 200 mM), 1 mmol esters (100 mM), and 10 U *Ro*CE (1 kU L^{-1}) at 30 °C and 180 rpm. Samples were withdrawn from the reaction mixtures at different time intervals, extracted with an equal volume of ethyl acetate, dried over anhydrous Na_2SO_4 for 6 h, and used for the determination of the conversion ratios and ee values based on an internal standard using GC equipped with a chiral CP7502 column as previously described.[30] Enantioselectivity including ee_s value and *E* value was calculated according to the following equation.[43]

$$ee_s = \frac{[S1]-[R1]}{[S1]+[R1]} \times 100\% \quad (1)$$

$$E\ \text{value} = \frac{\ln[(1-c)(1-ee_s)]}{\ln[(1-c)(1+ee_s)]} \quad (2)$$

where [R1] and [S1] are the peak areas corresponding to (*R*)-substrate and (*S*)-substrate. *c* denotes the conversion ratio (%).

Results and discussion

Identification of *Ro*CE for asymmetric synthesis of chiral acids

Previously, a carboxylesterase *Ro*CE was identified from *Rhodococcus opacus* with high catalytic efficiency whereas low enantioselectivity toward 3-cyclohexene-1-carboxylic ester. The application potential of *Ro*CE in the asymmetric synthesis of chiral acids and esters with near-symmetric structures, including cyclic esters ethyl 2,2-dimethylcyclopropane-1-carboxylate (DMCPE), and oxyheterocyclic esters such as ethyl oxirane-2-carboxylate (ORCE), ethyl oxetane-2-carboxylate (OTCE), ethyl tetrahydrofuran-2-carboxylate (THFCE) and ethyl tetrahydro-2*H*-pyran-2-carboxylate (THPCE), was explored as shown in Table 1. The specific activity of *Ro*CE toward DMCPE was 6.86 U mg^{-1}. As much as 50 mM DMCPE could be enantioselectively hydrolyzed to obtain (*S*)-CMCPA, which is an important building block for cilastatin.[44] At a conversion ratio of 50.1%, ee_s and *E* values >99.9% and >200, respectively, were achieved (Fig. S1†). This newly identified *Ro*CE displays excellent enantioselectivity toward cyclic carboxylic acid esters carrying asymmetric substituents such as DMCPE.

Considering the prominent performance of *Ro*CE toward DMCPE, oxyheterocyclic carboxylic esters with near-symmetric structure were further investigated, which are important chiral building blocks of pharmaceuticals such as danuglipron (PF-06882961) targeting type II diabetes.[31] The specific activities of *Ro*CE toward ORCE, OTCE, THFCE and THPCE were 143, 107, 302 and 148 U mg^{-1}, respectively, indicating that *Ro*CE is an efficient biocatalyst for the hydrolysis of oxyheterocyclic carboxylic esters. Enantioselective resolution reactions were further conducted at 50 mM and were not terminated until the ee_s values reached >99%. As shown in Table 1, for ORCE, OTCE, THFCE and THPCE, the ee_s values reached 95.5%, 95.8%, 96.0% and 95.3% at conversion ratios of 94.2%, 91.1%, 85.6%, 74.9%, with *E* values calculated to be 2.7, 3.2, 4.1 and 6.2, respectively (Fig. S2–S5†). *Ro*CE exhibited the highest activity toward THFCE and the highest enantioselectivity toward THPCE. To the best of our knowledge, this is the highest record for the enantioselective resolution of oxyheterocyclic carboxylic esters. To gain insight into the molecular basis of the enantioselectivity and provide guidance for further protein engineering, crystallization, theozyme calculation, MD simulations and pre-reaction state analysis of *Ro*CE were conducted.

Table 1 Enantioselective resolution of various esters with near-symmetric structure employing recombinant *Ro*CE

Substrate	Specific activity/U mg^{-1}	Conv./%	ee_s/%	*E* value
DMCPE	6.86 ± 0.12	50.1	>99.9	>200
ORCE	143 ± 3	94.2	95.5	2.7
OTCE	107 ± 2	91.1	95.8	3.2
THFCE	320 ± 5	85.6	96.0	4.1
THPCE	148 ± 4	74.9	95.3	6.2

Crystallization and structure elucidation of *Ro*CE

This novel carboxylesterase *Ro*CE belongs to the α/β hydrolase superfamily, which contains eight subfamilies.[45] According to phylogenetic analysis, *Ro*CE displays high sequence similarity toward carboxylesterases in family V and should be classified into this subfamily (Fig. 1A). Interestingly, enzymes grouped in family V usually originate from acid-adapted, cold-adapted or heat-adapted microorganisms, such as *Sto*Est from *Sulfolobus tokodaii*, *Ssh*Est1 from *Sulfolobus shibatae*, and PestE from *Pyrobaculum calidifontis*, which could tolerate harsh conditions.[45] Relatively few enzymes in family V were identified from mesophilic organisms such as *Rhodococcus*. Carboxylesterases in family V share significant amino acid sequence similarity of 20–25% with various bacterial non-lipolytic enzymes, including epoxide hydrolases, dehalogenases and haloperoxidases, rendering them with diverse applicability and catalytic promiscuity.[46,47]

Purified *Ro*CE was obtained through nickel affinity chromatography and gel-filtration chromatography (Fig. S6†), and the enzyme properties were characterized. *Ro*CE displayed the highest activity at pH 9.0 and 40 °C (Fig. S7†). No metal ions were found with capability in enhancing the activity of *Ro*CE. Moreover, the relative activity under addition of EDTA was about 86%. All the above indicated that *Ro*CE was not a metal ion-dependent enzyme. The substrate preference of *Ro*CE toward *p*-nitrophenyl (*p*NP) esters containing acyl chains with different chain lengths, including acetate ($pNPC_2$), butyrate ($pNPC_4$), caproate ($pNPC_6$), caprylate ($pNPC_8$), decanoate ($pNPC_{10}$), laurate ($pNPC_{12}$), myristate ($pNPC_{14}$) and palmitate ($pNPC_{16}$), was explored. The highest activity was found with $pNPC_2$. *Ro*CE preferred *p*NP esters with shorter chain length such as acetate and butyrate to *p*NP esters with longer chain length (Fig. 1B).

Purified *Ro*CE was further crystallized and used for X-ray diffraction. The crystal structure of the *apo* form of *Ro*CE was resolved at 1.78 Å resolution and deposited in the Protein Data Bank under accession number 7YII. The refinement statistics as well as model quality parameters are summarized in Table S1.† *Ro*CE is a homodimer, with each subunit comprising 275 amino acid residues. The two subunits interact in a back-to-back fashion to form a dimer, with the catalytic region oriented toward the opposite and open space (Fig. 2A). The subunit of *Ro*CE is globular with approximate dimensions of 44 Å × 42 Å × 40 Å. Each subunit consists of an eight-stranded, predominantly parallel, except for β_1 and β_4, central β sheet surrounded by eleven α-helices according to the topological analysis (Fig. 2B). Among the eight strands in the central β sheet of *Ro*CE, the β_3–β_8 strands provide the framework onto which the catalytic residue is placed, usually between β_5 and β_6 (Fig. S8†). The six loops at the C terminal ends of the parallel β sheets form the active

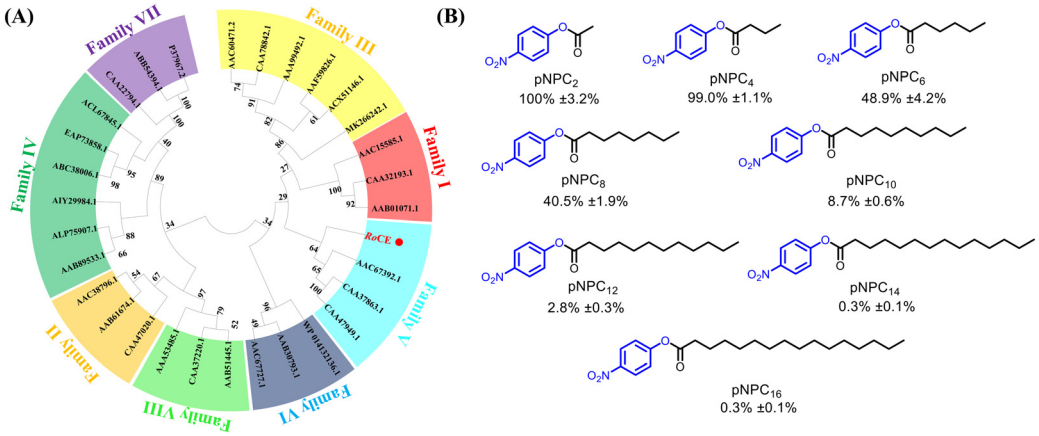


Fig. 1 Phylogenetic analysis of *Ro*CE with other carboxylesterases from family I to family VIII (A) and substrate spectrum analysis of *Ro*CE toward *para*-nitrophenol (*p*NP) esters with different chain lengths (B).

site cleft in each subunit. The bottom of the cleft is formed by three loops after the β_3, β_5 and β_6 strands.

According to the overall structure, each subunit of *Ro*CE comprises two modular domains, including a core domain harboring the α/β hydrolase fold and a V-shaped cap domain built up of four α-helices (Fig. 2C), similar to carboxylesterase from *Pseudomonas fluorescens*.[48] Moreover, the most notable structural feature of *Ro*CE is the nucleophile–acid–base catalytic triad, consisting of S101–D225–H253, in the core domain. This catalytic triad forms a charge relay system to generate a nucleophilic serine capable of attacking the carbonyl carbon atom of ester substrates. The catalytic triad is nestled on the top of the C-terminal β sheet. Catalytic S101 locates at the sharp turn between β_5 and α_3, namely the canonical nucleophilic elbow, in the distinct consensus motif Gly-X-Ser-X-Gly (X represents any amino acid residue), which is responsible for coordinating the enzymatic machinery. The conformation of S101 is stabilized by hydrogen bond interaction of the hydroxyl group with the imidazole side chain of H253 (Fig. 2D). D225 and H253 locate at loops between β_7 and α_9 and β_8 and α_1 in the C-terminal, respectively (Fig. 2C). The OD1 atom of D225 has hydrogen bond interaction with ND1 in H253 and the OD2 atom of D225 has hydrogen bond interaction with NE2 in H253 (Fig. 2D). The distances between S101 and H253 and D225 and H253 were calculated to be 3.5 Å and 3.1 Å, respectively, in the *apo* structure of *Ro*CE. Similar hydrogen bond networks are observed in the catalytic triad structures of other carboxylesterases. Moreover, an oxyanion hole was also identified adjacent to the catalytic nucleophile, which was composed of the main chain nitrogen atoms of W33 and Y102 (Fig. 2C). Through hydrogen network interaction, the oxyanion hole stabilizes the tertiary structure of the enzyme–substrate complex. Additionally, the oxyanion hole also takes part in organizing the spatial structure of the active center.

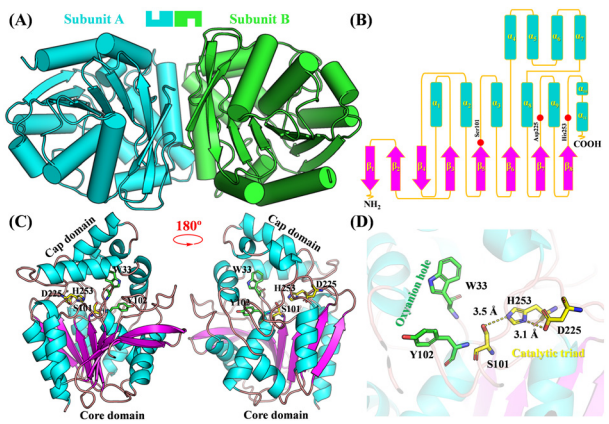


Fig. 2 Crystal structure analysis of *Ro*CE. (A) Overall structure, (B) topological structure analysis, (C) tertiary structure of one subunit, (D) enlarged view of the oxyanion hole and the catalytic triad.

The crystal structure of *Ro*CE was found to be highly similar to that of the esterase *Rh*Est1 from *Rhodococcus* sp. ECU1013 (PDB 4RNC) (with an RMSD of 1.25 Å), which also prefers esters with short-chain acyl substituents.[49] The overlapped structure of *Rh*Est1 and *Ro*CE showed that the catalytic triad and the oxyanion hole were almost in the same orientation (Fig. S9†). The largest deviation of these two enzymes originated from the cap domain, which might manipulate the entrance of substrates into the active center. Moreover, tremendous effort was also committed to obtaining the holo structure of *Ro*CE in complex with ester substrates through sitting-drop or soaking methods employing the deactivated mutant. However, no complex crystal was obtained, which might be attributed to its high flexibility. To further explore its molecular mechanism, MD simulation analysis was conducted.

Theozyme calculation and MD simulations of *Ro*CE toward DMCPE and THPCE

Pre-reaction state analysis is a generally accepted strategy for exploring the catalytic mechanism of enantioselective reactions. The definition of key parameters for pre-reaction

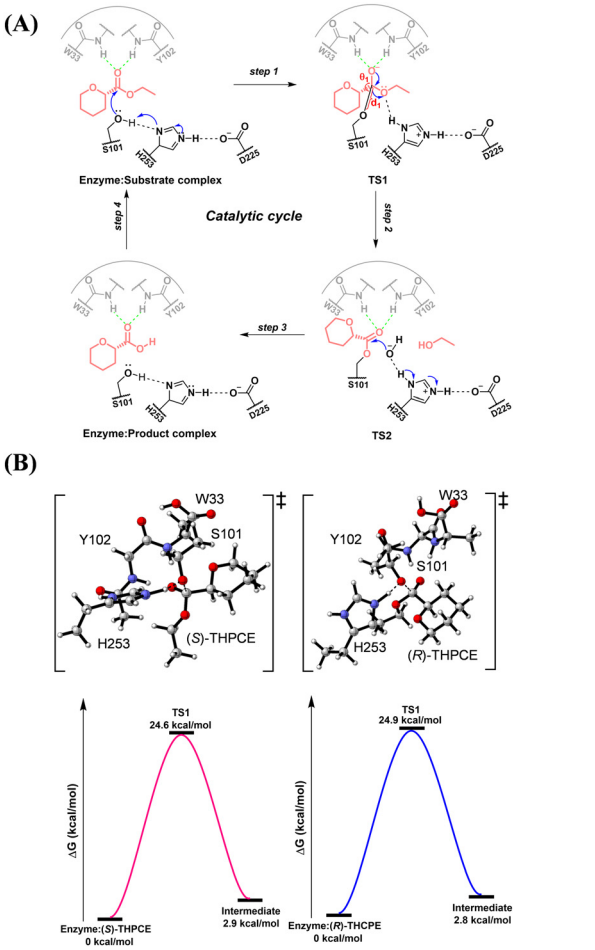


Fig. 3 Proposed catalytic cycle (A) and the theozyme calculations (B) of the *Ro*CE catalyzed enantioselective resolution of THPCE. Computation method: B3LYP-D3/6-31+G(d,p)//B3LYP-D3/6-311+G(d,p).

state analysis was conducted according to the catalytic cycle of carboxylesterases. Carboxylesterase-catalyzed hydrolysis is believed to proceed through several steps including formation and dissociation of two tetrahedral transition states (Fig. 3A).[50] Initially, a substrate such as THPCE enters the active center and is stabilized through hydrogen network interaction with W33 and Y102. Then, the hydroxyl H atom of the catalytic S101 is deprotonated by H253; simultaneously, the nucleophile O of S101 attacks the carbonyl C of THPCE. Then, the first tetrahedral transition state (TS1) is formed, with S101 ligating to the carbonyl C of the substrate and the negatively charged O of the substrate stabilized by hydrogen bond interactions with the oxyanion hole. This high-energy tetrahedral TS1 is highly unstable and the structure would be broken through proton transfer with the neighbouring H atom of H253. This step is a spontaneous process and produces ethanol to be released from the active center and the acyl–enzyme complex.

Furthermore, a water molecule in the active center would be stabilized and deprotonated by H253, producing a new nucleophile O of water, which would attack the carbonyl C of the acyl group of the substrate to form a second tetrahedral transition state (TS2). This TS2 is also highly unstable and is prone to be degraded into carboxylic acid to be released and a recovered catalytic S101 for the next round of the catalytic cycle. According to the free energy analysis, the formation of the TS1 is the rate-limiting step with the highest free energy barrier.[51] Hence, the parameters of the TS1 structure could be used for pre-reaction state analysis. Theozyme models of *Ro*CE toward (*S*)- and (*R*)-THPCE were constructed according to the catalytic mechanism. The TS1 was obtained and verified through intrinsic reaction coordinate analysis (Fig. S10†). The reaction energy profiles of *Ro*CE with (*S*)- and (*R*)-THPCE were also calculated as shown in Fig. 3B. The intrinsic activation energy barriers ($\Delta G^{\ddagger}$) of (*S*)- and (*R*)-THPCE were 24.6 and 24.9 kcal mol^{-1}, respectively, with an energy difference ($\Delta\Delta G^{\ddagger}$) of merely 0.3 kcal mol^{-1}. This demonstrates that there is no intrinsic difference for the hydrolysis of (*S*)- or (*R*)-THPCE. The enantioselectivity toward (*S*)- or (*R*)-THPCE should be attributed to the unique spatial structure of the substrate-binding pocket of *Ro*CE. Furthermore, two key parameters, the distance between the O atom of S101 and the carbonyl C atom of the substrate, and the angle among the O atom of S101, the carbonyl C atom and the O atom of the substrate, were designated as d_1 and θ_1, respectively. Concerning an effective attacking conformation, the d_1 and θ_1 should be lower than 3.0 Å and 90°. Hence, for identifying the trajectories satisfying the pre-reaction state, two cut-off parameters of d_1 ≤3.4 Å and θ_1 ≤90° ± 15° were employed, which are defined as reported.[52]

Multiple 100 ns MD simulations of *Ro*CE with protonated S101 and H253 were conducted. The trajectories were retrieved and used for pre-reaction analysis to explain the enantioselectivity of *Ro*CE. The distribution and population of d_1 and θ_1 of *Ro*CE with (*S*)- and (*R*)-DMCPE are illustrated in Fig. 4A and B. The

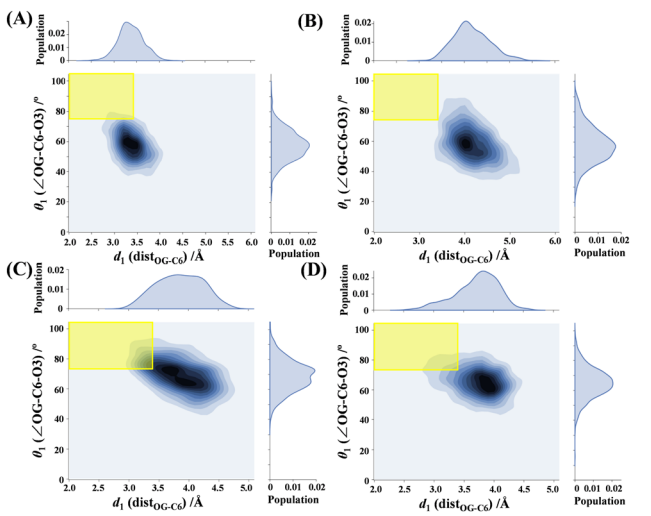


Fig. 4 Distribution and population of d_1 and θ_1 of *Ro*CE with substrates retrieved from MD trajectories. (A) (*S*)-DMCPE, (B) (*R*)-DMCPE, (C) (*S*)-THPCE, (D) (*R*)-THPCE. Yellow blocks denote pre-reaction state regions.

average d_1 and θ_1 of *Ro*CE with (*S*)- and (*R*)-DMCPE were 3.25 Å and 58° and 4.01 Å and 56° of the substrate, which holds promise for higher possibility of nucleophilic attack. The percentage of both d_1 and θ_1 satisfying the parameters of the pre-reaction state was 3.94% for (*S*)-DMCPE, significantly higher than 0.50% for (*R*)-DMCPE. The binding free energies (ΔG_{bind}) for (*S*)- and (*R*)-DMCPE were determined to be −14.8 ± 3.2 kcal mol^{-1} and −7.5 ± 2.1 kcal mol^{-1} employing the MM-PBSA method. The remarkable differences in pre-reaction state analysis and binding free energy are consistent with the excellent enantioselectivity (*E* value >200) of *Ro*CE toward DMCPE. The characteristic conformations of both (*S*)- and (*R*)-DMCPE were also retrieved. As shown in Fig. 5A and B, both (*S*)- and (*R*)-DMCPE could be accommodated in the active center owing to the interaction between W33 and the substrate. (*S*)-DMCPE is positioned with the dimethyl group toward the bottom of the active center. However, due to the existence of F166 which would introduce strong steric hindrance for the dimethyl group, (*R*)-DMCPE could only be positioned in an opposite pose with the dimethyl group toward Y102, which is also unfavorable in terms of steric and hydrophobic effects. As a result, an increase in distance d_1 from 3.0 Å in (*S*)-DMCPE to 3.5 Å in (*R*)-DMCPE was observed. An enlarged view of substrate alignment can be found in Fig. 5E. (*S*)- and (*R*)-DMCPE were nestled in the active center in completely reversed states. For THPCE, the distribution and population of d_1 and θ_1 of *Ro*CE with (*S*)- and (*R*)-THPCE are shown in Fig. 4C and D. The average d_1 and θ_1 of *Ro*CE with (*S*)- and (*R*)-THPCE were 3.75 Å and 71° and 3.83 Å and 64°, respectively. The percentages of both d_1 and θ_1 within the cut-off ranges of the pre-reaction state were 9.67% and 6.56% for (*S*)- and (*R*)-THPCE, respectively. With regard to the binding free energy, the ΔG_{bind} values for (*S*)- and (*R*)-THPCE were −16.2 ± 4.1 kcal mol^{-1} and −15.1 ± 3.8 kcal mol^{-1}, respectively. The high similarity in percentages of the pre-reaction states and binding free energies indicates that (*S*)- and (*R*)-THPCE could not be finely discriminated by *Ro*CE. According to the characteristic conformations, both (*S*)- and (*R*)-THPCE could be recognized and positioned in the active center in a quite similar conformation (Fig. 5C and D). Unlike DMCPE with a relatively larger difference due to the dimethyl group, the influence of F166 was trivial on the recognition of (*S*)- and (*R*)-THPCE. Because of the steric effect of Y102, (*S*)- and (*R*)-THPCE could only be nestled in the active center with a smaller alcohol group toward Y102 and a larger acyl group in the opposite direction. The distances d_1 between the O atom of S101 and the substrates were 2.3 Å and 2.4 Å, respectively. An enlarged view of substrate conformation is illustrated in Fig. 5F. Two enantiomers were situated in the active center in highly similar directions and poses. The only difference was the direction of the O atom in the oxyheterocyclic group of the substrates, which was upward in (*S*)-THPCE and downward in (*R*)-THPCE. The special characteristics and minor differences of these oxyheterocylic esters render them difficult to be discriminated by enzymes.

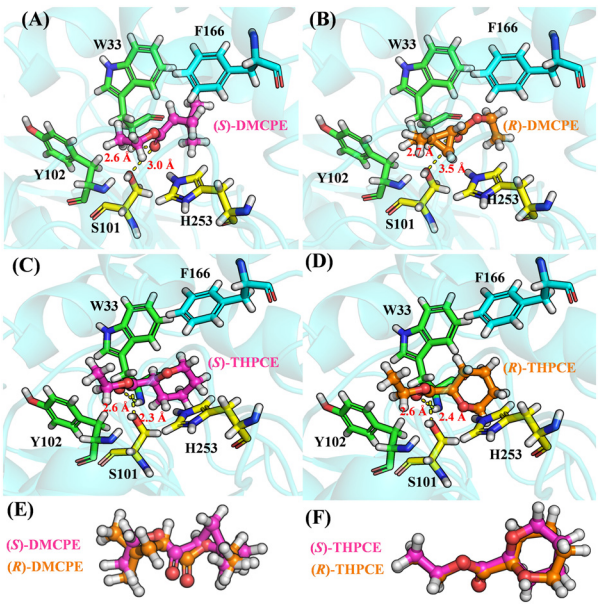


Fig. 5 Characteristic conformations retrieved from MD simulations. (A) *Ro*CE and (*S*)-DMCPE, (B) *Ro*CE and (*R*)-DMCPE, (C) *Ro*CE and (*S*)-THPCE, (D) *Ro*CE and (*R*)-THPCE, (E) alignment of (*S*)-DMCPE and (*R*)-DMCPE, (F) alignment of (*S*)-THPCE and (*R*)-THPCE.

Rational design of *Ro*CE and experimental verification

Inspired by the observations in MD simulations of *Ro*CE toward (*S*)- and (*R*)-THPCE, we envisioned rational engineering of enantioselectivity by manipulating residues in the active center. Considering the difference between the two enantiomers in the direction of the O atom, residues capable of donating a hydrogen bond should be introduced to form new hydrogen interactions with the O atom of (*S*)-THPCE. The strengthened interaction might improve the enantioselectivity of *Ro*CE toward (*S*)- and (*R*)-THPCE. Residues surrounding the acyl group of (*S*)-THPCE were analyzed as illustrated in Fig. 6A. The nearest residues to the O atom of the acyl group were M144 and V227 with a distance of less than 2.2 Å. Furthermore, these two residues were mutated into Thr (T), Ser (S), Cys (C), Gly (G), Gln (Q) and Tyr (Y),[53] which possess free hydroxy, thiol and amide groups and could function as hydrogen bond donors. The activity of the mutants decreased to some extent, such as M144T, M144Y and V227C with relative activities of 34%, 42% and 52%, respectively (Fig. S11†). The conversion ratios and ee_s values along the reaction progress of all the mutants were also monitored (Fig. 6B and D). Interestingly, some mutants were found with increased *E* values as expected such as mutants M144T and V227T, with *E* values as high as 15.1 and 12.1, which were about 2.44- and 1.95-fold that of wild-type *Ro*CE (WT). Under the same conditions, the yield was expected to increase by 5–10% compared with that of the WT. Consequently, the enantioselectivity of *Ro*CE toward the "hard-to-be-discriminated" substrates with a nearly symmetric structure was concisely and facilely increased by

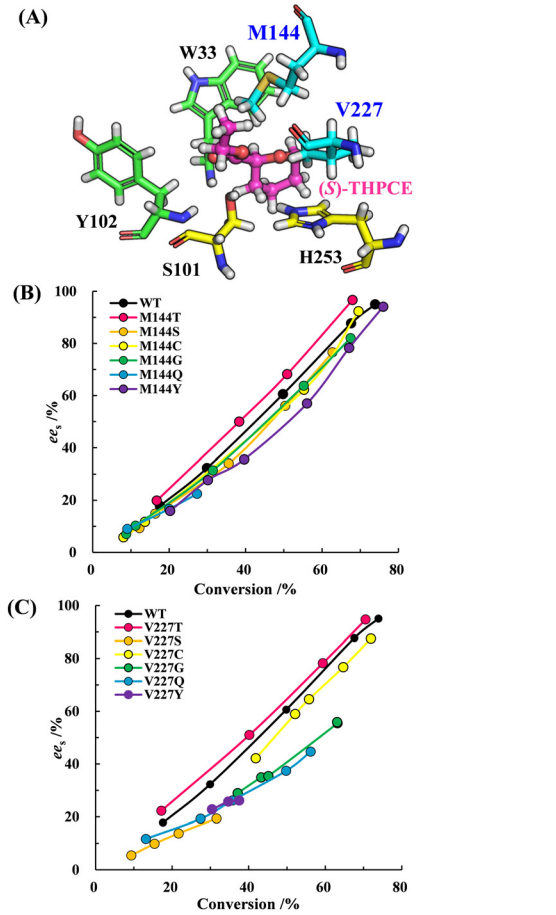


Fig. 6 Residues surrounding (*S*)-THPCE (A) and the conversion and ee_s of single mutants at M144 (B) and V227 (C) in the enantioselective resolution of *rac*-THPCE.

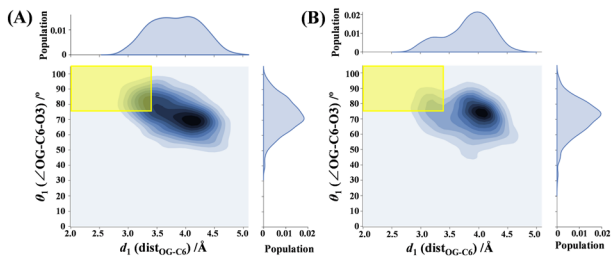


Fig. 7 Distribution and population of d_1 and θ_1 of M144T with substrates retrieved from MD trajectories. (A) (*S*)-THPCE, (B) (*R*)-THPCE. Yellow blocks denote pre-reaction state regions.

2.44-fold through structure- and computation-guided rational design.

Furthermore, multiple MD simulations were also performed to explore the difference in the enantioselectivity of M144T toward (*S*)- and (*R*)-THPCE (Fig. 7). The average d_1 and θ_1 of M144T toward (*S*)- and (*R*)-THPCE were 3.54 Å and 71° and 3.90 Å and 72°, respectively. The percentages of both d_1 and θ_1 satisfying the parameters of the pre-reaction state were 13.70% and 9.22% for (*S*)- and (*R*)-THPCE. With regard to the binding free energy, the ΔG_{bind} values toward (*S*)- and (*R*)-THPCE were −17.3 ± 3.6 kcal mol^{-1} and −14.9 ± 3.2 kcal mol^{-1}. The differences between (*S*)- and (*R*)-THPCE in terms of pre-reaction state and ΔG_{bind} of mutant M144T were 4.48% and 2.4 kcal mol^{-1}, higher than those of WT. To this end, the molecular evidence for the enantioselectivity of carboxylesterase *Ro*CE toward esters with a nearly symmetric structure was elucidated. More computationally guided rational and combinatorial design should be undertaken to further gear to the enantioselectivity of carboxylesterases.

Conclusions

In summary, a novel carboxylesterase *Ro*CE was identified with high enantioselectivity toward asymmetric esters and relatively high enantioselectivity toward esters with a nearly symmetric structure. The crystal structure of *Ro*CE was resolved at 1.78 Å and the consensus catalytic motifs were identified. Through theozyme calculation and multiple MD simulations, molecular evidence for the enantioselectivity of *Ro*CE was elucidated. For 'hard-to-be-discriminated' substrate THPCE with a nearly symmetric structure, mutant M144T with 2.44-fold increased *E* value was obtained by introducing hydrogen bond interactions through structure-guided computational design. This study provides essential molecular insights into the enantioselectivity of carboxylesterase toward esters with a nearly symmetric structure.

Author contributions

Z. D., G. X. and Y. N. conceived and designed the project. Z. D. conducted enzyme screening, characterization and mutation. P. J. performed the crystallization. X. C. carried out the computation analysis. Z. W. performed enantioselectivity analysis, Z. D., G. X. and Y. N. wrote the manuscript. G. X. and Y. N. obtained the financial support for the project leading to this publication.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

We are grateful to the National Key Research and Development Program (2019YFA0906400), the National Natural Science Foundation of China (22077054, 22078127), the Open Funding Project of the Key Laboratory of Industrial Biotechnology (KLIB-KF202101) and the Program of Introducing Talents of Discipline to Universities (111-2-06) for the financial support of this research.

References

1 C. Lamberth and J. Dinges, *Bioactive Carboxylic Compound Classes: Pharmaceuticals and Agrochemicals*, Wiley-VCH, 2016.

2 A. Halama and M. Zapadlo, *Org. Process Res. Dev.*, 2019, **23**, 102–107.
3 A. Böcker, P. R. Bonneau, O. Hucke, A. Jakalian and P. J. Edwards, *ChemMedChem*, 2010, **5**, 2102–2113.
4 S. E. David, P. Timmins and B. R. Conway, *Drug Dev. Ind. Pharm.*, 2012, **38**, 93–103.
5 J. Montoro, J. Sastre, J. Bartra, A. del Cuvillo, I. Davila, I. Jauregui, J. Mullol and A. L. Valero, *J. Invest. Allergol. Clin. Immunol.*, 2006, **16**, 24–28.
6 R. M. Lanigan and T. D. Sheppard, *Eur. J. Org. Chem.*, 2013, **33**, 7453–7465.
7 P. Yang, Y. X. Sun, K. Y. Fu, L. Zhang, G. Yang, J. Y. Yue, Y. Ma, J. R. S. Zhou and B. Tang, *Angew. Chem., Int. Ed.*, 2022, **61**, e202111778.
8 S. Feng and S. L. Buchwald, *J. Am. Chem. Soc.*, 2021, **143**, 4935–4941.
9 C. An, M. H. Shaw, A. Tharp, D. Verma, H. Li, H. Wang and X. Wang, *Org. Lett.*, 2020, **22**, 8320–8325.
10 F. Hollmann, D. J. Opperman and C. E. Paul, *Angew. Chem., Int. Ed.*, 2021, **60**, 5644–5665.
11 M. J. Hatfield, R. A. Umans, J. L. Hyatt, C. C. Edwards, M. Wierdl, L. Tsurkan, M. R. Taylor and P. M. Potter, *Chem.-Biol. Interact.*, 2016, **259**, 327–331.
12 H. Muller, S. P. Godehard, G. J. Palm, L. Bernndt, C. P. S. Badenhorst, P. S. Christoffel, A. K. Becker, M. Lammers and U. T. Bornscheuer, *Angew. Chem., Int. Ed.*, 2020, **60**, 2013–2017.
13 U. T. Bornscheuer, *Curr. Opin. Biotechnol.*, 2002, **13**, 543–547.
14 S. Sood, A. Sharma, N. Sharma and S. S. Kanwar, *Insights in Enzyme Research*, 2016, **1**, 1–11.
15 U. U. M. Johan, R. N. Z. R. Abd Rahman, N. H. A. Kamarudin and M. S. M. Ali, *Colloids Surf., B*, 2021, **205**, 111882.
16 U. T. Bornscheuer, *FEMS Microbiol. Rev.*, 2002, **26**, 73–81.
17 M. R. Franklin, *XPharm. Compr. Phamacol. Ref.*, 2007, pp. 1–3.
18 T. Satoh and M. Hosokawa, *Chem.-Biol. Interact.*, 2006, **162**, 195–211.
19 U. T. Bornscheuer, *Annu. Rev. Food Sci. Technol.*, 2018, **9**, 85–103.
20 S. K. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angew. Chem., Int. Ed.*, 2021, **60**, 88–119.
21 M. Schmidt, M. Baumann, E. Henke, M. Konarzycka-Bessler and U. T. Bornscheuer, *Methods Enzymol.*, 2004, **388**, 199–207.
22 H. Jochens, M. Hesseler, K. Stiba, S. K. Padhi, R. J. Kazlauskas and U. T. Bornscheuer, *ChemBioChem*, 2011, **12**, 1508–1517.
23 S. C. Lun and W. R. Bishai, *J. Biol. Chem.*, 2007, **282**, 18348–18356.
24 H. J. Rozeboom, L. F. Godinho, M. Nardini, W. J. Quax and B. W. Dijkstra, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 567–575.
25 D. Xu, Y. Gao, B. Sun, T. Ran, L. Zeng, J. He, J. He and W. Wang, *Appl. Environ. Microbiol.*, 2020, **86**, e02971-19.
26 M. Holmquist, *Curr. Protein Pept. Sci.*, 2005, **1**, 209–235.
27 M. Hosokawa, *Molecules*, 2008, **13**, 412–431.
28 D. T. Zhao, H. Yue, G. Chen, L. Y. Jiang, H. Zhang, Z. Wang and G. C. Liu, *Biotechnol. Appl. Biochem.*, 2014, **61**, 655–659.
29 X. Liu, J. H. Xu, J. Pan and J. Zhao, *Appl. Biochem. Biotechnol.*, 2010, **162**, 1574–1584.
30 Y. Zhang, J. Pan, Z. J. Luan, G. C. Xu, S. H. Park and J. H. Xu, *Appl. Environ. Microbiol.*, 2014, **80**, 7348–7355.
31 A. R. Saxena, D. N. Gorman, R. M. Esquejo, A. Bergman, K. Chidsey, C. Buckeridge, D. A. Griffith and A. M. Kim, *Nat. Med.*, 2021, **27**, 1079–1087.
32 Z. B. Cheng, C. M. Sun and Z. H. Tian, *Chin. Patent*, CN109868298, 2019.
33 R. Alajarin, J. J. Vaquero, J. Alvarez-Builla, M. Pastor, C. Sunkel, M. F. de Casa-Juana, J. Priego, P. R. Statkow, J. Sanz-Aparicio and I. Fonseca, *J. Med. Chem.*, 1995, **38**, 2830–2841.
34 Z. Dou, G. C. Xu and Y. Ni, *Enzyme Microb. Technol.*, 2020, **139**, 109580.
35 Z. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307–326.
36 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.
37 P. Emsley and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2004, **60**, 2126–2132.
38 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
39 A. W. Goetz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.
40 C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu and C. Simmerling, *J. Chem. Theory Comput.*, 2019, **16**, 528–552.
41 N. Homeyer and H. Gohlke, *Mol. Inf.*, 2012, **31**, 114–122.
42 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
43 Z. Dou, G. C. Xu and Y. Ni, *Bioresour. Technol.*, 2020, **317**, 123984.
44 Q. Wang, F. Yang, H. Du, M. Mahmum Hossain, D. Bennett and D. S. Grubisha, *Tetrahedron: Asymmetry*, 1998, **9**, 3971–3977.
45 J. L. Arpigny and K. E. Jaeger, *Biochem. J.*, 1999, **343**, 177–183.
46 K. H. Verschueren, F. Seljée, H. J. Rozeboom, K. H. Kalk and B. W. Dijkstra, *Nature*, 1993, **24**, 693–698.
47 E. Misawa, C. K. Chan Kwo Chion, I. V. Archer, M. P. Woodland, N. Y. Zhou, S. F. Carter, D. A. Widdowson and D. J. Leak, *Eur. J. Biochem.*, 1998, **1**, 173–183.
48 K. K. Kim, H. K. Song, D. H. Shin, K. Y. Hwang, S. Choe, O. J. Yoo and S. W. Suh, *Structure*, 1997, **5**, 1571–1584.
49 Z. J. Luan, F. L. Li, S. Dou, Q. Chen, X. D. Kong, J. H. Zhou, H. L. Yu and J. H. Xu, *Catal. Sci. Technol.*, 2015, **5**, 2622–2629.
50 T. Satoh and M. Hosokawa, *Annu. Rev. Pharmacol. Toxicol.*, 1998, **38**, 257–288.
51 L. Corici, A. Pellis, V. Ferrario, C. Ebert, S. Cantone and L. Gardossi, *Adv. Synth. Catal.*, 2015, **357**, 1763–1774.
52 T. D. Fonseca, K. B. Veja, M. R. da Silva, M. D. F. Ferreira de Oliveira, T. L. G. Gomes de Lemos, M. L. Contente, F. Molinari, M. Cespugli, S. Fortuna, L. Gardossi and M. C. de Mattos, *Mol. Catal.*, 2020, **485**, 110819.
53 L. M. Ramaniah, C. Kamal, R. J. Kshirsagar, A. Chakrabarti and A. Banerjee, *Mol. Phys.*, 2013, **111**, 3067–3076.