Crystallographic analysis of a novel aldo-keto reductase from *Thermotoga maritima* in complex with NADP⁺

Hai Hou,^{a,b} Ruiying Li,^a Xiaoyan Wang,^a Zhen Yuan,^a Xuemeng Liu,^a Zhenmin Chen^c and Xiaoling Xu^{a*}

Received 17 March 2015

Accepted 20 May 2015

Edited by I. Kuta Smatanova, University of South Bohemia, Czech Republic.

Keywords: aldo-keto reductases; thermostable; *Thermotoga maritima*.

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

^aInstitute of Ageing Research, School of Medicine, Hangzhou Normal University, Hangzhou 311121, People's Republic of China, ^bKey Laboratory for Space Bioscience and Space Biotechnology, School of Life Sciences, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, People's Republic of China, and ^cLaboratory of Biocatalysis, Hangzhou Normal University, Hangzhou 311121, People's Republic of China. *Correspondence e-mail: xuxl@hznu.edu.cn

Aldo-keto reductases (AKRs) are a superfamily of NAD(P)H-dependent oxidoreductases that catalyse the asymmetric reduction of aldehydes and ketones to chiral alcohols in various organisms. The novel aldo-keto reductase Tm1743 from *Thermotoga maritima* was identified to have a broad substrate specificity and high thermostability, serving as an important enzyme in biocatalysis and fine-chemical synthesis. In this study, Tm1743 was over-expressed in *Escherichia coli* BL21(DE3) cells with an N-terminal His₆ tag and was purified by Ni²⁺-chelating affinity and size-exclusion chromatography. Purified recombinant enzyme was incubated with its cofactor NADP⁺ and its substrate ethyl 2-oxo-4-phenylbutyrate (EOPB) for crystallization. Two X-ray diffraction data sets were collected at 2.0 and 1.7 Å resolution from dodecahedral crystals grown from samples containing Tm1743–NADP⁺–EOPB and Tm1743–NADP⁺, respectively. Both crystals belonged to space group *P*₃121, with similar unit-cell parameters. However, in the refined structure model only NADP⁺ was observed in the active site of the full-length Tm1743 enzyme. Degradation of the N-terminal vector-derived amino acids during crystallization was confirmed by Western blot and mass-spectrometric analyses.

1. Introduction

Aldo-keto reductases (AKRs) are a superfamily of NAD(P)H-dependent oxidoreductases that catalyse the reversible reduction of compounds containing aldehydes and ketones to the corresponding alcohols (Jez *et al.*, 1997; Jin & Penning, 2007; Penning, 2015). More than 190 AKRs that fall into 16 families are widely distributed among many prokaryotes and eukaryotes (Mindnich & Penning, 2009; Penning *et al.*, 2001; Penning, 2015). The wide spectrum of AKR substrates indicates their broad range of physiological roles in various organisms. They are important enzymes in many industrial and biological processes, including vitamin C biosynthesis (Anderson *et al.*, 1985), polyketide biosynthesis (Ikeda *et al.*, 1999), steroid metabolism (Takikawa *et al.*, 1987), diabetic complications (Kador, 1988) and xylose metabolism in yeast (Hahn-Hägerdal *et al.*, 1991). In addition, human AKR enzymes play central roles in the bioactivation or detoxication of drugs, carcinogens and reactive aldehydes (Jin & Penning, 2007; Chen & Zhang, 2012); they are considered to be phase I drug-metabolizing enzymes (Penning, 2015).

Owing to their excellent stereoselectivity, several microbial AKRs have been exploited in asymmetric reduction of prochiral ketones to chiral alcohols, which are the key building

Table 1

Macromolecule-production information.

Source organism	<i>T. maritima</i> (Gene ID 897867)
DNA source	Synthetic DNA (codon-optimized)
Forward primer†	CGGGATCCATGGGCAGCAGCCATCATCA
Reverse primer‡	CCGCTCGAGTTAGCCAGACTATCCAGCA
Cloning vector	pET-28a(+)
Expression vector	pET-28a(+)
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced§	MGSSHHHHSSGLVPRGSHMASMTGGQMGRGS- MLYKELGRTGEEIPALGLGTWIGGFETPDYS- RDEEMVELLKTAKMGYTHIDTAEEYGGGHT- ELIGKAIKDFRREDLFIVSKVWPTLRRDLL- RSLENTLKRLDTDYVDLYLIHWPNEIPLEET- LSAMAEQVRQGLIRYIGVSNFDRRLLEEAISK- SQEPICVDQVKYNIEDRDPERDGLLEFCQKNG- VTLVAYSPLRRTLLSEKTKRTLEEIAKNHGAT- IYQIMLAWLLAKPNVVAIPKAGRVEHLRENLK- ATEIKLSEEMKLLDSLGG

† The underlined sequence is the BamHI site. ‡ The underlined sequence is the XhoI site. § The underlined sequence indicates the vector-derived amino acids of the recombinant Tm1743 enzyme, including the thrombin cleavage site (LVPRGS).

blocks of many important intermediates in the synthesis of numerous pharmaceuticals and high-value fine chemicals (Ellis, 2002; Schweiger *et al.*, 2010; Ni *et al.*, 2011). In particular, some thermostable microbial AKRs have been extensively applied in organic synthesis and biocatalysis owing to their stability at high temperatures and pressures, as well as at high concentrations of chemical denaturants (Vieille & Zeikus, 2001).

In previous studies, we identified the novel aldo-keto reductase Tm1743 from *Thermotoga maritima* (Ma *et al.*, 2013). This enzyme demonstrated high thermostability, strong chemical tolerance and broad substrate specificity, which are the essential properties that are required for an excellent biocatalytic enzyme in fine-chemical and medicine synthesis. The optimal activity of Tm1743 was observed at 90°C and pH 9, and 63% of its activity was retained after 15 h incubation at 85°C. This high thermostability gives Tm1743 great potential in biocatalysis, since numerous reactions are performed at high temperatures and pressures, which allow higher substrate concentrations and reaction rates, low viscosity and a lower risk of microbial contamination (Vieille & Zeikus, 2001). Tm1743 also showed good tolerance of the organic solvents that are commonly used to improve the solubility of hydrophobic substrates, such as 10% (v/v) acetonitrile, methanol, isopropyl alcohol and DMSO. In addition, Tm1743 exhibited broad substrate specificity towards various keto esters, ketones and aldehydes. Excellent conversion rates were observed for both α - and β -keto esters. Its optimal substrate 2,2,2-trifluoroacetophenone was reduced to *S*-1-phenyl-2,2,2-trifluoroethanol with a 99.8% enantiomeric excess and 98% conversion (Ma *et al.*, 2013).

However, poor enantioselectivity was observed in the asymmetric reduction of ethyl 2-oxo-4-phenylbutyrate (EOPB) to ethyl (*R*)-2-hydroxy-4-phenylbutyrate (EHPB), a key chiral intermediate in the synthesis of angiotensin-converting enzyme inhibitors (ACEIs). In this reaction, a 73% enantiomeric excess of ethyl (*S*)-2-hydroxy-4-phenylbutyrate and a surprisingly high percentage of ethyl (*R*)-2-hydroxy-4-

phenylbutyrate (13% enantiomeric excess) were produced (unpublished data). According to the current enantioselectivity theory, EOPB is supposed to bind to the active site of Tm1743 in only one orientation and to be reduced to ethyl (*S*)-2-hydroxy-4-phenylbutyrate with excellent chiral purity. The production of a high percentage of ethyl (*R*)-2-hydroxy-4-phenylbutyrate triggered us to think about how these *R* enantiomers are produced: does a new stereorecognition mechanism exist in Tm1743? In addition, the target product for the synthesis of ACEIs is ethyl (*R*)-2-hydroxy-4-phenylbutyrate. Could we modify the enantioselectivity of Tm1743 through enzyme engineering to produce more ethyl (*R*)-2-hydroxy-4-phenylbutyrate for the synthesis of ACEIs?

However, no structural information is available to answer these questions. Here, we report the expression, purification, crystallization and preliminary X-ray crystallographic studies of Tm1743. The crystal structure of Tm1743 in complex with NADP⁺ or the substrate will serve as a template to study the enantioselectivity of this enzyme, contributing to the enzyme engineering of Tm1743 in fine-chemical and medicine synthesis.

2. Materials and methods

2.1. Cloning, expression and purification

The gene sequence encoding the full-length Tm1743 enzyme from *T. maritima* (Gene ID 897867) was codon-optimized and synthesized by GenScript (Nanjing, People's Republic of China). The optimized DNA sequence was introduced into pET-28a(+) vector (Novagen, USA) at the BamHI and XhoI restriction sites to construct a His₆-tagged recombinant expression vector (Table 1). The construct was checked by DNA sequencing and transformed into *Escherichia coli* BL21(DE3) cells (Novagen, USA). The transformed cells were grown in 1 l LB medium containing 100 mg ml⁻¹ kanamycin at 310 K until the OD₆₀₀ reached 0.6–0.8. The culture was then cooled and induced by adding isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM, and the bacteria were cultured for a further 120 h at 298 K.

Enzyme purification was performed according to a modification of a previously published procedure (Ma *et al.*, 2013). Cells containing the expressed Tm1743 enzyme were harvested by centrifugation at 4596g and 277 K for 15 min. Considering the high thermostability and chemical tolerance of Tm1743, the harvested cells were resuspended in 40 ml distilled water. The cells were then homogenized using a high-pressure homogenizer (Union, People's Republic of China) and the insoluble cell debris was removed by centrifugation at 34 541g for 40 min at 277 K. The supernatant containing the crude enzyme was boiled at 373 K for 10 min. Most of the bacterial proteins precipitated after heating, but the Tm1743 enzyme was still soluble. After centrifugation at 34 541g and 277 K for a further 20 min, the soluble fractions containing the Tm1743 enzyme were collected and diluted with an equal volume of nickel column binding buffer (25 mM Tris–HCl pH

Table 2Initial crystallization conditions and optimization results for the Tm1743–NADP⁺–EOPB sample.Tm1743 (20 mg ml^{−1}) was incubated with NADP⁺ and EOPB in a 1:1.5:1.5 molar ratio for 2 h at 277 K before crystallization. Crystallization trials were performed in 24-well plates using the hanging-drop vapour-diffusion method.

Temperature (K)	Condition	Initial condition	Crystal shape	Optimized condition	Diffraction
277	Crystal Screen No. 24	0.2 M calcium chloride dehydrate, 0.1 M sodium acetate trihydrate pH 4.6, 20%(v/v) 2-propanol	Cubic	0.2 M calcium chloride dehydrate, 0.1 M sodium acetate trihydrate pH 4.6, 25%(v/v) 2-propanol	No diffraction
289	PEGRx 2 No. 30	0.2 M magnesium formate dehydrate, 0.1 M sodium acetate trihydrate pH 4.0, 18%(v/v) PEG MME 5000	Irregular octahedral	Irreproducible	No diffraction
289	PEGRx 2 No. 45	5%(v/v) 2-propanol, 0.1 M citric acid pH 3.5, 6%(w/v) PEG 20 000	Irregular octahedral	Irreproducible	No diffraction
298	Crystal Screen 2 No. 34	0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.0 M sodium acetate trihydrate	Regular dodecahedral	0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.2 M sodium acetate trihydrate; enzyme concentration 15 mg ml ^{−1}	Diffacted to 2 Å resolution (Fig. 1a)

Figure 1

Crystal images and Western blot and mass-spectrometric analyses of the degraded enzyme and the crystals. (a) Crystal obtained from the Tm1743–NADP⁺–EOPB sample (1:1.5:1.5 molar ratio). The crystals have dimensions of about 1.2 × 1.2 × 0.6 mm. (b) Crystal of Tm1743 in complex with NADP⁺. (c) SDS–PAGE and Western blots of the dissolved crystals (lanes C1 and C2) alongside those of degraded recombinant enzyme (lane D). Lane C1 contains dissolved crystals from the Tm1743–NADP⁺–EOPB sample and lane C2 contains dissolved crystals from the Tm1743–NADP⁺ sample. Ponceau S staining and Western blots of these samples are shown below the SDS–PAGE. Lane M contains molecular-mass standards (labelled in kDa). (d) LC–ESI–MS analysis of the degraded enzyme (red) and dissolved crystals (blue). The molecular mass of each peak is indicated in Da.

Table 3

Crystallization of Tm1743 in complex with NADP⁺.

Sample	Tm1743–NADP ⁺ –EOPB	Tm1743–NADP ⁺
Method	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion
Plate type	24-well crystallization plates	24-well crystallization plates
Temperature (K)	298	298
Enzyme concentration (mg ml ^{−1})	15	12
Buffer composition of enzyme	25 mM Tris–HCl pH 8.5	25 mM Tris–HCl pH 8.5
Composition of reservoir solution	0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.2 M sodium acetate trihydrate	0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.3 M sodium acetate trihydrate
Volume and ratio of drop	3 µl, 1:1	3 µl, 1:1
Volume of reservoir (µl)	200	200
Crystal image	Fig. 1(a)	Fig. 1(b)

8.5, 20 mM imidazole). The diluted fractions were then loaded onto an Ni²⁺-chelating affinity column (GE Healthcare, USA) and rinsed with 100 ml binding buffer to remove nonspecifically bound proteins. The Tm1743 enzyme was eluted with a buffer consisting of 25 mM Tris–HCl pH 8.5, 50–100 mM imidazole. To improve the homogeneity of the enzyme, the eluate was dialyzed against 25 mM Tris–HCl pH 8.5 and was further purified using a HiLoad 16/600 Superdex 200 PG size-exclusion chromatography column (GE Healthcare).

2.2. Crystallization

The purified recombinant enzyme was concentrated to 30 mg ml^{−1} at 277 K using an Amicon Ultra centrifugal filter device (10 kDa molecular-weight cutoff; Millipore). The enzyme concentration was determined by the Bradford method. The recombinant enzyme was diluted to 20 mg ml^{−1} in 25 mM Tris–HCl pH 8.5 for crystallization. In parallel, the recombinant enzyme was incubated with the cofactor NADP⁺ and the substrate EOPB to attempt to obtain crystals of the enzyme–cofactor–substrate complex. It was incubated with NADP⁺ at a 1:1.5 molar ratio for 2 h at 277 K before crystallization; this sample is named Tm1743–NADP⁺. For crystallization of the ternary complex (Tm1743–NADP⁺–EOPB), a 1:1.5:1.5 molar ratio of Tm1743:NADP⁺:EOPB was used. Initial crystallization trials were performed in 24-well plates at three different temperatures (277, 289 and 298 K) using the hanging-drop vapour-diffusion method. 1 µl protein sample was mixed with an equal volume of reservoir solution and was equilibrated against 200 µl of the conditions from commercial crystallization screening kits from Hampton Research (Crystal Screen, Crystal Screen 2, Index, PEG/Ion, PEGRx 1, PEGRx 2 and SaltRx). Conditions that showed crystalline structures were optimized to yield suitable crystals by varying the pH and the precipitant and enzyme concentrations. The initial crystallization conditions and optimization results are shown in Tables 2 and 3.

2.3. X-ray data collection, processing and structure determination

The optimized crystals were cryoprotected by adding 25% glycerol to the reservoir solution and were flash-cooled in liquid nitrogen. Two data sets were collected from crystals obtained from the Tm1743–NADP⁺–EOPB and Tm1743–NADP⁺ samples. A 2.0 Å resolution data set was collected

from a crystal from the Tm1743–NADP⁺–EOPB sample at 100 K using an in-house X-ray source (Rigaku MicroMax-007 HF desktop rotating-anode X-ray generator with a Cu target operated at 40 kV and 30 mA) and an R-Axis VI⁺⁺ imaging-plate detector with a 130 mm crystal-to-detector distance at a wavelength of 1.5418 Å. 180 diffraction frames were collected with 1° oscillation per image. An X-ray diffraction data set was collected from a Tm1743–NADP⁺ crystal on beamline BL17U at Shanghai Synchrotron Radiation Facility (SSRF) using an ADSC Q315r detector. This crystal diffracted to 1.7 Å resolution with 1 s exposure time and a 450 mm crystal-to-detector distance at a wavelength of 0.97923 Å. Table 4 gives a summary of data collection and processing.

The diffraction data sets were indexed, integrated and subsequently scaled using the *HKL-3000* software suite (Minor *et al.*, 2006). The data quality was assessed using *SFHECK* (Vaguine *et al.*, 1999) and the solvent content was calculated using *MATTHEWS_COEF* (Matthews, 1968) from the *CCP4* package (Winn *et al.*, 2011). The initial phase of the structural model was searched for using *Phaser* (McCoy *et al.*, 2007) from the *CCP4* package. Iterative model building and refinement were performed using *Coot* (Emsley *et al.*, 2010) and *REFMAC5* (Murshudov *et al.*, 2011).

2.4. Peptide mass fingerprinting (PMF), liquid chromatography–electrospray ionization tandem mass spectrometry (LC-ESI-MS) and Western blotting

To identify the composition of the degraded enzyme and the crystals, peptide mass fingerprinting (PMF) was performed. The crystals obtained from the Tm1743–NADP⁺–EOPB and the Tm1743–NADP⁺ samples were picked out, dissolved in 25 mM Tris–HCl pH 8.5 and separated by SDS–PAGE alongside the degraded enzyme (Fig. 1c). The isolated bands were excised from the gel and digested with trypsin (1:100, Promega), and the peptides were extracted with 60% acetonitrile/0.1% trifluoroacetic acid, dried and resuspended in 0.1% formic acid solution. The peptide solutions were subjected to LC-MS/MS analysis with a Finnigan LTQ mass spectrometer (Thermo Fisher) and spectra were collected in the *m/z* range 400–2000. The search program *BioWorks* 3.3.1, developed by Thermo Fisher, was used for protein identification by peptide mass fingerprinting (PMF). The following parameters were used for the database search: trypsin as the cleaving enzyme, a maximum of one missed cleavage, iodo-

acetamide (Cys) as a complete modification, oxidation (Met) as a partial modification and a mass tolerance of ± 0.1 Da. The PMF acceptance criterion is probability scoring.

Liquid chromatography–electrospray ionization tandem mass spectrometry (LC-ESI-MS) was performed to validate the molecular weight of the degraded fragments in the enzyme and the crystals. 1 μ g of the enzyme and of dissolved crystals were loaded onto a C4 reverse-phase C18 column (Waters) connected to an Agilent 1290 HPLC system. The samples were then analysed with a 20 min HPLC gradient from 5 to 95% buffer B (buffer A, 0.1% formic acid in water; buffer B, 0.1% formic acid in acetonitrile) at 0.4 ml min⁻¹. The eluted fragments were ionized and directly introduced into an Agilent 6550 Q-TOF mass spectrometer using a Dual AJS ESI source. Full-scan MS spectra (m/z range 600–2000) were acquired in standard mode (2 GHz, 3200 m/z). The molecular weights of

Figure 2
SDS–PAGE gels (15%, Coomassie Blue-stained) for the most relevant purification stages of the His₆-tagged recombinant Tm1743 enzyme. (a) Ni²⁺-chelating affinity purification of the recombinant enzyme. Lane M, molecular-mass standards (labelled in kDa); lane C, cell lysates; lane S, supernatant of the cell lysates; lane H, soluble fraction of the heated supernatant; lane B, Ni²⁺-chelating agarose beads bound with recombinant enzyme; lane E20, eluted samples from binding buffer containing 20 mM imidazole; lanes E50 and E100, samples eluted with 50 and 100 mM imidazole, respectively. (b) Size-exclusion chromatogram and SDS–PAGE analysis (inset) of the gel-filtration fractions (lane G) and the degraded recombinant enzyme (lane D).

Table 4

X-ray diffraction data collection and processing.

Values in parentheses are for the highest resolution shell.

Sample for crystallization	Tm1743–NADP ⁺ –EOPB	Tm1743–NADP ⁺
Diffraction source	Rigaku MicroMax-007 HF	BL17U, SSRF
Wavelength (Å)	1.5418	0.97923
Temperature (K)	100	100
Detector	R-Axis VI ⁺⁺	ADSC Q315r
Crystal-to-detector distance (mm)	130	450
Rotation range per image (°)	1	1
Total rotation range (°)	180	180
Exposure time per image (s)	300	1
Space group	<i>P</i> ₃ 21	<i>P</i> ₃ 21
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 84.821, <i>c</i> = 93.727, α = β = 90, γ = 120	<i>a</i> = <i>b</i> = 84.922, <i>c</i> = 93.637, α = β = 90, γ = 120
Mosaicity (°)	1.10	0.47
Resolution range (Å)	50–2.00 (2.03–2.00)	50–1.70 (1.73–1.70)
Total No. of reflections	255930 (12490)	377861 (23371)
No. of unique reflections	25576 (1301)	39524 (2164)
Completeness (%)	95.8 (98.4)	90.8 (100.0)
Multiplicity	10.0 (9.6)	9.6 (10.8)
$\langle I/\sigma(I) \rangle$	9.8 (3.3)	15.2 (5.9)
<i>R</i> _{merge} † (%)	10.9 (0.0)	8.7 (99.3)
Overall <i>B</i> factor from Wilson plot (Å ²)	40.51	24.5
Matthews coefficient <i>V</i> _M (Å ³ Da ⁻¹)	2.79	2.79
Solvent content (%)	55.9	56.0

† $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the intensity of the i th measurement of reflection hkl and $\langle I(hkl) \rangle$ is the mean intensity of all symmetry-related reflections.

the degraded fragments in the enzyme and the crystals are shown in Fig. 1(d).

The SDS–PAGE gels of separated enzymes and dissolved crystals were transferred to a nitrocellulose membrane, where they were stained with 0.1% (w/v) Ponceau S in 5% acetic acid solution (Fig. 1c). Western blotting was performed using anti-His₆ tag antibody (Abcam), which recognizes proteins that contain either an N-terminal or a C-terminal HHHHHHGS epitope tag. N-terminal sequencing was performed by GeneCore BioTechnologies (Shanghai, People's Republic of China) using a PVDF membrane onto which the SDS–PAGE band of the dissolved crystals has been transferred.

2.5. Temperature-dependence of the activity of the enzyme in EOPB reduction

The activity of the enzyme in catalysing the reduction of EOPB was assayed at temperatures from 298 to 358 K. The assays were performed with a reaction mixture (3 ml) consisting of 3 mM EOPB, 30 mM dimethyl sulfoxide (DMSO), 50 mM Tris–HCl pH 9.0, 12 mM NADPH and an appropriate amount of enzyme at a specific temperature for 3 min. One unit (U) of the enzyme was defined as the amount of enzyme that was used to catalyse the oxidation of 1 μ mol of NADPH per minute. The change in the absorbance of NADPH at 340 nm ($\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$) was used to calculate the activity of the enzyme. The specific activity (units per

milligram of enzyme; U mg^{-1}) was plotted against the temperature to show the temperature-dependence of the enzyme activity.

3. Results and discussion

3.1. Enzyme expression and purification

The codon-optimized gene sequence of Tm1743 was introduced into pET-28a(+) vector at the BamHI and XhoI restriction sites. The encoded recombinant enzyme contains a 34-amino-acid vector-derived peptide followed by the coding region for the complete Tm1743 enzyme (Met1–Gly274), giving rise to a total of 308 residues (Table 1). The recombinant enzyme was expressed in *E. coli* BL21(DE3) cells and purified using an Ni^{2+} -chelating affinity column. After homogenization, the supernatant of the cell lysates (Fig. 2*a*, lanes S and C) was heated at 373 K for 10 min to separate the Tm1743 enzyme from other bacterial proteins. Owing to its high thermostability, Tm1743 was stable after heating (Fig. 2*a*, lane H), but most of the bacterial proteins were denatured and precipitated. The soluble fraction of the heated supernatant (Fig. 2*a*, lane H) was diluted in binding buffer and loaded onto an Ni^{2+} -chelating affinity column (Fig. 2*a*, lane B). The Tm1743 enzyme was eluted with binding buffer (Fig. 2*a*, lane E20) and with elution buffer containing 50 mM (Fig. 2*a*, lane E50) and 100 mM imidazole (Fig. 2*a*, lane E100).

To improve the homogeneity of the enzyme for crystallization, the elutes E50 and E100 were further purified using a HiLoad 16/600 Superdex 200 PG size-exclusion column. A single absorption peak at 280 nm was observed at 83.04 ml; this elution volume corresponds to a monomer of the Tm1743 enzyme. 6 mg purified recombinant enzyme was obtained from 1 l of bacterial culture. The molecular weight of the recombinant enzyme was estimated to be about 35 kDa by SDS–PAGE (Fig. 2*b*, lane G), coinciding with the calculated molecular weight of 35 032 Da. However, an additional band gradually appeared after storage of the enzyme at 277 K for 2 d (Fig. 2*b*, lane D). PMF analysis of this band showed excellent sequence coverage of Tm1743 (Supplementary Fig. S2), indicating that this band is a proteolytic fragment of the recombinant enzyme.

3.2. Crystallization and optimization

Initial crystallization trials were performed using the purified recombinant enzyme (Fig. 2*b*, lane G) and the Tm1743–NADP⁺ and Tm1743–NADP⁺–EOPB samples in parallel. However, no crystals were observed for the free enzyme. After 3–5 d of incubation, microcrystals were first observed for the Tm1743–NADP⁺–EOPB sample in Crystal Screen condition No. 24 at 277 K, PEGRx 2 condition Nos. 45 and 30 at 289 K and Crystal Screen 2 condition No. 34 at 298 K (Table 2). Optimization of the crystals grown from the Tm1743–NADP⁺–EOPB sample was performed based on these four initial conditions. However, the crystals grown in PEGRx 2 condition Nos. 45 and 30 could not be reproduced. Additionally, the crystals obtained in Crystal Screen condition No.

24 at 277 K were very small and showed no X-ray diffraction after optimization. Only the dodecahedral crystals grown in Crystal Screen 2 condition No. 34 were successfully optimized to diffraction quality (Table 2 and Fig. 3*a*). The optimized crystals were grown in a reservoir solution consisting of 0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.2 M sodium acetate trihydrate with a decreased enzyme concentration of 15 mg ml^{-1} (Tables 2 and 3). These crystals grew to dimensions of about $1.2 \times 1.2 \times 0.6 \text{ mm}$ (Fig. 1*a*). Based on the optimized crystallization condition, crystals were obtained from the Tm1743–NADP⁺ sample in 0.05 M cadmium sulfate hydrate, 0.1 M HEPES pH 7.5, 1.3 M sodium acetate trihydrate using an enzyme concentration of 12 mg ml^{-1} (Fig. 1*b* and Table 3).

3.3. Mass-spectrometric and Western blot analyses

SDS–PAGE of dissolved crystals from the Tm1743–NADP⁺–EOPB and the Tm174–NADP⁺ samples both showed a single band with a similar molecular weight to the proteolytic fragment of the recombinant enzyme (Fig. 1*c*, lanes C1 and

Figure 3
Representative X-ray diffraction images of the crystals. (*a*) Diffraction image of a crystal from the Tm1743–NADP⁺–EOPB sample collected using an in-house X-ray facility. The frame edge is at 2.0 Å resolution; an inset is presented for easier visualization of the spots. (*b*) Diffraction image of a Tm1743–NADP⁺ crystal at SSRF. The crystal diffracted to 1.55 Å resolution.

C2). PMF of these two bands also showed excellent sequence coverage of Tm1743 (Supplementary Fig. S3 and S4). Thus, only the proteolytic fragment of the recombinant enzyme was crystallized. LC-ESI-MS analysis of the degraded enzyme (Fig. 1c, lane D) and the dissolved crystals (Fig. 1c, lane C1) was applied to validate the degradation of the recombinant enzyme. LC-ESI-MS of the enzyme showed peaks for two major fragments with molecular weights of 34 772 and 33 488 Da and two small fragments with molecular weights of 32 539 and 32 450 Da. The largest fragment (34 772 Da) corresponds to the recombinant enzyme with the first three residues (MGS) missing, while the other fragments represent proteolytic fragments of different lengths, indicating gradual degradation of the recombinant enzyme. However, the dissolved crystals only contained one fragment, with a molecular weight of 32 236 Da. The calculated molecular weight of Tm1743 is 31 488 Da, suggesting that the crystals contain some of the vector-derived amino acids.

To further investigate whether the N- or C-terminus of the recombinant enzyme was degraded, Western blotting of the degraded enzyme and dissolved crystals were performed using an anti-His₆ antibody. Ponceau S staining of the nitrocellulose membrane was performed to indicate the transfer efficiency of the samples (Fig. 1c). Only the recombinant enzyme was blotted by the anti-His₆ tag; the proteolytic fragments in neither the degraded enzyme nor the crystals were blotted (Fig. 1c), suggesting that degradation occurred at the N-terminus of the recombinant enzyme. The crystals contain

an N-terminal His₆ tag-degraded proteolytic fragment of the recombinant enzyme. N-terminal sequencing of the dissolved crystals confirmed that the fragment starts with the sequence QQM. Thus, the crystallized proteolytic fragment (32 236 Da) contains full-length Tm1743 enzyme fused to an N-terminal vector-derived sequence QQMGRGS. The thrombin cleavage site (LVPRGS) included in the 34 vector-derived amino acids may provide a protease-attack site that results in degradation of the recombinant enzyme.

3.4. X-ray data collection, processing and structure determination

The crystal obtained from the Tm1743–NADP⁺–EOPB sample diffracted beyond 2.0 Å resolution in-house (Fig. 3a) and belonged to space group *P*₃₁21, with unit-cell parameters *a* = *b* = 84.821, *c* = 93.727 Å, $\alpha = \beta = 90^\circ$, $\gamma = 120.00^\circ$. The completeness of the overall data set is 95.8% and that in the highest resolution shell is 98.4%. The solvent content and Matthews coefficient are 55.9% and 2.79 Å³ Da^{−1}, respectively (Matthews, 1968). A 1.7 Å resolution data set was collected from a crystal obtained from the Tm1743–NADP⁺ sample (Fig. 3b). This crystal belonged to the same space group, *P*₃₁21, with similar unit-cell parameters *a* = *b* = 84.922, *c* = 93.637 Å, $\alpha = \beta = 90^\circ$, $\gamma = 120.00^\circ$ (Table 4).

Tm1743 shares 41% sequence identity with a putative oxidoreductase from *Sinorhizobium meliloti* 1021 (PDB entry 4pmj; New York Structural Genomics Research Consortium,

Figure 4

Sequence alignment of Tm1743 with a putative oxidoreductase from *S. meliloti* 1021 (S. meliloti 1021 OR; PDB entry 4pmj) and Akr11a from *Bacillus subtilis* (B. subtilis Akr11a; PDB entry 1pyf; Ehrensberger & Wilson, 2004). The conserved active-site residues Asp53, Tyr58, Lys84 and His117 are indicated by red stars.

unpublished work) and 33% sequence identity with Akr11a from *Bacillus subtilis* (PDB entry 1pyf; Ehrensberger & Wilson, 2004) (Fig. 4). The BLAST scores of PDB entries 4pmj and 1pyf are 165 and 156, respectively. The coordinates of Akr11a were used as a search model for molecular replacement (MR) using Phaser (McCoy *et al.*, 2007) from the CCP4 package (Winn *et al.*, 2011) to determine the initial phase. By deleting the non-aligned amino acids (Gly91–Pro100 and Thr219–Lys247) from the coordinates of Akr11a, a solution with RFZ = 4.1 and TFZ = 8.5 was obtained. After several cycles of refinement, a structure model with an R_{work} of 25.6% and an R_{free} of 28.3% was obtained.

The structure covers the full-length Tm1743 enzyme (Met1–Gly274); the vector-derived amino acids are not observed.

Figure 5
(a) The NADP⁺ cofactor bound in the active site of Tm1743. The overall crystal structure of Tm1743 is displayed in ribbon representation, with the NADP⁺ molecule shown as sticks in magenta. The $2F_o - F_c$ electron-density map of NADP⁺ is contoured at 1.0σ in blue. Conserved active-site residues are shown as sticks and the hydrogen bond between Asp53 and NADP⁺ is shown as a dashed line. This figure was prepared with PyMOL v.0.99 (DeLano, 2002). (b) The temperature-dependence of catalysis of the reduction of EOPB by Tm1743. The specific activity (U mg^{-1}) is plotted against temperature (K). Error bars represent the standard error of the mean of triplicate experiments.

Clear electron density for the cofactor NADP⁺ was observed in the active site of Tm1743 from the Tm1743–NADP⁺–EOPB sample (Fig. 5a). However, no EOPB molecule was found in the structure, although we incubated Tm1743 with NADP⁺ and EOPB before crystallization. NADP⁺ was also resolved in the structure model from the Tm1743–NADP⁺ sample. Thus, the crystal structure that we determined here is of Tm1743 in complex with its cofactor NADP⁺.

3.5. The specific activity of Tm1743 in reducing EOPB is temperature-dependent

The specific activity of Tm1743 in catalysing the reduction of EOPB was determined in order to investigate the underlying reasons for the lack of EOPB in the structure. The specific activity is temperature-dependent; it increased from 298 to 333 K and then slightly declined to about 4 U mg^{-1} at 358 K (Fig. 5b). The optimum activity was observed at 333 K, which is much higher than the crystallization temperature (298 K). The low specific activity of Tm1743 at 298 K could cause weak binding of EOPB in the active site. This structure of Tm1743 bound to NADP⁺ will be used as a template to study the enantioselectivity of Tm1743 and its application in biocatalysis through molecular-dynamics simulations and enzyme engineering.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 31400630, the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY14C050002, the Setup Foundation for Scientific Research from Hangzhou Normal University (Grant No. PE13002004007) and an HZNUARI-Pilot Research Grant (No. PD11001006007011). We thank the technical support of Chao Peng at the National Center for Protein Science (Shanghai) in the mass-spectrometric analysis.

References

- Anderson, S., Marks, C. B., Lazarus, R., Miller, J., Stafford, K., Seymour, J., Light, D., Rastetter, W. & Estell, D. (1985). *Science*, **230**, 144–149.
- Chen, W.-D. & Zhang, Y. (2012). *Front. Pharmacol.* **3**, 35.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Ehrensberger, A. H. & Wilson, D. K. (2004). *J. Mol. Biol.* **337**, 661–673.
- Ellis, E. M. (2002). *FEMS Microbiol. Lett.* **216**, 123–131.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Hahn-Hägerdal, B., Lindén, T., Senac, T. & Skoog, K. (1991). *Appl. Biochem. Biotechnol.* **28–29**, 131–144.
- Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Omura, S. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 9509–9514.
- Jez, J. M., Bennett, M. J., Schlegel, B. P., Lewis, M. & Penning, T. M. (1997). *Biochem. J.* **326**, 625–636.
- Jin, Y. & Penning, T. M. (2007). *Annu. Rev. Pharmacol. Toxicol.* **47**, 263–292.
- Kador, P. F. (1988). *Med. Res. Rev.* **8**, 325–352.
- Ma, Y.-H., Lv, D.-Q., Zhou, S., Lai, D.-Y. & Chen, Z.-M. (2013). *Biotechnol. Lett.* **35**, 757–762.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mindnich, R. D. & Penning, T. M. (2009). *Hum. Genomics*, **3**, 362–370.
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* **D62**, 859–866.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Ni, Y., Li, C.-X., Ma, H.-M., Zhang, J. & Xu, J.-H. (2011). *Appl. Microbiol. Biotechnol.* **89**, 1111–1118.
- Penning, T. M. (2015). *Chem. Biol. Interact.* **234**, 236–246.
- Penning, T. M., Ma, H. & Jez, J. M. (2001). *Chem. Biol. Interact.* **130–132**, 659–671.
- Schweiger, P., Gross, H. & Deppenmeier, U. (2010). *Appl. Microbiol. Biotechnol.* **87**, 1415–1426.
- Takikawa, H., Ookhtens, M., Stolz, A. & Kaplowitz, N. (1987). *J. Clin. Invest.* **80**, 861–866.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.