the FEBS Journal

Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries

Journal:	*FEBS Journal*
Manuscript ID:	Draft
Manuscript Type:	Regular Paper
Subdiscipline:	Enzymes and catalysis
Date Submitted by the Author:	n/a
Complete List of Authors:	Ferrandi, Erica; ICRM, CNR Sayer, Christopher; University of Exeter, Biosciences, College of Life and Environmental Sciences Isupov, Michail; University of Exeter, Biosciences, College of Life and Environmental Sciences Annovazzi, Celeste; ICRM, CNR Marchesi, Carlotta; ICRM, CNR Iacobone, Gianluca; ICRM, CNR Peng, Xu; University of Copenhagen, Department of Biology Bonch-Osmolovskaya, Elizaveta; Russian Academy of Sciences, Winogradsky Institute of Microbiology Wohlgemuth, Roland; Sigma-Aldrich, Research Specialties Littlechild, Jenny; University of Exeter, Biosciences, College of Life and Environmental Sciences Monti, Daniela; ICRM, CNR
Key Words:	Limonene-1,2-epoxide hydrolases, metagenomics, industrial biocatalysis, stereoselectivity, protein structure

Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries

Erica Elisa Ferrandi[1,3], Christopher Sayer[2], Michail N. Isupov[2], Celeste Annovazzi[1], Carlotta Marchesi[1], Gianluca Iacobone[1], Xu Peng[3], Elizaveta Bonch-Osmolovskaya[4], Roland Wohlgemuth[5], Jennifer A. Littlechild[2], and Daniela Monti[1]

1 Istituto di Chimica del Riconoscimento Molecolare, C.N.R., Milano, Italy

2 The Henry Wellcome Building for Biocatalysis, College of Life and Environmental Sciences, Biosciences, University of Exeter, Exeter, UK

3 University of Copenhagen, Copenhagen , Denmark

4 Winogradsky Institute of Microbiology of the Russian Academy of Sciences, Moscow, Russia

5 Sigma Aldrich, Buchs, Switzerland

Correspondence

Daniela Monti or Jennifer A. Littlechild, Istituto di Chimica del Riconoscimento Molecolare, C.N.R., Via Mario Bianco 9, 20131 Milano, Italy; The Henry Wellcome Building for Biocatalysis, College of Life and Environmental Sciences, Biosciences, University of Exeter, Stocker Road, Exeter EX4 4QD, UK.
Fax: +39 02 28901239; + 44 (0) 1392 723434
Tel: +39 02 28500025; + 44 (0) 1392 263468
e-mails: daniela.monti@icrm.cnr.it ; J.A.Littlechild@exeter.ac.uk

E.E.F. and C.S. contributed equally.

Abbreviations

EH, epoxide hydrolase; LEH, Limonene-1,2-epoxide hydrolases; *Re*-LEH, *Rhodococcus erythropolis* Limonene-1,2-epoxide hydrolases; *Mt*-LEH, *Mycobacterium tuberculosis* Limonene-1,2-epoxide hydrolases; PEG, poly(ethylene glycol).

Running title

Novel epoxide hydrolases from hot environments.

Database

The atomic coordinates and structure factors of the crystal structures have been deposited in the Protein Data Bank with the codes **5AIF** (Tomsk-LEH native structure), **5AIG** (Tomsk-LEH valpromide complex), **5AIH** (CH55-LEH native structure), **5AII** (CH55-LEH PEG complex).
Nucleotide sequence data are available in the GenBank databases under the accession numbers: **KP765711** (Tomsk-LEH) and **KP765710** (CH55-LEH).

Keywords

Limonene-1,2-epoxide hydrolases, metagenomics, industrial biocatalysis, stereoselectivity, protein structure

ABSTRACT

The epoxide hydrolases (EHs) represent an attractive option for the synthesis of chiral epoxides and 1,2 diols which are valuable building blocks for the synthesis of several pharmaceutical compounds. A metagenomic approach has been used to identify two new members of the atypical EH limonene-1,2-epoxide hydrolase (LEHs) family of enzymes. These two LEHs (Tomsk-LEH and CH55-LEH) show EH activity toward different epoxide substrates, differing in most cases from the previously identified *Rhodococcus erythropolis* (*Re*-LEH) in terms of stereoselectivity. Tomsk-LEH and CH55-LEH, both from thermophilic sources, have higher optimal temperatures and apparent melting temperatures than *Re*-LEH. The new LEH enzymes have been crystallized and their structures solved to high resolution in the native form and in complex with the inhibitor valpromide for Tomsk-LEH and in complex with polyethylene glycol for CH55-LEH. The structural analysis has provided insights into the LEH mechanism and the substrate specificity and stereoselectivity of these new LEH enzymes.

Introduction

Epoxide hydrolases (EHs) are enzymes that catalyze the hydrolysis of an oxirane (epoxide) ring by addition of a molecule of water, thereby forming a vicinal diol as a product [1,2]. They are ubiquitously expressed in all living organisms and play important physiological roles, *e.g.*, in the detoxification of reactive xenobiotics or endogenous metabolites and in the formation of biologically active mediators. Thus, mammalian EHs, such as microsomal epoxide hydrolases (mEHs, **EC 3.3.2.9**) and soluble epoxide hydrolases (sEHs, **EC 3.3.2.10**) are widely studied to gain an understanding of their function in regulatory processes and disease [3].

On the contrary, the interest of studying microbial EHs is mostly related to their potential use as biocatalysts for the production of optically pure epoxides and diols which are important synthons for the preparation of fine chemicals and drugs, for example the chiral β-blockers precursors [4,5]. In fact, these cofactor-independent enzymes can be applied in enantioselective kinetic resolution processes of racemic epoxides as well as in the stereoselective hydrolysis of *meso*-epoxides, in many cases with high selectivity and conversion yields [6]. Moreover, they are also capable of performing enantio-convergent hydrolytic processes through regio-complementary attack at the different oxirane carbon atoms, in the best cases catalyzed by the same enzyme [7].

As far as enzyme structure and catalytic mechanism is concerned, the majority of microbial EHs described to date from different sources (fungi, yeasts, bacteria) belong to the α/β-hydrolase superfamily, showing the typical fold of other hydrolytic enzymes such as lipases and proteases. The catalytic mechanism of this superfamily is highly conserved and uses an Asp-His-Asp (or Glu) catalytic triad which in EHs catalyzes the epoxide ring opening in two main steps through the formation of an alkyl-enzyme intermediate. As a distinct feature not found in other hydrolases, two highly conserved

tyrosine residues in the EHs play an essential role in the polarization of the epoxide ring, resulting in an increase in its reactivity [1].

More recently, a structurally different family of EHs has been discovered and named the limonene-1,2-epoxide hydrolases (LEHs, **EC 3.3.2.8**) according to the natural substrate of the first isolated member of this family [8,9]. Specifically, this enzymatic activity was first shown from the *Rhodococcus erythropolis* DCL14 strain able to grow on both (+)- and (-)-limonene as the sole carbon and energy source. The isolation and cloning of this enzyme and its subsequent biochemical and structural characterization has shown that *R. erythropolis* LEH (*Re*-LEH) does not belong to the α/β-hydrolase superfamily. This enzyme has a smaller subunit molecular weight (17 kDa) and has no sequence identity to the previously characterized LEHs. The X-ray structure of this enzyme shows it to have a completely different protein fold from other EHs, which resembles the nuclear transport factor NTF2 domain [10]. The catalytic mechanism is also different from that of the EHs belonging to the α/β-hydrolase superfamily. The LEH enzyme active site contained three residues (Asp^{101}, Arg^{99}, and Asp^{132}) that were proposed to act in a concerted fashion to activate a water molecule which is able to open the epoxide ring without the formation of a covalently bound alkyl-enzyme intermediate [10,11]. Investigations on the substrate specificity and stereospecificity of *Re*-LEH for the hydrolysis of the natural substrate limonene-1,2-epoxide and other related substrates showed sequential and enantio-convergent conversion of the substrate isomers [12]. Despite these attractive features, *Re*-LEH has been relatively unexploited in synthetic organic chemistry, due to its narrow substrate scope and lack of stereoselectivity [12], as well as its low thermal stability [9]. However, recent studies have demonstrated that these limitations can be overcome by protein engineering. For example, Zheng and Reetz significantly improved (up to > 99% enantiomeric excess) and even inverted the stereoselectivity of *Re*-LEH towards the model substrate cyclopentene oxide by using iterative saturation mutagenesis

[13]. In another recent study, the *Re*-LEH thermostability was significantly improved by screening computationally designed libraries, resulting in an increase of the melting temperature from 50 to 85°C [14].

In addition to protein engineering approaches, it might be expected that novel LEH variants with different properties could be found in other organisms. So far, only one LEH homologue, the enzyme Rv2740 from *Mycobacterium tuberculosis*, has been identified and studied [15]. Although the substrate specificity shown by Rv2740 was fairly different from that of *Re*-LEH, the available information about this enzyme is quite preliminary and no real application for its synthetic exploitation has been made. Sequence-based mining of (meta)genomes for novel EHs has been so far limited to the discovery of enzymes showing a α/β-hydrolase fold [16-18].

The potential of metagenomics, *i.e.*, the direct extraction and cloning of DNA from natural environments without culturing isolated microorganisms, to discover novel enzymes has been widely demonstrated in recent years [19-22]. Thermostable hydrolases with different activities, such as lipases, proteases and glycosidases, have already been identified from (hyper)thermophiles [23,24], however there is still no evidence for naturally thermostable epoxide hydrolases, or the occurrence of this enzymatic activity in hot environments.

A bioinformatics search on 10 Gbp of our metagenomic assemblies using known LEHs as queries has identified two ORFs with good homology to the *Re*-LEH. Here we present the cloning, biochemical and structural characterization of these novel LEHs from hot terrestrial environments from Russia and China growing at 46°C and 55 °C, respectively, and at neutral pH.

Results and discussion

Discovery and cloning of novel limonene-1,2-epoxide hydrolases from hot springs metagenomic libraries

Hot springs samples were collected by the partners of the Hotzyme project from thirteen different terrestrial environments (from China, Iceland, Italy, Russia, and Yellowstone National Park, USA), at temperatures ranging from 46°C to 92°C and pH between 1.8 to 10.2 (Table S1). Metagenomic DNA was extracted from the samples and sequenced using either Roche/454 Titanium FLX or Illumina HiSeq methods, thus affording about 10 Gbp of total sequence information, which was assembled into contigs for further analyses. As shown in a recent comparative metagenomic study including several of the aforementioned samples [25], the choice of the selected environmental sites has allowed for access to a wide biodiversity. This has included a global distribution of geographic locations at a variety of temperatures and pHs, as well as the presence of different organic substrates. Thus, a different distribution of thermophilic and hyperthermophilic microbial communities, belonging to both Archaeal and Bacterial domains, would be expected to be found in each sample.

Assembled contigs of the metagenomes were then submitted to the *in silico* screening of the target enzymes, *i.e.*, limonene-1,2-epoxide hydrolases (LEHs), using the program LAST (http://last.cbrc.jp/). Specifically, the coding sequences present in the metagenome assemblies were first traced and translated into the corresponding ORFs which were then submitted to multiple alignment with the two available amino acid sequences of characterized LEHs (query sequences), *i.e.*, those from *Rhodococcus erythropolis* (*Re*-LEH) (GenBank Q9ZAG3.3) and *Mycobacterium tuberculosis* (*Mt*-LEH) (GenBank CCP45539.1).

Only two sequences showing good similarity to the query sequences were found in the above metagenomes. These were in samples collected from the hot springs in the West Siberian Plain of Russia (Tomsk sample) and in the Yunnan region of China (Ch2-EY55S sample), at a geographic distance of about 4000 km. Interestingly, both samples were collected at around neutral pH and at moderate temperatures (46°C and 55 °C, respectively) in relatively organic rich environments, whereas no LEH homologous (or even partial sequences) were identified from metagenomes prepared from samples collected under more extreme conditions (very high temperatures and/or low/high pH). One possible explanation of this finding is related to the limited chemical stability of epoxides at high temperatures and extreme pH, thus making the production of EH enzymes unnecessary for the catabolism of environmental xenobiotics [1], since they can spontaneously hydrolyze in these extreme environments. However, since EH enzymes can play a role in the more protected intracellular metabolism, we cannot exclude that EH enzymes with low or no homology to the bacterial counterparts, could have evolved in the thermophilic Archaea. A more detailed description of the sampling sites of Tomsk and Ch2-EY55S samples is included in Supporting information (Description S1 and S2).

The 375 bp-long ORF found in the Tomsk metagenomic DNA, designated Tomsk-*limA*, encodes for a protein (Tomsk-LEH) of 125 amino acid residues, while the ORF (CH55-*limA*, 387 bp) identified in the metagenomic DNA from the Chinese sample, codes for a 129 amino acids-long protein (CH55-LEH). Sequence analysis revealed that both Tomsk-LEH and CH55-LEH display an identity of around 30% with the previously characterized LEHs. Specifically, Tomsk-LEH shows 30% identity with *Mt*-LEH and 25% identity with *Re*-LEH at the deduced amino acid level, while CH55-LEH shows 31% identity with both *Mt*-LEH and *Re*-LEH amino acid sequences. The two novel proteins share a 48% amino acid identity, demonstrating that they are two distinct enzymes, although related. A deeper

analysis in the BLASTp database confirmed the novelty of the two sequences as the closest relative to Tomsk-LEH is a hypothetical protein from *Porphyrobacter* sp. AAP82, showing only 53% identity and 95% query cover sequence, while in the case of CH55-LEH the best match is with a hypothetical protein from *Erythrobacter* sp. JL475 (84% identity, 98% query cover).

The multiple sequence alignment between the two novel LEH homologues and the *Re*-LEH and *Mt*-LEH sequences shows the presence of LEHs conserved amino acid residues (Fig. 1) [10,15].

<< *Insert Figure 1 here* >>

In particular, active site residues Arg99, Asp101 and Asp132 (amino acids numbers refer to the *Re*-LEH sequence), involved in the catalytic mechanism of epoxide ring opening, are present in both of the LEH homologues. In addition, as far as the amino acid residues involved in water activation are concerned, Asn57 is conserved in both sequences, while Tyr55 is present only in the CH55-LEH sequence. This information together with comparisons the novel proteins with previously characterized LEHs, suggest that the two identified enzymes could show epoxide hydrolase activity, but could have different functional properties in terms of substrate specificity and stereoselectivity.

Thus, the genes encoding the LEH homologues were amplified by PCR starting from the original Tomsk or Ch2-EY55S metagenomic DNA samples and were inserted into a suitable cloning vector for sequence confirmation and subsequent manipulation.

Recombinant expression in *E. coli* of the LEH homologues and functional characterization

The novel LEH-coding genes were cloned in the pRham expression vector in frame with the C-terminal His6x-tag sequence, and the His-tagged fusion proteins were successfully over-expressed in

Escherichia coli 10G in a soluble form. To allow direct comparison of the functional properties during enzyme characterization, the *Re*-LEH was expressed under similar conditions starting from its synthetic gene. Purification to homogeneity of the LEH homologues was subsequently achieved by Ni-NTA chromatography of cell extracts recovered from 1-L cultures. About 17 mg of recombinant Tomsk-LEH, 50 mg of CH55-LEH and 277 mg of *Re*-LEH were recovered, as assessed by protein determination and SDS-PAGE analysis (Fig. S1).

The functional characterization of the recombinant proteins was started by assessing their catalytic activity in the hydrolysis of selected epoxide substrates (Table 1).

<< *Insert Table 1 here* >>

Commercially available mixtures of *cis* and *trans* isomers of (4*R*)- and (4*S*)-limonene-1,2-epoxide ((+)-(**1**) and (-)-(**1**), respectively) were tested as substrates by monitoring the bioconversion reactions by chiral GC analyses at scheduled times. Both the Tomsk-LEH and CH55-LEH enzymes showed appreciable activity toward these substrates, demonstrating that the novel proteins are indeed functional epoxide hydrolases, but the maximum specific activity values were about two orders of magnitude lower than that estimated for *Re*-LEH and its preferred substrate, *i.e.*, (+)-(**1**). Therefore, it is likely that, differing from *Re*-LEH, (**1**) is not the natural substrate of the novel LEHs. This is consistent with the expected low probability of the occurrence of this organic compound at the geographical collection sites of Tomsk and Ch2-EY55S (Description S1 and S2). Nonetheless, in all cases the mixture of *cis* (1*R*,2*S*,4*R*) and *trans* (1*S*,2*R*,4*R*) isomers of (+)-(**1**) and the mixture of *cis* (1*S*,2*R*,4*S*) and *trans* (1*R*,2*S*,4*S*) isomers of (-)-(**1**) were quantitatively converted into the diaxial (1*S*,2*S*,4*R*)- and (1*R*,2*R*,4*S*)-

limonene-1,2-diols, respectively (Scheme S1). Thus, Tomsk-LEH and CH55-LEH showed the same enantio-convergent process previously described for *Re*-LEH [12].

Interestingly, the three LEHs showed a markedly different stereo-preference for the limonene-1,2-epoxide isomers. *Re*-LEH confirmed the previously reported strict preference for the *cis* form of (+)-(**1**) (86% diastereomeric excess (*de*) at 36% conversion) and for the *trans* form of (-)-(**1**) (45% *de* at 49% conversion) [12]. The Tomsk-LEH showed the opposite stereospecificity by preferring the *trans* isomer of (+)-(**1**) (55% *de* at 48% conversion) and the *cis* form of (-)-(**1**) (86% *de* at 41% conversion). The behavior of CH55-LEH was even more divergent from that observed with *Re*-LEH. In fact, in the case of (-)-(**1**) the fastest reacting diastereomer was definitely the *cis* form (85% *de* at 47% conversion), thus showing the same stereo-preference as Tomsk-LEH. Instead, both isomers of (+)-(**1**) were hydrolyzed by CH55-LEH with approximately the same rate (Fig S3-S4). The so-called sequential hydrolysis phenomenon, *i.e.*, the hydrolysis of the slowest reacting substrate only after complete consumption of the preferred one, which has been previously described for *Re*-LEH and other EHs [12,26], was evident only in those biotransformations with high diastereoselectivity, e.g., in the hydrolysis of (-)-(**1**) catalyzed by either Tomsk-LEH or CH55-LEH.

Subsequently, the performances of the three LEHs in the stereoselective hydrolysis of the *meso*-epoxides (**2**)-(**4**) were compared by monitoring the conversions and the enantiomeric excesses of the formed diol (ee_p) by chiral GC analyses. The three LEHs showed a marked preference for the C_6-ring derivative (**3**), CH55-LEH being about 6 times more active than the other homologues under the standard assay conditions, i.e., pH 8.0 and 20°C. Cyclopentene oxide (**2**) was the worst substrate for both *Re*-LEH and Tomsk-LEH and was not hydrolyzed by CH55-LEH. Interestingly, as far as stereoselectivity is concerned, the novel LEHs behaved differently than *Re*-LEH. In particular, a stereoselectivity switch was observed in the case of (**2**) with Tomsk-LEH and in the case of (**3**) with

both Tomsk-LEH and CH55-LEH, in all cases the (1*S*,2*S*)-diol being the preferred product. Moreover, for all the three *meso*-epoxides, the ee_p were higher with the novel enzymes than those obtained in the presence of *Re*-LEH.

Finally, we compared the behavior of the three LEHs in the kinetic resolution of racemic mixtures of the terminal epoxides (**5**) and (**6**) by determining the respective specific activity and the enantiomeric ratio (*E* value) [27]. Significantly higher enzyme activity was observed with the susbstrate having an aromatic ring, thus suggesting the possible presence of specific hydrophobic pockets in the active sites of each enzyme. Although at quite moderate levels, the hydrolysis of (**5**) was catalyzed by the novel LEHs with an appreciable enantioselectivity toward the (*S*) enantiomer, whereas the *Re*-LEH was not enantioselective (*E* = 1). Also in the case of substrate (**6**), the novel LEHs showed opposite enantioselectivity of *Re*-LEH, although again with quite low *E* values.

The hydrolysis reaction of cyclohexene oxide (**3**) was then chosen as a model reaction to assess the influence of pH and temperature on the activity of LEH homologues (Fig. 2).

<< *Insert Figure 2 here* >>

The specific activities (U/mg) of *Re*-LEH, Tomsk-LEH and CH55-LEH were compared in the pH range 5.0-9.5 (Fig. 2a) due to the limited chemical stability of (**3**), and in general of epoxide substrates, at lower and higher pH values. The novel LEHs showed a broader pH activity profile than the *Re*-LEH, being active in the whole range of tested conditions, while *Re*-LEH was not active at pH 5.0. Maximal specific activity was observed at pH 6.5 and 8.0 with Tomsk-LEH and CH55-LEH, respectively.

The effect of temperature on LEHs activity was tested by carrying out the hydrolysis reactions of (**3**) in the range 20-90°C at pH 8.0. As shown in Fig. 2b, the novel LEHs showed optimal activity, under the

used assay conditions, at higher temperature values than the *Re*-LEH (30°C), specifically at 40°C and 60°C. In particular, the activity of the CH55-LEH at 60°C was about 6 times higher (11500 U/g) than that observed at 20°C (Table 1), the latter being comparable to the specific activity estimated at 70°C. Additionally, to assess the thermostability of Tomsk-LEH and CH55-LEH, the apparent melting temperatures (T_M) were estimated by using the Thermofluor assay [28] in the presence of Sypro orange fluorescent dye (Fig. 3).

<< *Insert Figure 3 here* >>

The thermal shift analysis of the CH55-LEH defines clearly an apparent T_M of 79.7°C, which is, at the best of our knowledge, the highest T_M value estimated up to now for a natural epoxide hydrolase and significantly higher than that estimated for the *Re*-LEH (50°C) [14]. Instead, although the apparent T_M of the Tomsk-LEH is 74.5°C (Fig. 3a), the plot profile at lower temperatures suggests the possible occurrence of a gradual unfolding of the protein during the temperature gradient, which could be related to the lower optimal temperature.

Structural characterization of the LEH homologues

Both Tomsk-LEH and CH55-LEH show a dimeric oligomeric state under native conditions according to gel filtration analysis (Fig. S2).

Overall fold

Although both the Tomsk-LEH and CH55-LEH epoxide hydrolases lack around 20 residues at the N-termini when compared to the previously described *Re*-LEH [10], their ternary and quaternary

structures are all very similar to each other (Fig. 4). All three proteins form stable dimers, with the N-terminal extension of *Re*-LEH involved in the intersubunit interface which increases the buried surface area. The LEH monomer fold contains a curved six-stranded mixed β-sheet, with three α-helices packed on to its concave side to form the active site pocket.

<< *Insert Figure 4 here* >>

Active site

The LEH active site in Tomsk-LEH and CH55-LEH is predominantly hydrophobic as was previously observed in *Re*-LEH, with a cluster of polar/charged residues at the bottom of the active site cavity consisting of Asn34/34, Arg78/80, Asp111/112, Asp80/82 for the Tomsk-LEH and CH55-LEH enzymes respectively (Fig. 5a). These four residues are all conserved in the *Re*-LEH [10], the arginine and both aspartic acid residues forming the salt bridge network involved in the activation of the catalytic water [11].

<< *Insert Figure 5 here* >>

The LEH substrate binding pocket appears to have high affinity for polar molecules and additional electron density has been observed in the active site pocket in the different LEH structures. The active site of native Tomsk-LEH crystals contains an imidazole molecule coming from the nickel affinity chromatography. The active site of the CH55-LEH structure crystallized at higher pH contains a DMSO molecule which was used for solubilisation of the ligand. However the CH55-LEH crystallized at a low pH contains a bound polyethylene glycol (PEG; Fig. 5b) molecule from the crystallization

conditions, which does not bind in the similar conditions at higher pH due to the Asp82 (CH55-LEH) deprotonation.

Valpromide, an inhibitor of epoxide hydrolases [10], was found bound in several conformations in the active site of the co-crystallized complex of Tomsk-LEH crystals. The carboxamide group of valpromide is hydrogen bonded to Asp80 and to a catalytic water molecule (Fig 5c). These interactions are conserved in the previously reported *Re*-LEH-valpromide structure [10]. The catalytic water is also coordinated by NE1 of Trp32 in the Tomsk-LEH enzyme instead of OH of Tyr53 in the *Re*-LEH (Tyr32 in CH55-LEH). The rest of the active site is hydrophobic in nature consisting of several aromatic amino acids in both the Tomsk and CH55-LEHs.

A comparison of the position of α3 helix, which forms part of the substrate binding site, shows several differences between the three LEH enzymes. This helix adopts a different conformation in the CH55-LEH compared to both the Tomsk-LEH and *Re*-LEH (Fig. 6).

<< *Insert Figure 6 here* >>

In the CH55-LEH this helix, although located in a similar position, is displaced so that it does not superimpose with the equivalent helix in the other two LEH enzymes. This results in Gln52 in the CH55-LEH to point away from the LEH active site, in contrast to Phe53 (Tomsk-LEH) and Leu74 (*Re*-LEH) which point directly into the active site pocket. In the Tomsk-LEH the Phe53 adopts two positions to accommodate the binding of valpromide. The other main difference in this loop, which is also in close proximity to the active site, is the location of Phe49 (CH55-LEH), Ala49 (Tomsk-LEH) which are close to the substrate binding pocket compared to Thr70 in the *Re*-LEH enzyme. Overall

both these differences on the α3 helix result in a slightly more open and less hydrophobic active site in the *Re*-LEH.

Structure-based rationalization of functional properties

The LEH catalysis proceeds through attack by an activated catalytic water onto the epoxide carbon of the substrate which binds in the hydrophobic pocket with the epoxide oxygen bound to the sequentially first aspartic acid of the active site. As hydrolysis of both *cis* and *trans* isomers of compound (**1**) by the enzyme produces the same product (enantio-convergent reaction), the catalytic water molecule has to attack the most substituted carbon (C1) in the *cis* isomer of (+)-(**1**) and the less substituted carbon of its *trans* isomer (C2). This was shown by quantum chemical calculations[11] to be due to the influence of the conformation of the isopropenyl group of the substrate. This makes attack at either C1 or C2 energetically unfavorable for each of the different stereoisomers. Attack on the most substituted carbon (C1) is chemically preferable, therefore it is understandable why the *cis* (+)-(**1**) isomer and the *trans* (-)-(**1**) isomer are better substrates for *Re*-LEH which has a larger active site and can accommodate all four stereoisomers of compound (**1**). However, both the Tomsk-LEH and CH55-LEH demonstrate lower activity in comparison to *Re*-LEH with opposite stereoselectivity having a preference for attack at the C2 position of this substrate (Table 1). This is clearly due to the features of the more restricted active site in the Tomsk-LEH and CH55-LEH enzymes.

Four stereoisomers of compound (**1**) were modelled into the active sites of Tomsk-LEH and CH55-LEH so that the substrate carbon under attack (C1 in *cis* (+)-(**1**) and *trans* (-)-(**1**)) or (C2 *cis* (-)-(**1**) and *trans* (+)-(**1**)) was aligned between the catalytic water and the epoxide oxygen. This was positioned at the location of the valpromide nitrogen in the Tomsk-LEH complex structure which is H-bonded to Asp80/82. The substrates were then rotated around the epoxide bond subject to attack and re-orientated

to a position where no clashes with the active site residues were observed. This modelling approach revealed three steric factors which could restrict the catalytically competent binding of stereoisomers of compound (**1**) in the Tomsk-LEH and CH55-LEH enzymes.

The most noticeable difference between the Tomsk-LEH and CH55-LEH enzymes compared to the *Re*-LEH is that the residue Trp 62/60 (Tomsk-LEH/CH55-LEH) sterically hinders access of all stereoisomers of compound (**1**) to the side chain of the active site Asp80/82. This explains the significant decrease in activity compared to the *Re*-LEH which has a less bulky Val83 residue at this position.

The side chains of the residues Phe49/Phe53 of Ch55-LEH/Tomsk-LEH are in the same position as the smaller Leu74 in *Re*-LEH due to the relative movement of the α3 helix. This further limits the possibility of binding of the *cis* (+)-(**1**) isomer and the *trans* (-)-(**1**) isomer in both of the new LEHs, thereby changing their selectivity to the opposite of that observed with *Re*-LEH. The bulkier side chain of Trp32, which in Tomsk–LEH replaces Tyr32/53 of CH55-LEH/*Re*-LEH, further sterically hinders the binding of the *cis* (+)-(**1**) isomer and the *trans* (-)-(**1**) isomer. This would explain the stronger preference of the Tomsk-LEH for the stereoisomers of (**1**) which are attacked at the C2 position.

The PEG molecule was located bound in the CH55-LEH active site at low but not at high pH which indicates Asp82 is deprotonated at physiological pH. However for the catalytic reaction to proceed, this residue (Asp101 in *Re*-LEH) should be protonated to donate a proton to the epoxide oxygen as suggested by Hopmann et al. [11]. These authors further propose the protonation of this aspartic acid residue from bulk solvent and a preference of low pH for the reaction. The absence of activity of *Re*-LEH at pH 5.0 and the high optimal pH for CH55-LEH reported above (pH 8.0) does not fully support their theory. On the basis of the Tomsk-LEH valpromide complex reported here it seems that solvent molecules except the catalytic water are excluded from the hydrophobic LEH active site upon substrate

binding. It would appear more likely that the proton exchange proceeds from the catalytic water to the aspartic acid via the salt bridge network described above.

Conclusion

In this study a metagenomic approach has been successfully exploited to discover two novel EHs belonging to the relatively unexplored class of limonene-1,2-epoxide hydrolases. The new LEHs isolated from hot springs metagenomic libraries showed a different activity and stereoselectivity in the hydrolysis of a variety of epoxide substrates when compared to that reported for the previously described LEH from *R. erythropolis*. By using a metagenomics approach we have been able to access enzymes from organisms that are difficult or impossible to cultivate in the laboratory.

Interestingly, the two new LEH enzymes, which have been cloned and over-expressed in *E. coli*, have a higher thermal stability than any other EH enzyme that has been characterized to date that has been isolated from a natural environmental source. Both Tomsk-LEH and CH55-LEH have a melting temperature in excess of 70°C. These two new LEHs are also active within a broad pH range from 5.0 to 9.5 giving them properties that are more suitable for application in industrial biocatalysis.

The high resolution crystallographic structures of the Tomsk-LEH and the CH55-LEH have been compared to the previously studied *Re*-LEH structure. This has allowed us to obtain an insight into the basis for the different substrate specificity observed between these three related enzymes. One main difference between the new LEH enzymes compared to the *Re*-LEH is that a tyrosine residue in these enzymes sterically hinders access of all stereoisomers of limonene-1,2-epoxide (**1**) to the side chain of the active site aspartic acid residue required for catalysis. This explains the significant decrease in activity of the new LEHs towards this substrate when compared to the *Re*-LEH which has a less bulky valine residue at this position.

Additional information on enzyme mechanism has also been obtained from the Tomsk-LEH valpromide complex where solvent molecules with exception of the catalytic water are excluded from the hydrophobic active site upon substrate binding. It would therefore appear more likely that the proton exchange proceeds from the catalytic water to the aspartic acid by a salt bridge network rather than from bulk solvent as previously proposed.

This increased understanding of substrate specificity/stereoselectivity and the overall enzyme mechanism will pave the way for the rational design of mutant LEH enzymes with optimized specificity and selectivity and other improved properties that will be required for their application as new synthetic tools for the chemical and pharmaceutical industries.

Materials and Methods

Bacterial strains and materials

Racemic and enantiopure epoxides, diols, and rhamnose were from Sigma-Aldrich (St. Louis, USA) or Alfa Aesar (Karlsruhe,Germany). *Escherichia coli* TOP 10 [F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G] was purchased from Invitrogen (Darmstadt, Germany). *E. coli* 10G (mcrA Δ(mrr-hsdRMS-mcrBC) endA1 recA1 φ80dlacZΔM15 ΔlacX74 araD139 Δ(ara,leu) 7697 galU galK rpsL (StrR) nupG λ−tonA) was from Lucigen (Middleton, WI, USA). Tryptone and yeast extract were from Sigma-Aldrich (St. Louis, USA). All other reagents were of analytical grade and commercially available.

Analytical methods

Enantiomeric excesses of epoxides and (acetylated) diols and conversions were routinely determined by gas-chromatographic (GC) analyses on a AGILENT 6850 (Network GC System) gaschromatograph equipped with a chiral capillary column (MEGA DEX DAC-BETA , Legnano, Italy), having 0.25 mm-diameter, 25 m length and 0.25 μm-thickness, and with a Flame Ionization Detector (FID). The stereochemical outcome of the transformations was expressed as enantiomeric excess (*e.e.*) of the major enantiomer or as enantiomeric ratio (*E*) [27]. Optical rotations were determined on a Jasco P-2000 polarimeter (Cremella, IT). Details of analytic conditions, derivatization procedures and retention times are reported in Table S3.

General molecular biology techniques

Gene cloning into the pJet 1.2 vector was carried out with the CloneJET™ PCR Cloning Kit (Thermo Fisher Scientific, Waltham,USA). Plasmid DNA was purified by using the HiSpeed Plasmid Midi Kit from Qiagen (Hilden, Germany). DNA sequencing was performed by Bio-Fab Research (Rome, Italy).

Gene cloning in the pRham C-His Kan vector and *E. coli* 10G chemically competent cells transformations were carried out with the Expresso® Rhamnose Cloning & Protein Expression System (Lucigen, Middleton, WI, USA). If not stated otherwise, standard PCR amplifications were performed on 50 µL reaction mixtures containing plasmid DNA (10 ng) or metagenomic DNA (100 ng), primers (1 µM each), dNTPs (0.2 mM each), 4 U of XtraTaq Pol and 5 µL of the XtraTaq buffer (both from Genespin, Milan, Italy). PCR conditions were as follows: 95 °C for 2 min, followed by 40 cycles at 94 °C for 30 s, 55-76 °C for 30 s, 72 °C for 1 min, and then 72 °C for 10 min. Annealing temperatures vary according to primers melting temperatures.

Extraction, purification and sequencing of environmental DNA

Extraction, purification and sequencing of DNA from the environmental samples CH1102, Sun Spring, It-6, It-3, Ch2-EY65S, NL-10, Is3-13, Is2-5S, collected during the Hotzyme project is described in Menzel *et al*. (2015) [25]. Mw-2 and RC-2 samples were obtained following the same procedure described for the NL-10 sample, while recovery of environmental DNA from the Ch2-EY55S and Tomsk samples was performed as described for Ch2-EY65S and Sun Spring samples, respectively. After extraction of total genomic DNA, sequencing was carried out either by Roche/454 Titanium FLX or Illumina HiSeq according to Table S1.

***In silico* screening of metagenomic libraries and cloning of novel LEHs**

The results of the metagenomic sequencing of the environmental samples were assembled *de novo* into contigs using Meta-Velvet v.1.2. as previously described [25]. The resulting contigs were analyzed by using the ORF finder "getorf" program (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf.). Alignment of the query sequences (GenBank Q9ZAG3.3 and CCP45539.1) with the ORFs was performed using LAST (http://last.cbrc.jp/). The 375 bp-long ORF designated Tomsk-*limA* was found in the Tomsk metagenomic DNA (contig 165_6798_tomsk-sample-1_na, 1630 bp), while the ORF

CH55-*limA* (387 bp) was identified in the Ch2-EY55S metagenomic DNA (contig 5785 1630_Ch2-EY55S_meta-500, 6817 bp). Tomsk-LEH and CH55-LEH gene regions were amplified by PCR from the corresponding metagenomic DNA as a template in the presence of primers F1/R1 or primers F2/R2 (Table S2), respectively. PCR amplifications were carried out under the previously described standard conditions. The amplified fragments were subcloned into the pJet1.2 vector (Thermo Fisher Scientific, Waltham, USA) and the resulting plasmids pJetTomskLEH and pJetCH55LEH were transformed into *E. coli* TOP10. The cloned PCR amplicon was verified by DNA sequencing on both strands. The *Re*-LEH gene (GenBank CAC20854.1) was synthesized and cloned into the pUC57 vector by BaseClear (Leiden, The Netherlands).

Recombinant expression, purification and functional characterization of LEHs

The *Expresso*® Rhamnose Cloning and Protein Expression kit (Lucigen, Middleton, WI, USA) was used to construct expression vectors. Specifically, LEHs genes were amplified in the presence of plasmids pJetTomskLEH, pJetCH55LEH, and pUC57ReLEH as templates using primers which include 18 nt overlap with the ends of the pRham vector and that were suitable for the gene cloning in frame with the C-terminal His-tag (Table S2). PCR were performed under standard conditions and the amplified products were purified from the agarose gel using the Wizard SV Gel & PCR CleanUp System (Promega, Madison, WI, USA). Gene cloning into the pRham vector and transformations of *E. coli* 10G chemically competent cells were carried out according to the manufacturer's instructions. The resulting plasmids pRhamTomskLEH, pRhamCH55LEH, and pRhamReLEH were purified and sequenced. Enzyme expression was performed as follows: LB medium (40 mL) supplemented with 30 μg mL^{-1} kanamycin was inoculated with a single colony of recombinant *E. coli* 10G harboring one of the pRham plasmids developed as described above, and cultivated at 37 °C and 200 rpm overnight. The

culture was then diluted into 1 L of LB medium containing 30 μg mL^{-1} kanamycin and growth was propagated until an OD_{600} of 0.2-0.8 was reached. Subsequently the temperature was lowered to 30 °C, enzyme expression was induced with 10 mL of rhamnose 20% solution (w/v in water) and the culture was grown for a further 20 h. After recovery by centrifugation (5,000 rpm for 30 min, at 4°C), cells were resuspended in 10 mL of washing buffer (20 mM potassium phosphate (KP) buffer, pH 7.0, 500 mM NaCl, 20 mM imidazole) and lysed by ultrasonication. Cell debris was removed by centrifugation and the clear lysate was incubated with 8 mL of the Ni Sepharose 6 Fast Flow agarose resin (GE Healthcare, Milano, Italy) at 4°C under mild shaking. The resin was then loaded into a glass column (10 x 110 mm) and washed with 10 mL of washing buffer. Bound proteins were eluted using a 3 step gradient, washing the column with a mixture of washing buffer and elution buffer (20 mM KP buffer, pH 7.0, 500 mM NaCl, 500 mM imidazole) in 4:1, 3:2, and 2:3 ratio (8 mL each step), respectively. Fractions (2 mL) were collected and the protein concentration was determined. The samples showing the highest protein concentration were pooled and dialyzed overnight against 20 mM KP buffer, pH 7.2, at 4°C. Dialyzed samples were diluted with the dialysis buffer to a maximum concentration of 1 mg/mL to avoid protein precipitation and stored at -80°C. Molecular mass under native conditions was determined by comparison of LEHs retention times with those of standard proteins on a gel filtration analytical column (see Supplementary information Fig. S2). Enzyme purity was monitored by SDS-PAGE (17% T, 2.6% C) according to the method of Laemmli [29]. Protein bands were visualized by Coomassie blue staining and molecular mass under denaturing conditions was determined by comparison with standard markers (Bio-Rad, Hercules, CA, USA). Protein concentration was determined according to the method of Bradford [30] using the Bio-Rad Protein Assay and bovine serum albumin as a standard.

Biotransformations were carried out with the purified LEHs (0.25 mg of Tomsk-LEH or 0.125 mg of CH55-LEH or 0.002 mg of *Re*-LEH for the hydrolysis of (+)-(**1**) and (-)-(**1**), 0.25 mg of each LEH for the hydrolysis of (**2**) - (**5**), 0.125 mg of each LEH for the hydrolysis of (**6**)) in 1 mL KP buffer, pH 8.0, 10% (v/v) CH_3CN, containing 10 mM substrate, at 20°C. At scheduled time points, samples (100 µL) were extracted with an equal volume of a 0.025 mg/mL benzophenone solution in AcOEt in the presence of saturating NaCl and analyzed by chiral GC analyses (Table S3). Substrates and products peak areas were normalized to the internal standard benzophenone and concentrations were calculated using calibration curves obtained with authentic substrate/product standards (2.5-20 mM). One unit of activity (U) is defined as the enzyme activity that hydrolyzes 1 µmol of substrate per min under the assay conditions described above. Influence of pH and temperature on LEHs activity was evaluated by assaying the hydrolysis of (**3**) at pH values ranging from 6.5 to 9.0 and temperatures from 20 to 90°C. Spontaneous hydrolysis of (**3**) under different reaction conditions was taken into account and subtracted from enzymatic hydrolysis. Melting temperatures were determined by thermal shift experiments performed with a StepOne^{TM} Real-time PCR system and analysed using the Protein Thermal Shift^{TM} Software version 1.0 (Applied Biosystems). The final mixture composition consisted of 20 µL of 0.1 mg/mL protein, 5 mM K_2HPO_4/KH_2PO_4 buffer and 4 × SYPRO Orange dye (Invitrogen).

Protein Crystallization

The Tomsk-LEH and CH55-LEH enzymes were both concentrated to approximately 10 mg/mL using a 10 kDa membrane Vivaspin (Vivascience, MA, USA) and microbatch crystallization trials were set up using an Oryx 6 crystallization robot (Douglas Instruments, Hungerford, Berks, UK). The droplet

contained a 50:50 ratio of protein solution to precipitant solution and was covered with a 50:50 mix of silicon and paraffin oils.

Crystals of the Tomsk-LEH enzyme were grown in condition B11 of the Stura Footprint Screen (Molecular dimensions) consisting of 0.6 M sodium citrate, 5 mM sodium borate, pH 8.5. Crystals of the Tomsk-LEH enzyme complex were grown from the same condition in the presence of 10 mM valpromide dissolved in DMSO (final concentration 5%).

Crystals of the CH55-LEH enzyme were grown from F1 of the Morpheus screen (Molecular dimensions) consisting of 15 % PEG MME550, PEG20K, 60 mM monosaccharide mix and 50 mM MES buffer, pH 6.5. A second crystal form was grown from D11 from the Stura Footprint Screen consisting of 11.25 % PEG 10,000 and 50 mM ammonium acetate at pH 4.5.

Data collection and structure determination

Data were collected on beamlines I03 and I04-1 at the Diamond Synchrotron light source (Oxford, UK) under cryo conditions (100 K in a stream of gaseous nitrogen). Data were processed and scaled using XDS [31] and AIMLESS [32] in the Xia2 [33] pipeline. All further data and model manipulation was carried out using the CCP4 suite of programs [34].

The structure of the Tomsk-LEH enzyme was solved by molecular replacement using the program MOLREP [35]. The dimer of the *Re*-LEH (PDB code 1NWW) was used as a search model. The rotation and translation functions were calculated at 2.5 Å resolution. All potential space-groups were checked for with the solution being found in $P6_5$. The rotation solution peak height was 3.8 σ and the translation solution peak was 5.6 σ. The resulting MR solution was subjected to restrained refinement in REFMAC5 [36] resulting in an R-factor of 40.3 % and a R-free of 43.2 % after 30 cycles of refinement.

The structure of the CH55-LEH enzyme was solved using the previously solved Tomsk-LEH enzyme structure with the dimer being used as the search model. The rotation and translation functions were calculated at 2.0 Å resolution. The rotation solution peak height was 9.0 σ and the translation function 15.5 σ. The resulting MR solution was subjected to restrained refinement in REFMAC5 resulting in an R-factor of 32.3 % and a R-free of 38.3 % after 30 cycles of refinement. The asymmetric unit of structure of the CH55-LEH PEG complex contains 16 monomers which are organised as two independent dimers multiplied further by three pseudotranslation vectors. The original MR solution had a wrong choice of origin and did not refine. ZANUDA [37] was used to find a true origin out of four possible ones.

The Tomsk-LEH and CH55-LEH refined models were submitted to automated refinement using ARP/WARP [38]. The models were subsequently manually rebuilt in COOT [39] and refined with REFMAC5. The statistics of the data processing and refinement are given in Table 2.

<< *Insert Table 2 here* >>

After refinement the quality of the model was checked using the program PROCHECK [40]. Images were created using the molecular graphics program PyMOL [41]. Limonene epoxide stereoisomer models were built using JLIGAND [42] for further docking into the enzyme active site.

Acknowledgments

This work has been carried out in the framework of the HotZyme Project (http://hotzyme.com, grant agreement no. 265933) financed by the European Union 7th Framework Programme FP7/2007-2013, an EU FP7 Collaborative programme that aims to use a metagenomic approach to identify new thermostable hydrolases which have improved performances and/or novel functionalities for different industrial processes from diverse hot environments. D.M. would like to thank the "SusChemLombardia: prodotti e processi sostenibili per l'industria lombarda" project, Accordo Quadro Regione Lombardia-CNR, 16/07/2012 (protocol no. 18096/RCC) for additional support and Dr. Michael Kotik (AVCR, Prague, Czech Republic) for valuable suggestions during the assessment of assay methods. Work in the laboratories of J.A.L. is supported from the Biological and Biotechnology Research Council, UK and the University of Exeter, UK. The Exeter group would like to thank the Diamond Synchrotron Light Source, Reading, UK for access to beamline I03 and I04-1 (proposal No. MX8889) and the beamline scientists for their assistance. Sergey Gavrilov (INMI RAS) is thanked for the field work in the Tomsk region and Nikolai Ravin and Vitaly Kadnikov (the “Bioengineering” Center, RAS, Moscow, Russia) for the metagenomic sequencing of the thermophilic microbial mat. The authors would like to thank all of the partners of the Hotzyme project for all of their assistance and suggestions.

Author contributions

EEF and CS were equally responsible for the majority of the work. EEF performed the discovery, cloning and expression of the novel enzymes, while CS performed the crystallization and structure determination studies. MNI contributed to structure refinement and docking studies. CA performed the

biochemical characterization of functional properties. CM performed the substrate specificity studies. GI contributed to the set-up of GC analytical methods. XP and EB-O collected the environmental DNA, prepared and sequenced the metagenomic libraries from the Chinese and Russian samples, respectively. RW contributed to substrate profiling by suggesting and providing target epoxide substrates. JAL and DM wrote the manuscript with input from all authors.

References

1. Widersten M, Gurell A, Lindberg D (2010) Structure-function relationships of epoxide hydrolases and their potential use in biocatalysis. *Bioch Biophys Acta Gen Subj* **1800**, 316-326.
2. Kotik M, Archelas A, Wohlgemuth R (2012) Epoxide hydrolases and their application in organic synthesis. *Curr Org Chem* **16**, 451-482.
3. ElSherbeni AA, ElKadi AOS (2014) The role of epoxide hydrolases in health and disease. *Arch Toxicol* **88**, 2013-2032.
4. Nestl BM, Hammer SC, Nebel BA, Hauer B (2014) New generation of biocatalysts for organic synthesis. *Angew Chem Int Ed* **53**, 3070-3095.
5. Kong X-D, Ma Q, Zhou J, Zeng B-B, Xu J-H (2014) A smart library of epoxide hydrolase variants and the top hits for synthesis of (*S*)-β-blocker precursors. *Angew Chem Int Ed* **53**, 6641-6644.
6. Wohlgemuth R (2015) Epoxide hydrolysis. In: *Science of Synthesis, Biocatalysis in Organic Synthesis 2* (Faber K, Fessner W-D, Turner NJ, eds), pp 529-555. Georg Thieme Verlag KG, Stuttgart, NY.
7. Schober M, Faber K (2013) Inverting hydrolases and their use in enantioconvergent biotransformations. *Trends Biotechnol* **31**, 468-478.
8. Barbirato F, Verdoes JC, de Bont JAM, van der Werf MJ (1998) The *Rhodococcus erythropolis* DCL14 limonene-1,2-epoxide hydrolase gene encodes an enzyme belonging to a novel class of epoxide hydrolases. *FEBS Lett* **438**, 293-296.
9. van der Werf MJ, Overkamp KM, de Bont JAM (1998) Limonene-1,2-epoxide hydrolase from *Rhodococcus erythropolis* DCL14 belongs to a novel class of epoxide hydrolases. *J Bacteriol* **180**, 5052-5057.

10. Arand M, Hallberg BM, Zou J, Bergfors T, Oesch F, van der Werf MJ, deBont JAM, Jones TA, Mowbray SL (2003) Structure of *Rhodococcus erythropolis* limonene-1,2-epoxide hydrolase reveals a novel active site. *EMBO J* **22**, 2583-2592.
11. Hopmann KH, Hallberg BM, Himo F (2005) Catalytic mechanism of limonene epoxide hydrolase, a theoretical study. *J Am Chem Soc* **127**, 14339-14347.
12. van der Werf MJ, Orru RVA, Overkamp KM, Swarts HJ, Osprian I, Steinreiber A, de Bont JAM, Faber K (1999) Substrate specificity and stereospecificity of limonene-1,2-epoxide hydrolase from *Rhodococcus erythropolis* DCL14; an enzyme showing sequential and enantioconvergent substrate conversion. *Appl Microbiol Biotechnol* **52**, 380-385.
13. Zheng H, Reetz MT (2010) Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis. *J Am Chem Soc* **132**, 15744-15751.
14. Wijma HJ, Floor RJ, Jekel PA, Baker D, Marrink SJ, Janssen DB (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng Des Sel* **27**, 49-58.
15. Johansson P, Unge T, Cronin A, Arand M, Bergfors T, Jones A, Mowbray SL (2005) Structure of an atypical epoxide hydrolase from *Mycobacterium tubercolosis* gives insights into its function. *J Mol Biol* **351**, 1048-1056.
16. van Loo B, Kingma J, Arand M, Wubbolts MG, Janssen DB (2006) Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis. *Appl Environ Microbiol* **72**, 2905-2917.
17. Zhao L, Han B, Huang Z, Miller M, Huang H, Malashock DS, Zhu Z, Milan A, Robertson DE, Weiner DP, Burk MJ (2004) Epoxide hydrolase-catalyzed enantioselective synthesis of chiral 1,2-diols via desymmetrization of *meso*-epoxides. *J Am Chem Soc* **126**, 11156-11157.

18. Kotik M, Štěpánek V, Grulic M, Kyslìk P, Archelas A (2010) Access to enantiopure aromatic epoxides and diols using epoxide hydrolases derived from total biofilter DNA. *J Mol Cat B-Enzym* **65**, 41-48.

19. Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V, Batra N (2009) Metagenomics: concept, methodology, ecological inference and recent advances. *Biotechnol J* **4**, 480-494.

20. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P (2005) Metagenomic gene discovery: past, present and future. *Trends Biotechnol* **23**, 321-329.

21. Mocali S, Benedetti A (2010) Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res Microbiol* **161**, 497-505.

22. Wilson MC, Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* **20**, 636-647.

23. Littlechild JA, Guy J, Connelly S, Mallett L, Waddell S, Rye CA, Line K, Isupov M (2007) Natural methods of protein stabilization: thermostable biocatalysts. *Biochem Soc Trans* **35**, 1558-1563.

24. Yeoman CJ, Han Y, Dodd D, Schroeder CM, Mackie RI, Cann IK (2010) Thermostable enzymes as biocatalysts in the biofuel industry. *Adv Appl Microbiol* **70**, 1-55

25. Menzel P, Gudbergsdóttir SR, Rike AG, Lin L, Zhang Q, Contursi P, Moracci M, Kristjansson J K, Bolduc B, Gavrilov S, Ravin N, Mardanov A, Bonch-Osmolovskaya E, Young M, Krogh A, Peng X (2015) Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb Ecol*, DOI 10.1007/s00248-015-0576-9.

26. Rink R, Janssen DB (1998) Kinetic mechanism of the enantioselective conversion of styrene oxide by epoxide hydrolase from *Agrobacterium radiobacter* AD1. *Biochemistry* **37**, 18119 - 18127.

27. Chen CS, Fujimoto Y, Girdaukas G, Sih CJ (1982) Quantitative analysis of biochemical resolution of enantiomers. *J Am Chem Soc* **104**, 7294-7299.

28. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298.

29. Laemml, UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

30. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254.

31. Kabsch W (2010) XDS, *Acta Crystallogr* **D66**, 125-132.

32. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204-1214.

33. Winter G, Lobley CMC, Prince SM (2013) Decision making in xia2. *Acta Crystallogr D* **69**, 1260–1273.

34. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235–242.

35. Vagin A, Teplyakov A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr D* **66**, 22–25.

36. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long

F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* **67**, 355-367.

37. Lebedev AA, Isupov MN (2014) Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr D* **70**, 2430-2443.

38. Langer GG, Hazledine S, Wiegels T, Carolan C, Lamzin VS (2013) Visual automated macromolecular model building. *Acta Crystallogr D* **69**, 635-641.

39. Emsley P, Lohkamp B, Scott W G, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D* **66**, 486-501.

40. Laskowski RA, MacArthur MW, Moss DS Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**, 283–291.

41. DeLano WL (2002) The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

42. Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA, Murshudov GN (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D* **68**, 431-440.

43. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D* **55**, 191-205.

44. Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Mode*. **51**, 2778-2786.

Supporting information

Additional supporting information may be found in the online version of this article publisher's web site:

Description S1. Tomsk sampling site.

Description S2. Ch2-EY55S sampling site.

Fig. S1. Recombinant expression in *E. coli* and purification of LEH homologues.

Fig. S2. Gel filtration analysis of Tomsk-LEH and CH55-LEH under native conditions.

Fig. S3. Stereochemical analysis of (+)-limonene oxide hydrolysis catalyzed by LEHs.

Fig. S4. Stereochemical analysis of (-)-limonene oxide hydrolysis catalyzed by LEHs.

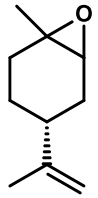
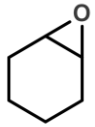
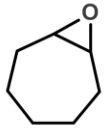
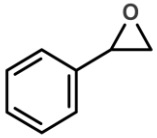
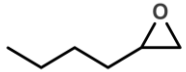
Table S1. Sampling sites and sequencing methods of samples collected during the HotZyme project.

Table S2. Primers used in this study.

Table S3. GC analytic conditions .

Tables

Table 1. Substrate scope and selectivity of the novel LEHs in comparison with *R. erythropolis* LEH (*Re*-LEH)

Substrate	**Specific activity (U/g)/ Selectivity**[a]						
	 **(+)-(1)**[b]	 **(-)-(1)**[b]	 **(2)**	 **(3)**	 **(4)**	 **(5)**	 **(6)**
***Re*-LEH**	30575 (*cis*) 2400 (*trans*)	1500 (*cis*) 6475 (*trans*)	4.5 ee_p (%) = 16 (*1R,2R*)	360 ee_p (%) = 7 (*1R,2R*)	50 ee_p (%) = 13 (*1S,2S*)	470 *E* = 1	60 *E* = 3 (*R*)
Tomsk-LEH	22 (*cis*) 200 (*trans*)	150 (*cis*) 4 (*trans*)	2.5 ee_p (%) = 33 (*1S,2S*)	320 ee_p (%) = 50 (*1S,2S*)	40 ee_p (%) = 67 (*1S,2S*)	150 *E* = 6 (*S*)	12 *E* = 2.6 (*S*)
CH55-LEH	293 (*cis*) 440 (*trans*)	365 (*cis*) 71 (*trans*)	-	2088 ee_p (%) = 39 (*1S,2S*)	200 ee_p (%) = 57 (*1S,2S*)	1230 *E* = 2.5 (*S*)	60 *E* = 2.5 (*S*)

[a)] Selectivity is indicated as enantiomeric excesses of the products (ee_p) in the case of *meso*-epoxides (**2**)-(**4**), and *E* values in the case of the racemic substrates (**5**) and (**6**) (see Supporting Information for details); [b)] substrates (+)-(**1**) and (-)-(**1**) were commercially available mixtures of *cis* and *trans* isomers of (4*R*)- and (4*S*)-limonene-1,2-epoxide, respectively.

Table 2: Summary of the data processing and refinement statistics

Crystal	Tomsk-LEH	Tomsk-LEH Valpromide complex	Ch55-LEH	CH55-LEH PEG complex
Beamline	I04-1	I03	I03	I03
Wavelength (Å)	0.9200	0.9763	0.9763	0.9763
Space Group	$P6_5$	$P6_5$	$C222_1$	$P2_1$
Unit Cell a,b,c (Å) α,β,γ (°)	96.4,96.4,57.8 90,90,120	96.5,96.5,56.9 90,90,120	91.5,100.8,56.9 90,90,90	69.0,104.2,148.0 90,102.4,90
Number of protein monomers per asymmetric unit	2	2	2	16
V_m ($Å^3Da^{-1}$)	2.75	2.71	2.37	2.35
Resolution range (Å)	41.7 - 1.26 (1.29-1.26)[a]	48.2 – 1.16 (1.19-1.16)	45.8-1.42 (1.46-1.42)	104.2-1.47 (1.51-1.47)
Multiplicity	10.1 (9.2)	9.7 (7.6)	6.3 (4.7)	3.4 (3.4)
Unique reflections	82450	104136	49831	336972
Completeness (%)	99.9 (99.1)	99.9 (99.4)	99.9 (99.3)	97.5 (94.2)
R_{sym}[b] (%)	5.7 (103.6)	6.4 (96.7)	4.7 (72.7)	4.3 (77.1)
I/σ (*I*)	20.4 (2.1)	15.9 (2.0)	19.4 (2.0)	18.3 (1.9)
Wilson B factor ($Å^2$)	18.5	16.0	24.3	25.5
Refined residues	248	248	250	2011
Refined water molecules	449	491	204	1830
Refined ligand* molecules	2	2	2	16
R_{cryst}[c] %	11.9	10.7	15.3	17.5
R_{free} % (5% of total data)	13.8	13.2	17.3	20.9
R.m.s.d. bond	0.008 [0.020][d]	0.012 [0.020]	0.011 [0.019]	0.013 [0.019]

lengths (Å)				
R.m.s.d. bond angles (^{o})	1.30 [1.93]	1.62 [1.94]	1.60 [1.97]	1.65 [1.98]
Occupancy of ligand	1.0	0.9	0.7-1.0	1.0
Average B factor (Å^2)				
protein	18.4	16.7	21.1	25.8
solvent	34.0	31.7	36.2	38.1
ligands	22.2	24.9	21.0	43.2
Ramachandran plot analysis, residues in (%)				
Most favoured regions	95.5	95.5	92.3	92.1
Additional allowed regions	4.5	4.5	7.7	7.8
Generously allowed regions	0	0	0	0.1
Disallowed regions	0	0	0	0

[a] Values in parentheses are given for the outer resolution shell.

[b] $R_{sym} = \Sigma_h \Sigma_J |<I_h> - I_J(h)| / \Sigma_h \Sigma_J I_J(h)$, where I(h) is the intensity of reflection h. Σ_h is the sum over all reflections and Σ_J is the sum over J measurements of the reflection.

[c] $R_{cryst} = \Sigma ||Fo| - |Fc|| / \Sigma |Fo|$

[d] Target values are given in square brackets.

* Only molecules within the active site cavity are considered ligands.

Wilson B-factor was estimated by SFCHECK [43]. Ramachandran plot analysis was performed by PROCHECK [40].

Figure legends

Figure 1. Multiple sequence alignment of Tomsk-LEH and CH55-LEH amino acidic sequences and those of other LEHs including those from *M. tuberculosis* (*Mt*-LEH) and *R. erythropolis* (*Re*-LEH). Catalytic residues are shown in green background, while residues involved in the interaction with the hydrolytic water are shown in light blue background.

Figure 2. Influence of pH (a) and temperature (b) on the activity of LEH homologues.

Figure 3. Determination of melting temperatures of Tomsk-LEH (a) and CH55-LEH (b) by the Thermofluor assay. For each sample, both the melting curve plot (upper side) and the derivative plot (bottom side) are shown.

Figure 4. The superposition of the three LEH dimer structures displayed as ribbons. *Re*-LEH is shown in black, Tomsk-LEH in blue and CH55-LEH in red.

Figure 5. LEH Active site. a) Superposition of active sites of the three LEHs with residues and the valpromide molecule shown as stick models. The *Re*-LEH is colored green, the Tomsk-LEH in magenta and the CH55-LEH in cyan. The catalytic water molecules are shown as red spheres. Hydrogen bonds are shown as black dashed lines. b) the electron density observed for the PEG molecule in the active site of CH55-LEH at low pH. The 2Fo-Fc map (blue) is contoured at 1.0 σ and the Fo-Fc map is contoured at 3.0 σ (green) and -3.0 σ (red). The amino acid residues are shown as

stick models. The solvent molecules are shown as red stars. c) key interactions of the valpromide inhibitor in the Tomsk-LEH active site demonstrate the hydrophobic nature of this enzyme active site. This figure was prepared by the program LigPlot+ [44].

Figure 6. Structural features responsible for the Tomsk-LEH and CH55-LEH stereoselectivity. The key residue side chains are shown as stick models and the protein backbone is shown as ribbons. The *Re*-LEH is colored green, Tomsk-LEH as magenta and the CH55-LEH as cyan. The catalytic water molecules are shown as red spheres.

Figures

Figure 1

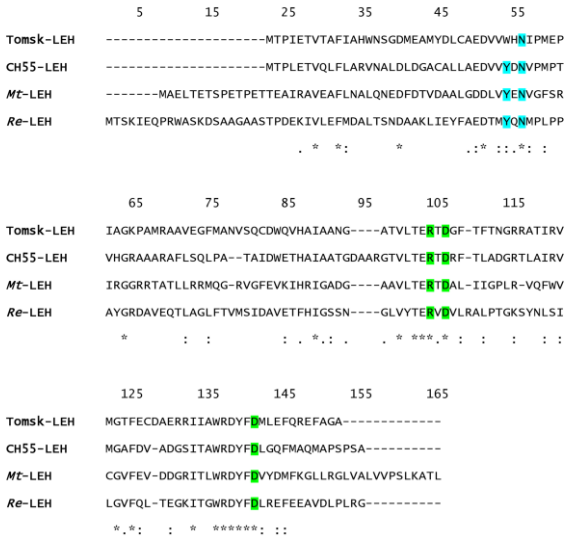


Figure 2

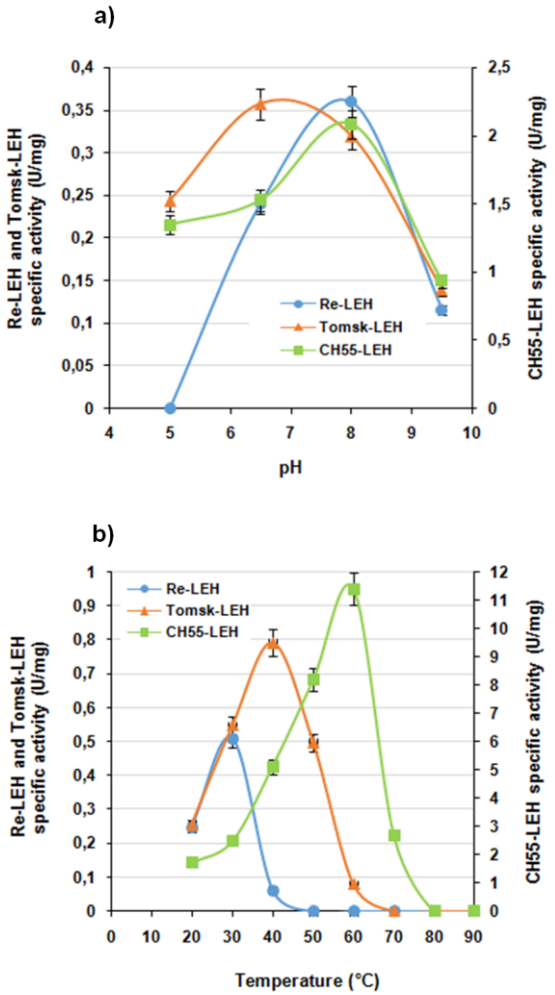

a)
Re-LEH and Tomsk-LEH specific activity (U/mg)
CH55-LEH specific activity (U/mg)
Re-LEH
Tomsk-LEH
CH55-LEH
pH
b)
Re-LEH and Tomsk-LEH specific activity (U/mg)
CH55-LEH specific activity (U/mg)
Re-LEH
Tomsk-LEH
CH55-LEH
Temperature (°C)

Figure 3

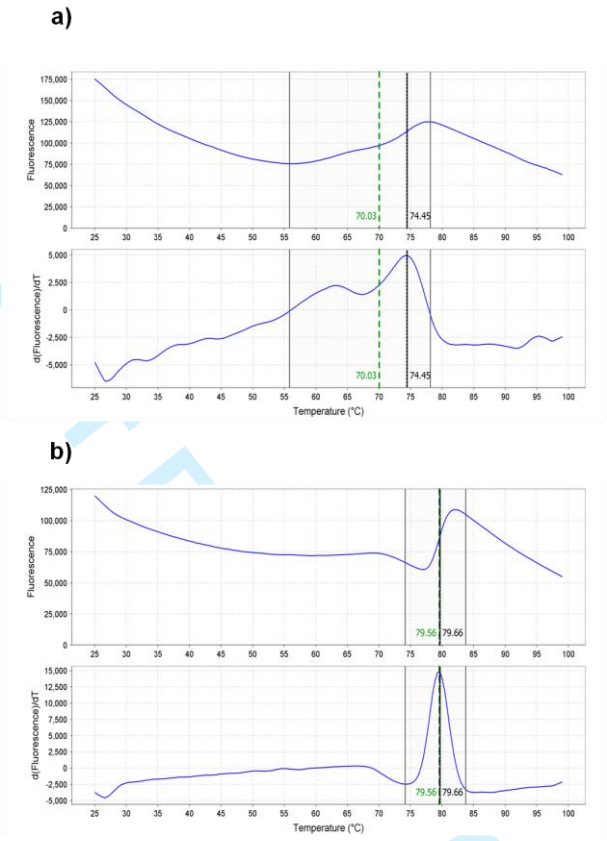

a)
Fluorescence
d(Fluorescence)/dT
70.03
74.45
Temperature (°C)
b)
Fluorescence
d(Fluorescence)/dT
79.56
79.66
Temperature (°C)

Figure 4

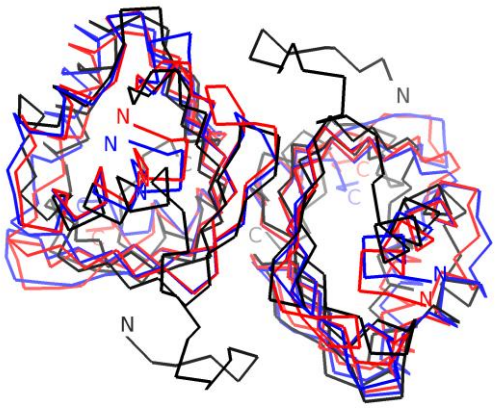


Figure 5

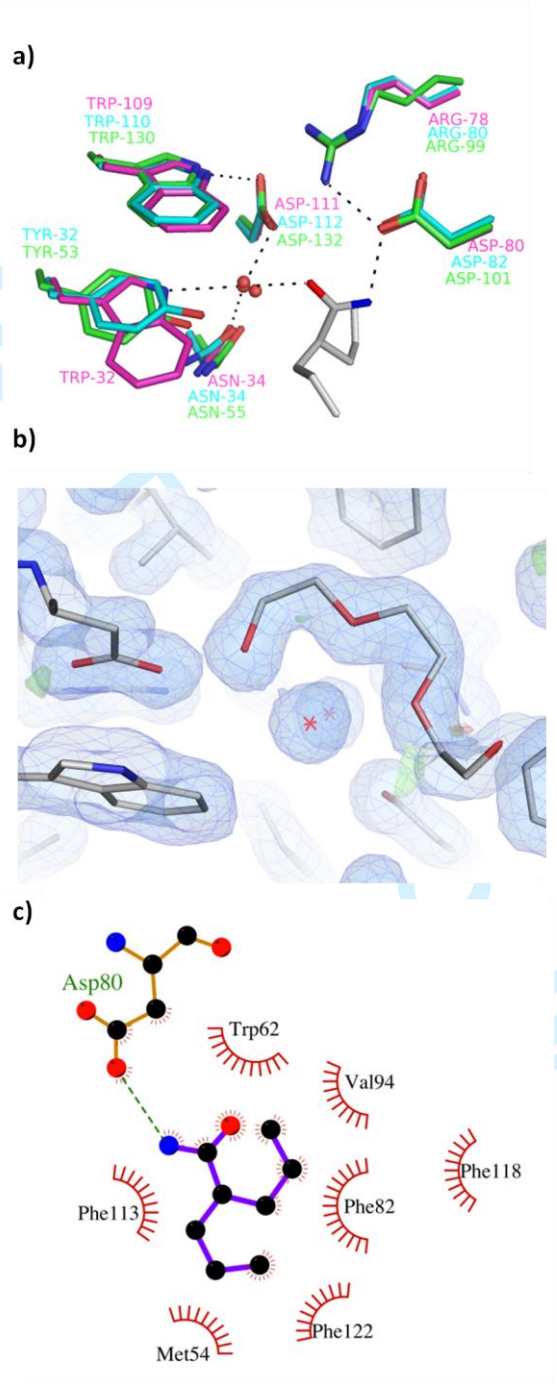

a)
TRP-109
TRP-110
TRP-130
ARG-78
ARG-80
ARG-99
ASP-111
ASP-112
ASP-132
ASP-80
ASP-82
ASP-101
TYR-32
TYR-53
TRP-32
ASN-34
ASN-34
ASN-55
b)
c)
Asp80
Trp62
Val94
Phe118
Phe113
Phe82
Phe122
Met54

Figure 6