



Subscriber access provided by Weizmann Institute of Science

## Structure-Guided Triple Code Saturation Mutagenesis: Efficient Tuning of the Stereoselectivity of an Epoxide Hydrolase

Zhoutong Sun, Richard Lonsdale, Lian Wu, Guangyue Li, Aitao Li, Jianbo Wang, Jiahai Zhou, and Manfred T. Reetz

ACS Catal., Just Accepted Manuscript • DOI: 10.1021/acscatal.5b02751 • Publication Date (Web): 22 Jan 2016 Downloaded from http://pubs.acs.org on January 28, 2016

## Just Accepted

Article

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Catalysis is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

# Structure-Guided Triple Code Saturation Mutagenesis: Efficient Tuning of the Stereoselectivity of an Epoxide Hydrolase

Zhoutong Sun<sup>1,2</sup>, Richard Lonsdale<sup>1,2</sup>, Lian Wu<sup>3</sup>, Guangyue Li<sup>1,2</sup>, Aitao Li<sup>1,2</sup>, Jianbo Wang<sup>1,2</sup>, Jiahai Zhou<sup>3,\*</sup>, Manfred T. Reetz<sup>1,2,\*</sup>

<sup>1</sup>Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

<sup>2</sup>Fachbereich Chemie der Philipps-Universität, Hans-Meerwein-Strasse, 35032 Marburg, Germany

<sup>3</sup>State Key Laboratory of Bio-organic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

**ABSTRACT:** Directed evolution of enzymes promises to eliminate the long-standing limitations of biocatalysis in organic chemistry and biotechnology - the often observed limited substrate scope, insufficient activity and poor regio- or stere-oselectivity. Saturation mutagenesis at sites lining the binding pocket with formation of focused libraries has emerged as the technique of choice, but choosing the optimal size of the randomization site and reduced amino acid alphabet for minimizing the labor-determining screening effort remains a challenge. Here we introduce structure-guided triple-code saturation mutagenesis (TCSM) by encoding three rationally chosen amino acids as building blocks in the randomization of large multi-residue sites. In contrast to conventional NNK codon degeneracy encoding all 20 canonical amino acids and requiring the screening of more than 10<sup>15</sup> transformants for 95% library coverage, TCSM requires only small libraries not exceeding 200–800 transformants in one library. The triple-code utilizes structural (X-ray) and consensus-derived sequence data, and is therefore designed to match the steric and electrostatic characteristics of the particular enzyme. Using this approach, limonene epoxide hydrolase has been successfully engineered as stereoselective catalysts in the hydrolytic desymmetrization of chiral substrates. Crystal structures and docking computations support the source of notably enhanced and inverted enantioselectivity.

**KEYWORDS**: amino acid alphabet, directed evolution, epoxide hydrolases, saturation mutagenesis, stereoselectivity, triple code.

## INTRODUCTION

Directed evolution of stereoselective enzymes constitutes a prolific source of biocatalysts for asymmetric reaction.<sup>1</sup> In this endeavor the most commonly applied gene mutagenesis methods are error-prone polymerase chain reaction (epPCR), DNA shuffling and saturation mutagenesis (SM).<sup>1</sup> SM at sites lining an enzyme binding pocket is a particularly powerful way to evolve stereo- and regioselective enzymes as catalysts in fine chemicals manufacturing.<sup>1a,e</sup> If the initial libraries fail to harbor hits showing optimal catalytic profiles, iterative saturation mutagenesis (ISM) can be invoked, which consists of consecutive rounds of SM. SM has been used in other important applications, including the degradation of phosphorusbased pesticides and chemical warfare agents,<sup>2,3</sup> and in potential universal blood production.<sup>4</sup> When evolving stereo- and/or regioselectivity, selection platforms<sup>5</sup> or FACS-based display systems,<sup>6</sup> normally capable of "handling" super large libraries (≈10<sup>10</sup>), are generally not possible,7 which means that screening assays need to be applied.7,8

Since screening is the labor-intensive step (bottleneck) when evolving enhanced or inverted stereoselectivity,<sup>1,8</sup> various attempts have been made for generating higherquality mutant libraries,<sup>1,9</sup> including the use of reduced amino acid alphabets in SM,<sup>10,10,11</sup> Application of the Patrick/Firth algorithm<sup>12</sup> or the Nov metric<sup>13</sup> enables the estimation of oversampling as a function of the size of the randomization site and of the number of amino acids used as building blocks<sup>10,10,11</sup>(Supporting Information Table S1). Extensive multi-residue sites in combination with large amino acid alphabets (up to 20 canonical amino acids as defined by NNK codon degeneracy) ensure maximal structural diversity, but such scenarios maximize the screening effort, e.g., up to  $\approx 10^{15}$  transformants in the case of a 10-residue site for 95% library coverage. To date, uncertainty persists concerning the optimal choice of a reduced amino acid alphabet and the number of residues in a randomization site. When opting for reduced amino acids in SM experiments, two different strategies can be applied: 1) Use of one and the same codon degeneracy for the whole multi-residue randomization site,<sup>10a-b</sup> or 2) use of a different codon degeneracy at each position of such a

60

In the present study we apply "triple code saturation mutagenesis" (TCSM) in SM of relatively large multi-residue randomization sites using strategy 1 as a practical strategy for generating focused high-quality libraries requiring minimal screening (Fig. 1). Utilizing structural, mechanistic and/or sequence<sup>10C</sup> data of an enzyme under study, a rational choice is made regarding three appropriate amino acids as building blocks in SM, in addition to the wildtype (WT) amino acid. Then, 10 residues surrounding the binding pocket are identified on the basis of the crystal structure followed by grouping them into typically 3or 4-residue randomization sites. This is necessary because application of TCSM at a 10-residue site requires the screening of 3.14 x 10<sup>6</sup> transformants for 95% library coverage. These choices constitute a viable "compromise" between structural diversity and amount of screening as shown by statistical analyses.12 Following SM and screening, ISM can be applied if necessary (Fig. 1). In order to illustrate the efficacy of this approach, we have applied the method to limonene epoxide hydrolase (LEH)<sup>14</sup> which has been characterized by X-ray crystallography.<sup>15</sup> The choice of this particular enzyme as the model system was made in view of previous directed evolution studies of LEH using other SM-based strategies,<sup>16,17</sup> thereby allowing direct comparison of the different approaches.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26 27

28

29 30 31

32

33

34

35 36

37

38

39 40

41

42

43

44

45

46 47

48

49

50

51

52

53

54

55

56

57

58

59 60



**Figure 1.** Steps in structure-guided triple code saturation mutagenesis (TCSM) as illustrated in the directed evolution of limonene epoxide hydrolase (LEH) as the catalysts in stereoselective transformations. The green letters in the parentheses (top) indicate the chosen triple code which defines the reduced amino acid alphabet used in SM of LEH. In this work, one and the same triple code (V-F-Y) was cho-

sen for each of the randomization sites: A (I80/V83/L14/I16, B (L74/M78/L147) and C (M32/L35/L103).

## RESULTS AND DISCUSSION

**Evolving enantioselective mutants of the epoxide hydrolase LEH.** As the model reaction we chose the hydrolytic desymmetrization of cyclohexene oxide (1) with formation of (R,R)- and (S,S)-2 (Scheme 1). WT LEH results in minimal enantioselectivity (ee = 4%) in favor of (S,S)-2.



**Scheme 1.** Hydrolytic desymmetrization of substrate 1 catalyzed by LEH.

Based on the crystal structure of LEH,<sup>15</sup> 10 residues lining the binding pocket were chosen for potential SM. Applying TCSM over such an extensive site would require the screening of about 10<sup>6</sup> transformants for 95% library coverage. Therefore, they were grouped into three randomization sites A (I80/V83/L114/I116, B (L74/M78/L147) and C (M32/L35/L103) (Fig. 2).



**Figure 2.** The 10 residues lining the binding pocket of LEH, assigned to three randomization sites A (blue), B (green) and C (yellow). Residues were selected based on docking of substrate **1** to the WT X-ray structure (PDB: 1NU<sub>3</sub>).<sup>15</sup>

At this point a decision had to be made concerning the choice of the three amino acids as building blocks in TCSM. We were primarily guided by the LEH crystal structure<sup>15</sup> which shows that the binding pocket is lined by residues having hydrophobic character. This suggests that amino acids with hydrophobic sidechains should be chosen, as in previous SM-based studies.<sup>16,17</sup> In order to gain further confidence, we applied the consensus approach in directed evolution<sup>18</sup> by first aligning 100 limonene epoxide hydrolases and then focusing on the respective 10 residues lining the corresponding binding pockets.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17 18

19

20

21

22

23

24

25

26

27

28

29

30 31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59 60

## **ACS Catalysis**

The dominance of hydrophobic amino acids was clearly confirmed (Fig. S1), such amino acids as valine, leucine, isoleucine and phenylalanine being typical. Tyrosine is not included in the list, but we considered it as well because its para-hydroxyl group could engage in H-bonding. In our previous study,<sup>16</sup> substitution using this amino acid had moderately positive effects.

Based on the sum of these information-based guides, we chose valine (V), phenylalanine (F) and tyrosine (Y) as the triple code. Randomization at sites A, B, and C requires a screening effort for 95% library coverage of only 576, 192 and 192 transformants, respectively.

Catalytically active variants in libraries A, B and C were rapidly identified by the adrenaline on-plate pretest,<sup>19</sup> followed by GC analysis for enantioselectivity of the active hits. Library A contained the best mutants as shown in **a**) Figure 3a, the data for the other two libraries are summarized in the SI (Table S2). It can be seen that three variants show (*S*,*S*)-selectivity in the range of ee = 96-99%, while the best (*R*,*R*)-selective variant I8oV/L114F is not quite as selective (ee = 89%). Therefore, the latter was used as a template for ISM at sites B and C employing the triple code Val/Phe/Tyr, which led to distinctly improved (*R*,*R*)-selective mutants (ee = 97%) (Fig. 3b). Kinetic data of the best mutants can be found in the SI Table S3. The total search required the screening of only 1728 transformants. As shown below in the Conclusions and Perspectives, the TCSM method is more efficient than previous approaches based on other SM techniques for engineering stereoselective LEH variants.<sup>16,17</sup>



**Figure 3.** (a) LEH variants originating from library A as catalysts in the model reaction of epoxide **1**. (b) Improved (*R*,*R*)-selective variants resulting from ISM at sites B (purple arrow) and C (light green arrow).

#### Understanding the source of evolved LEH enantiose-

lectivity. Structural and theoretical analyses of the evolved mutants provide detailed insight into the reshaping of the active site and the effect on substrate binding. X-ray structures were obtained for the greatly improved (S,S)-selective variant SZ348 (I8oY/L114V/I116V - 3.0 Å resolution) and the stereo-inverted (R,R)-selective variant SZ529 (M32V/M78V/I80V/L114F - 2.25 Å resolution). Structural comparison between the wild-type (WT, PDB code 1NU3) and the two mutants shows very little difference in tertiary structure (SI Fig. S2). The overall structural elements can be superimposed with a root-meansquare deviation (r.m.s.d.) of 0.24 and 0.688 Å for all atoms between the WT and the SZ348 and SZ529 mutants, respectively. The greater deviation in the case of SZ529 is mainly due to the C-terminal helix (residues 133-143 are 6residues shorter than the corresponding helix residues 133-149 in WT and SZ348), which has rotated by  $\sim$  35 degrees towards the active site pocket (Fig. 4a and SI Fig. S2). Figures 4c and 4d reveal that the variants SZ529 and SZ348 have undergone a significant change in shape of their active sites relative to WT LEH (Fig. 4b). These results clearly show that the correlation between reshaped binding pockets and altered stereoselectivity. Changes in hydrophobicity at the residues lining the binding pocket are shown in the SI (Fig. S<sub>3</sub>).



**Figure 4**. (a) An enlarged view of the structural alignments of WT, SZ529 and SZ348. The highlighted C-terminal helix was colored in magenta for WT, cyan for SZ529, and green for SZ348, respectively. Mutations resulted in 6 residues missing and a  $35^{\circ}$  rotation of the C-terminal helix in the

SZ529 structure. Alterations in the shape of the LEH binding pocket induced by SM-based directed evolution as revealed by crystal structures. Surface representation of the binding pockets of WT LEH (b) and variants SZ529 (c) and SZ348 (d) reveal the changes in shape; D101 marks the aspartate that is known to form an activating H-bond to the O-atom of epoxides.

In order to shed light on the origin of reversed enantioselectivity, substrate 1 was docked into the active site of the SZ529 and SZ348 crystal structures. In order for the substrate to undergo epoxide ring-opening by the active site water molecule, activation must occur via a hydrogen bond from the protonated D101 to the epoxide oxygen. For each of the mutants, only a single reactive docking pose was found in which this requirement is satisfied (as shown in Fig. 5). As discussed previously,<sup>15-17, 20</sup> the carbon atom of the epoxide ring at which nucleophilic attack occurs determines the stereochemistry of the diol product. In the case of both mutants, the epoxide carbon atom closest to the nucleophilic water molecule in the docking poses is consistent with the experimentally observed stereoselectivity. The positioning of Y80 in the SZ348 mutant (Fig. 5b) suggests that this residue may hydrogen bond to the epoxide oxygen, in addition to D101, further affecting the binding of the substrate in this mutant. Docking of substrate 1 was also performed to the WT enzyme: Two reactive docking poses were observed with similar scores, one favoring formation of (R, R)-2 and the other (S, S)-2 (as shown in SI Fig. S4). This is consistent with the low stereoselectivity observed for reaction of substrate 1 with WT LEH.



**Figure 5**. Highest ranked docking poses for substrate 1 in the crystal structures of the SZ529 (a) and SZ348 (b) mutants of LEH. Nucleophilic attack at C1 and C2 of substrate 1 results in formation of the (R,R)-2 and (S,S)-2, respectively. Geometric parameters for these complexes are displayed in SI Table S4.

**Expanding the substrate scope of LEH.** Some of the best variants evolved for substrate 1 were tested as catalysts in the hydrolytic desymmetrization of structurally different epoxides 3, 5 and 7 without performing any additional mutagenesis experiments (Scheme 2). It can be seen that in most, but not in all cases, high stereoselectivity was achieved (Table 1). The sterically smaller cyclopentene oxide (3) is not accepted by any of the variants, an observation that was made previously when testing LEH variants that had been evolved for desymmetrization of substrate 1.<sup>16</sup>



**Scheme 2.** Hydrolytic desymmetrization of further mesoepoxides as substrates using the best variants evolved for epoxide **1**.



**Scheme 3.** Hydrolytic kinetic resolution of styrene oxide (*rac-***9**) using the best variants evolved for epoxide **1**.

The best variants were also tested for hydrolytic kinetic resolution of styrene oxide (*rac-9*) (Scheme 3). All of them induce reversed enantioselectivity compared to WT, vari-

## ACS Catalysis

E = 43 (Table 2).

ant SZ351 reaching maximum (S)-selectivity amounting to

**Table 1.** Hydrolytic desymmetrization of different epoxides catalyzed by the best LEH variants evolved for substrate **1**.

	1		3		5		7	
Code	ee%	c%	ee%	c%	ee%	c%	ee%	c%
WT	4 ( <i>S</i> , <i>S</i> )	>99	2 (R,R)	80	17 ( <i>S</i> , <i>S</i> )	97	94 (R,R)	>99
SZ347	96 ( <i>S</i> , <i>S</i> )	91	nd	<5	81 ( <i>S</i> , <i>S</i> )	16	0.2 ( <i>R</i> , <i>R</i> )	7
SZ348	99 ( <i>S</i> , <i>S</i> )	97	10 ( <i>S</i> , <i>S</i> )	5	98 ( <i>S</i> , <i>S</i> )	81	4 ( <i>S</i> , <i>S</i> )	89
SZ351	97 ( <i>S</i> , <i>S</i> )	98	10 ( <i>S</i> , <i>S</i> )	5	93 ( <i>S</i> , <i>S</i> )	63	14 ( <i>S</i> , <i>S</i> )	93
SZ366	93 (R,R)	97	nd	<5	68 (R,R)	8	>99 ( <i>R</i> , <i>R</i> )	<5
SZ369	95 (R,R)	98	nd	<5	86 (R,R)	26	>99 ( <i>R</i> , <i>R</i> )	28
SZ380	96 (R,R)	98	nd	<5	95 (R,R)	74	>99 ( <i>R</i> , <i>R</i> )	73
SZ386	94 (R,R)	91	nd	<5	nd	<5	nd	<5
SZ388	96 (R,R)	>99	nd	<5	87 (R,R)	46	>99 ( <i>R</i> , <i>R</i> )	49
SZ389	97 (R,R)	>99	10 ( <i>R</i> , <i>R</i> )	<5	91 (R,R)	80	>99 ( <i>R</i> , <i>R</i> )	68
SZ390	95 (R,R)	96	nd	<5	93 (R,R)	5	nd	<5
SZ391	95 (R,R)	98	10 ( <i>R</i> , <i>R</i> )	<5	85 (R,R)	58	97 (R,R)	57
SZ398	96 (R,R)	98	nd	<5	95 (R,R)	36	nd	<5
SZ529	97 (R,R)	>99	10 (R,R)	8	94 (R,R)	83	>99 (R,R)	92

nd: not determined.

Table 2.	Hydrolytic	kinetic	resolution	of ep	oxide	rac- <b>9</b>		
with best LEH variants evolved for substrate 1.								

Code	ee%	c%	Favored enantiomer	Ε
WT	26	17	( <i>R</i> )	1.8
SZ348	92	15	( <i>S</i> )	28
SZ351	94	23	( <i>S</i> )	43
SZ386	39	36	( <i>S</i> )	2.8
SZ390	51	46	( <i>S</i> )	4.6
SZ391	44	20	( <i>S</i> )	2.9

## CONCLUSIONS AND PERSPECTIVES

The purpose of this study was to increase the efficacy of directed evolution of stereo- and regioselective enzymes. Lessons have been learned in terms of methodology development and in uncovering the reasons for enhanced and reversed enantioselectivity of LEH mutants. We conclude that triple code saturation mutagenesis (TCSM) is a simple and effective means to explore protein sequence space at large randomization sites lining the binding pockets of enzymes, provided guidance is sought from crystal structures, consensus sequence data and mechanistic considerations. The inherent knowledge-driven strategy constitutes a practical compromise between restricted structural diversity and screening effort. The results of the present study contrast with those of the initial report utilizing conventional ISM, in which 8 CAST residues in WT LEH were chosen for randomization and grouped into four 2-residue sites.<sup>17</sup> Unfortunately, such a scenario poses the problem of determining which of the 4! = 24 ISM pathways should be considered, since exploring all of them requires excessive screening.<sup>21</sup> Several pathways were arbitrarily chosen, and following the screening of 5,000 transformants, which is twice as much as in the present study, the best hits showed the following improved but not fully satisfactory catalytic profiles.<sup>17</sup> For epoxide **1** : 90% ee (*R*,*R*); 97% ee (*S*,*S*); epoxide **5** : 77% ee (*R*,*R*) versus 92% ee of WT LEH.

The present approach also stands in contrast to the most recent study in which a 10-residue site in LEH was randomized using a single amino acid alphabet in what can be called single code saturation mutagenesis (in that case valine), which required more than three times the screening effort for 95% library coverage (>3072 transformants) and provided only mediocre enantioselectivities (82-86% ee for (*S*,*S*)-variants and 70-76% ee for (*R*,*R*)-variants) in one and the same library.<sup>16</sup> About 828 additional transformants had to be screened to obtain variants that show 96% ee (*R*,*R*-2) and 92% ee (*S*,*S*-2). A comparison with the previously published results<sup>16</sup> based on single code saturation mutagenesis is presented in Table 3.

0		0	. ,		
Library (code usage)	Residues included in initial library	Library size for 95% coverage	Best ee obtained	Reference	
Valine	L74,F75, M78,I80, L103,L114, I116,F134, F139,L147	3072	86% (S,S) 76% (R,R)	16	
ISM		828	92% (S,S) 96% (R,R)	16	
A (V-F-Y)	180,V83, L114,1116	576	99% (S,S) 89% (R,R)	This work	
ISM		384	97% (R,R)	This work	

**Table 3.** Comparison of screening effort of this work with single code saturation mutagenesis used in previous study.

Full library coverage is not mandatory in any SM experiment, but it gives the researcher confidence that the best variants have not been missed.<sup>16,13</sup> TCSM-designed libraries are so small that they can be screened by automated GC or HPLC, not only for notably reduced library coverage (e.g., 50%<sup>10a,b</sup> corresponding to the 53<sup>rd</sup> best variant according to the Nov algorithm<sup>13,22</sup>), which would require considerably less screening, but even when aiming for essentially full coverage (95%). Thus, the bottleneck of directed evolution of enzymes is no longer the screening step. When an on-plate pretest is used for identifying active variants, as demonstrated in this study, the actual ee-screening effort is even less.

Relative to other methods such as epPCR or DNA shuffling, CAST-based SM defines pre-determined boundaries in the vast protein sequence space ensuring high-quality focused libraries. Among the various SM strategies presently available,<sup>ia,e,i0,11,16</sup> structure-driven TCSM appears to be a logical choice. The number of recursive cycles of mutagenesis/expression/screening is limited. We recommend that step-economy, a term borrowed from synthetic organic chemistry,<sup>23</sup> should be considered in future directed evolution studies more so than in the past. In future TCSM-based work, the choice of the triple code can also be guided by in silico design and screening.

Crystal structure analyses of stereoselective variants derived from directed evolution are rarely performed,<sup>1,16,24</sup> yet they provide a unique opportunity to analyze on a more sound basis the effect of the mutations on the catalytic profile. The crystal structures reported in this study, combined with docking computations, shed light on the origin of enhanced and reversed stereoselectivity and on the intricacies of the catalytic machinery of LEH.

Finally, we comment on the *hydrolytic* desymmetrization of meso-configurated epoxides as a synthetic method for accessing enantiopure vicinal diols in organic chemistry. This asymmetric transformation remains a challenge using chiral transition metal complexes such as the Jacobsen chiral cobalt-catalysts, because these do not work well when water is the nucleophile, in contrast to other O- or N-nucleophiles.<sup>25</sup> The same applies to chiral organocatalysts<sup>26</sup> The state of the art appears to be List's chiral organocatalysts based on BINOL-derived chiral phosphoric acids as catalysts using benzoate as the nucleophile at low temperatures (-5 to -40° C) and reaction times of several days.<sup>26c</sup> For example, in the desymmetrization of epoxides **1**, **5** and **7**, enantioselectivities of 93%, 91% and 82% were reported, as assessed by measuring the respective monobenzoates of the 1,2-diols. Base-mediated deprotection then provides the respective 1,2-diols. Since EHs do not deliver O-protected derivatives,<sup>27</sup> which may actually be desired in some circumstances, chiral organocatalysts and transition metal catalysts can be considered to be complementary in this type of asymmetric catalysis.

## MATERIALS AND METHODS

#### Materials

KOD Hot Start DNA Polymerase was obtained from Novagen. Restriction enzyme *Dpn* I was bought from NEB. The oligonucleotides were synthesized by Life Technologies. Plasmid preparation kit was ordered from Zymo Research, and PCR gel extraction kit was bought from QIAGEN. DNA sequencing was conducted by GATC Biotech. All commercial chemicals were purchased from Sigma-Aldrich, Tokyo Chemical Industry (TCI) or Alfa Aesar. Lysozyme and DNase I were purchased from AppliChem.

#### PCR based methods for library construction

Libraries were constructed using the Over-lap PCR and megaprimer approach with KOD Hot Start polymerase. 50 µL reaction mixtures typically contained 30 µL water, 5 µL KOD hot start polymerase buffer (10×), 3 µL 25 mM MgSO4, 5 µL 2 mM dNTPs, 2.5 µL DMSO, 0.5 µL (50~100 ng) template DNA, 100 µM primers mix 0.5 µL each and 1 µL KOD hot start polymerase. The PCR conditions for short fragment: 95 °C 3 min, (95 °C 30 sec, 56 °C 30 sec, 68 °C 40 sec) × 32 cycles, 68 °C 120 sec, 16 °C 30 min. For mega-PCR: 95 °C 3 min, (95 °C 30 sec, 60 °C 30 sec, 68 °C 5 min 30 sec) × 24 cycles, 68 °C 10 min, 16 °C 30 min. The PCR products were analyzed on agarose gel by electrophoresis and purified using a Qiagen PCR gel extraction kit. 2 µL NEB CutSmart<sup>™</sup> Buffer and 2 µL Dpn I were added in 50 µL PCR reaction mixture and the digestion was carried out at 37 °C for more than 3 h. After Dpn I digestion, the PCR products  $(1 \mu L)$  were directly transformed into electro-competent E. coliBL21(DE3) to create the final library for Quick Quality Control<sup>28</sup> and screening.

#### Primer design and Library creation

Primer design and library construction depend upon the particular amino acid chosen, and in the case of LEH this involves ten residues which were divided into three groups (SI Fig. S<sub>5</sub>): 1) Amplification of the short fragments of LEH using mixed primers F1/R1, F2/R2 and F3/R3 for Library A, B and C, respectively; 2) Amplification of the

whole plasmid pET22bLEHwt<sup>16</sup> using the products of step 1 as megaprimers, leading to the final variety plasmids for library generation. Primers are listed in SI Table S5. The PCR products were digested by *Dpn* I and transformed into electro-competent *E. coli* BL21(DE3) to create the library for screening.

Libraries A-B, A-C, A-B-C1, A-B-C2 and A-C-B were created using the same procedure as mentioned above. All the primers used are listed in SI Table S6.

## Screening Procedures

Colonies were picked and transferred into deep-well plates containing 300 µL LB medium with 50 µg/ml carbenicillin and cultured overnight at 37 °C with shaking. An aliquot of 120 µL was transferred to glycerol stock plate and stored at -80 °C. Then, 800 µL TB medium with 0.5% (m/v) lactose and 50 µg/mL carbenicillin was added directly to the culture plate for 8 h at 28 °C with shaking for protein expression. The cell pellets were harvested and washed with 400 µL 50 mM pH 7.4 potassium phosphate buffer and centrifuged for 10 min 4000 rpm at 4 °C. Then, the pellets were resuspended in 400 µL of the same buffer with 6 U DNase I and 1 mg/mL lysozyme for breaking the cell at 30 °C for 1 h with shaking. The crude lysate was centrifuged for 30 min 4000 rpm at 4 °C. 40 µL of the supernatant was used for an adrenaline assay,<sup>19</sup> then 110 µL potassium phosphate buffer (50 mM, pH 7.4) with 5% acetonitrile and substrate 1 (final concentration 10 mM) were added. The plates were incubated at 30 °C, 500 rpm, 3~4 h. Afterwards, NaIO4 solution (20 µL, 77 mg in 24 mL of water) was added and the plates were further incubated at 30 °C, 500 rpm, 10 min. Subsequently, adrenaline solution 20 µL (epinephrine 132 mg, water 24 mL, concentrated HCl (5 drops) for solubilizing the adrenaline) was added, which caused the immediate formation of a red color in the inactive reactions. Active variants gave colorless wells. 300 µL rest supernatant of the active transformants was transferred into new deep-well plates for reaction with 5 mM substrate 1 and 5% acetonitrile as cosolvent for 14~16 h at 30 °C 800 rpm, the final volume was 400 µL. The product and remaining substrate were extracted using equal volumes of ethyl acetate (EtOAc) for GC analysis by chiral column (SI Table S7). The screening results are shown in SI Table S2.

## Protein expression and purification

LEH and mutants SZ<sub>34</sub>8 and SZ<sub>529</sub> were inoculated into 250 mL shaking flasks of 40 mL LB containing 50  $\mu$ g/mL ampicillin and cultured overnight at 37 °C with shaking, and then scaled up to 4 flasks of 500 mL TB containing 50  $\mu$ g/mL ampicillin. When the OD600 reached to 0.8, 0.4 mM IPTG was added to induce the protein expression. The cell cultures continued to grow overnight at 16 °C before being harvested by centrifugation at 6,000 × g and resuspended in a Tris lysis buffer (25 mM, pH 8.0) containing 300 mM NaCl, 10 mM imidazole and 5 mM  $\beta$ mercaptoethanol. The cell pellets were disrupted by sonicator and the cell debris was removed by centrifugation at  $31,000 \times g$  for 40 min. The soluble protein sample was loaded onto a nickel affinity column (GE Healthcare) and washed with 10~500 mM imidazole solution containing 300 mM NaCl and 25 mM Tris-HCl (pH 8.0). Proteins from the flow through were pooled and concentrated, and then loaded onto a Superdex 75 Hiload 16/60 column (GE Healthcare) and eluted with 25 mM Tris-HCl (pH 7.5), 150 mM NaCl and 1 mM DTT. Fractions containing highly pure LEH proteins were pooled and concentrated with Centricon filtration devices (Amicon). The protein concentrations were determined by the Bradford procedure.

## Measurement of kinetic parameters

The enzymatic hydrolysis rate was measured by monitoring the conversion of substrate by GC using 1-heptanol as the internal standard. Pure enzymes were added to potassium phosphate buffer (50 mM, pH 7.4) with a total volume of 400  $\mu$  L containing epoxide of varied concentration (2-64 mM) and 2.5% (v/v) of acetonitrile, and the reaction was performed at 30° C for 10 min with shaking (800 rpm). The reaction was terminated by the addition of 400  $\mu$  L EtOAc containing 2.0 mM 1-heptanol. The mixture was vigorously mixed and the organic layer was separated by centrifugation at 4,000 rpm for 15 min. The EtOAc layer was analyzed by GC, the results are shown in Table S3.

## X-ray structural analysis

The SZ529 crystals were grown in 1.4 M sodium citrate and 0.1 M Hepes (pH 7.0) by the sitting-drop vapor diffusion method at 4 °C. The SZ348 variant was crystallized in 2.4 M sodium/potassium phosphate, o.1 M Tris-HCl (pH8.5) by the sitting-drop vapor diffusion method at 18 °C. Proteins (SZ529 at 21 mg/mL while SZ348 at 16 mg/mL) were mixed in a 1:1 ratio with the reservoir solution in a final volume of  $4 \mu L$  and equilibrated against the reservoir solution. Single crystals of SZ529 were obtained directly, whereas single crystals of SZ348 were achieved by adding additive such as 3% xylitol or 3% D-Sorbitol. All crystals were mounted in nylon loops and flash-frozen in liquid nitrogen. The cryoprotectant for SZ529 crystals contained 1.6 M sodium citrate, 0.1 M Hepes (pH7.0) and 30% sucrose while the cryoprotectant for SZ348 crystals consisted of 2.4 M sodium/potassium, 0.1 M Tris-HCl (pH8.5). Diffraction data of SZ529 were collected at the wavelength of 1.5418 Å on a Raxis IV<sup>++</sup> imaging plate detector at 100 K. Diffraction data of SZ348 was collected at the wave length of 0.97853 Å using a Pilatus detector at beamline BL19U of Shanghai Synchrotron Radiation Facility (SSRF) at 100 K. All data sets were indexed, integrated, and scaled using the HKL2000 package.<sup>29</sup> The structures of SZ529 and SZ348 were solved by molecular replacement method using the program PHASER<sup>30</sup> and the coordinate of wild-type LEH (PDB code 1NU<sub>3</sub>) as a search model. Rounds of automated refinement were performed with PHENIX<sup>31</sup> and the models were extended and rebuilt manually with COOT.32 The structures of SZ529 and SZ348 were refined to 2.25 and 3.0 Å, respectively. The statistics for data collection and crystallographic refinement are summarized in Table S8. All structural figures were prepared using Pymol (http://www.pymol.org/) except for Figure S3 which was drawn by Discovery Studio Visualizer.<sup>33</sup>

### Docking calculations

The crystal structure of WT LEH was used as the starting structure for the docking of substrate 1 to the WT (PDB entry 1NU3).<sup>15</sup> The apo crystal structures of the SZ529 and SZ348 mutants were used for docking to the respective mutants. The protein structures were prepared for docking using the Protein Preparation Wizard in Maestro.<sup>34</sup> The positions of all hydrogen atoms were energy minimized using Prime. Preparation of substrate 1 for docking was performed using LigPrep.<sup>35</sup> Docking was performed using Clide,<sup>36</sup> with a rigid protein and standard precision (SP) settings. The docking calculations were restricted such that a hydrogen bond must be formed between the substrate and the protonated D101. A maximum of 10 docking poses were allowed for each model, but no more than 8 were found for a given mutant.

#### ASSOCIATED CONTENT

**Supporting Information**: including figures and tables for multi-sequence alignment, library design and creation, primers list, X-ray structure, docking analysis, kinetic parameters measurement and GC analytical conditions. This material is available free of charge via the Internet at http://pubs.acs.org."

#### AUTHOR INFORMATION

#### **Corresponding Author**

\* Correspondence should be addressed to (M.T.R.) reetz@mpi-muelheim.mpg.de or (J.Z.) jiahai@mail.sioc.ac.cn.

#### ACKNOWLEDGMENT

Support from the Max-Planck-Society and the LOEWE Research cluster SynChemBio is gratefully acknowledged in addition to grants from Science and Technology Commission of Shanghai Municipality (15JC1400403).

#### REFERENCES

(1) Recent reviews of directed evolution: (a) Bommarius, A. S. Annu. Rev. Chem. Biomol. Eng. 2015, 6, 319-345; (b) Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B. Chem. Soc. Rev. 2015, 44, 1172-239; (c) Denard, C. A.; Ren, H.; Zhao, H. Curr. Opin. Chem. Biol. 2015, 25, 55-64; (d) Gillam, E. M. J.; Copp, J. N.; Ackerley, D. F. Methods in Molecular Biology, Humana Press, Totowa, 2014, Vol. 1179; (e) Goldsmith, M.; Tawfik, D. S. Methods. Enzymol. 2013, 523, 257-283; (f) Brustad, E. M.; Arnold, F. H. Curr. Opin. Chem. Biol. 2011, 15, 201-10; (g) Reetz, M. T. Angew. Chem. Int. Ed. 2011, 50, 138-74; (h) Jäckel, C.; Hilvert, D. Curr. Opin. Biotechnol. 2010, 21, 753-9; (i) Turner, N. J. Nat. Chem. Biol. 2009, 5, 567-73; (j) Lutz, S.; Bornscheuer, U. T. Protein Engineering Handbook, Wiley-VCH, Weinheim, 2009.

(2) Hill, C. M.; Li, W. S.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. J. Am. Chem. Soc. 2003, 125, 8990-8991.

(3) Gupta, R. D.; Goldsmith, M.; Ashani, Y.; Simo, Y.; Mullokandov, G.; Bar, H.; Ben-David, M.; Leader, H.; Margalit, R.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *Nat. Chem. Biol.* **2011**, *7*, 120-125.

(4) Kwan, D. H.; Constantinescu, I.; Chapanian, R.; Higgins, M.

A.; Kötzler, M. P.; Samain, E.; Boraston, A. B.; Kizhakkedathu, J.

N.; Withers, S. G. J. Am. Chem. Soc. 2015, 137, 5695-5705.

(5) Taylor, S. V.; Kast, P.; Hilvert, D. Angew. Chem. Int. Ed. . 2001, 40, 3310-3335.

(6) Yang, G.; Withers, S. G. Chembiochem. 2009, 10, 2704-2715.

(7) Acevedo-Rocha, C. G.; Agudo, R.; Reetz, M. T. J. Biotechnol. **2014**, *191*, 3-10.

(8) Reymond, J. L. Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting; Wiley-VCH, Weinheim , **2006**.

(9) Lutz, S.; Patrick, W. M. Curr. Opin. Biotechnol. 2004, 15, 291-297.

(10) (a) Reetz, M. T.; Carballeira, J. D. *Nat. Protoc.* **2007**, *2*, 891-903; (b) Reetz, M. T.; Kahakeaw, D.; Lohmer, R. *Chembiochem.* **2008**, *9*, 1797-1804 (c) Reetz, M.T.; Wu, S. *Chem. Commun.* **2008**, 43, 5499-5501.

(11) (a) Evans, B. S.; Chen, Y.; Metcalf, W. W.; Zhao, H.; Kelleher, N. L. Chem. Biol. 2011, 18, 601-607; (b) Chen, M. M. Y.; Snow, C. D.; Vizcarra, C. L.; Mayo, S. L.; Arnold, F. H. Prot. Eng. Des. Sel. 2012, 25, 171-178; (c) Sandström, A. G.; Wikmark, Y.; Engström, K.; Nyhlén, J.; Bäckvall, J.-E. Proc. Natl. Acad. Sci. USA. 2012, 109, 78-83; (d) Swe, P. M.; Copp, J. N.; Green, L. K.; Guise, C. P.; Mowday, A. M.; Smaill, J. B.; Patterson, A. V.; Ackerley, D. F. Biochem. Pharmacol. 2012, 84, 775-783; (e) Dudek, H. M.; Fink, M. J.; Shivange, A. V.; Dennig, A.; Mihovilovic, M. D.; Schwaneberg, U.; Fraaije, M. W. Appl. Microbiol. Biotechnol. 2014, 98, 4009-4020; (f) Wikmark, Y.; Svedendahl Humble, M.; Bäckvall, J.-E. Angew. Chem. Int. Ed. 2015, 54, 4284-4288; (g) Zhang, L.; Lu, L.; Fan, S.; Jin, L.; Gu, G.; Xu, L.; Xiao, M. RSC Adv. 2015, 5, 22361-22364; (h) Chuang, H.-Y.; Suen, C.-S.; Hwang, M.-J.; Roffler, S. R. Prot. Eng. Des. Sel. 2015, 28, 519-530.

(12) (a) Patrick, W. M.; Firth, A. E. *Biomol. Eng.* 2005, 22, 105-112;
(b) Denault, M.; Pelletier, J. N. In *Protein Engineering Protocols;*Arndt, K. M.; Müller, K. M., Eds.; Humana Press, NJ, 2007, Vol. 352, pp 127-154.

(13) Nov, Y. Appl. Environ. Microbiol. 2012, 78, 258-262.

(14) van der Werf, M.J.; Overkamp, K. M.; de Bont, J. A. M. J. Bacteriol. **1998**, 180, 5052-5057.

(15) Arand, M.; Hallberg, B. M.; Zou, J.; Bergfors, T.; Oesch, F.; van der Werf, M. J.; de Bont, J. A. M.; Jones, T. A.; Mowbray, S. L. *EMBO. J.* **2003**, **22**, 2583-2592.

(16) Sun, Z.; Lonsdale, R.; Kong, X.-D.; Xu, J.-H.; Zhou, J.; Reetz, M. T. *Angew. Chem. Int. Ed.* **2015**, 54, 12410-12415.

(17) Zheng, H.; Reetz, M. T. J. Am. Chem. Soc. 2010, 132, 15744-15751.

(18) Examples of the consensus approach to protein engineering: (a) Steipe, B.; Schiller, B.; Plückthun, A.; Steinbacher, S. *J. Mol.Biol.* **1994**, 240, 188-192; (b) Vázquez-Figueroa, E.; Chaparro-

Riggers, J.; Bommarius, A. S. *ChemBioChem.* **2007**, *8*, 2295-2301. (19) Wahler, D.; Reymond, J. L. *Angew. Chem. Int. Ed.* **2002**, 41,

(19) Wanier, D.; Reymond, J. L. Angew. Chem. Int. Ed. **2002**, 41, 1229-1232.

(20) Theoretical studies of LEH: (a) Hou, Q. Q.; Sheng, X.; Wang, J. H.; Liu, Y. J.; Liu, C. B. Biochim. Biophys. Acta Proteins Proteomics 2012, 1824, 263-268; (b) Lind, M. E. S.; Himo, F. Angew. Chem. Int. Ed. 2013, 52, 4563-4567.

(21) Gumulya, Y.; Sanchis, J.; Reetz, M. T. *ChemBioChem* **2012**, 13, 1060-1066.

(22) Hoebenreich, S.; Zilly, F. E.; Acevedo-Rocha, C. G.; Zilly, M.; Reetz, M. T. *ACS Synth. Biol.* **2015**, 4, 317-331.

(23) Wender, P. A.; Verma, V. A.; Paxton, T. J.; Pillow, T. H. Acc. Chem. Res. 2008, 41, 40-49.

1

2

3

4

22

23

24

25

26

(24) Recent examples of X-ray structures of stereoselective variants generated by directed evolution: (a) Pompeu, Y. A.; Sullivan, B.; Stewart, J. D. *ACS Catal.* **2013**, *3*, 2376-2390; (b) Reetz, M. T.; Bocola, M.; Wang, L.-W.; Sanchis, J.; Cronin, A.; Arand, M.; Zou, J.; Archelas, A.; Bottalla, A.-L.; Naworyta, A.; Mowbray, S. L. J. Am. Chem. Soc. **2009**, *131*, 7334-7343.

5 (25) (a) Jacobsen, E. N. Acc. Chem. Res. 2000, 33, 421-431; (b) 6 Schaus, S. E.; Brandes, B. D.; Larrow, J. F.; Tokunaga, M.; Han-7 sen, K. B.; Gould, A. E.; Furrow, M. E.; Jacobsen, E. N. J. Am. 8 Chem. Soc. 2002, 124, 1307-1315; (c) Jacobsen, E. N.; Kakiuchi, F.; 9 Konsler, R. G.; Larrow, J. F.; Tokunaga, M. Tetrahedron Lett. 10 1997, 38, 773-776; (d) Kawthekar, R. B.; Kim, G.-J. Helv. Chim. Acta 2008, 91, 317-332; (e) Ready, J. M.; Jacobsen, E. N. J. Am. 11 Chem. Soc. 2001, 123, 2687-2688; (f) Ready, J. M.; Jacobsen, E. N. 12 Angew. Chem. 2002, 114, 1432-1435. Angew. Chem., Int. Ed. 2002, 13 41, 1374-1377; (g) Private communication of Jacobsen, E. N. (Har-14 vard University, Cambridge/MA) regarding the performance of 15 Jacobsen-catalysts in the hydrolytic desymmetrization of meso-16 epoxides (2010). 17

(26) Reviews and key papers of organocatalyzed desymmetrization of meso-epoxides: (a) Wang, P. A.; Beilstein. J. Org. Chem.
2013, 9, 1677-1695; (b) Wang, Z.; Law, W. K.; Sun, J. Org. Lett.
2013, 15, 5964-5966; (c) Monaco, M. R.; Prévost, S.; List, B. Angew. Chem. Int. Ed. 2014, 53, 8142-8145.

(27) Recent reviews of EHs: (a) Wu, S.; Li, A.; Chin, Y. S.; Li, Z. ACS Catal. 2013, 3, 752-759. (b) Schober, M.; Faber, K.; Trends Biotechnol. 2013, 31, 468-478. (c) Kotik, M.; Archelas, A.; Wohlgemuth, R. Curr. Org. Chem. 2012, 16, 451-482; (d) Widersten, M.; Gurell, A.; Lindberg, D. Biochim. Biophys. 2010, 1800, 316-326; (28) (a) Bougioukou, D. J.; Kille, S.; Taglieber, A.; Reetz, M. T. *Adv. Synth. Catal.* **2009**, *351*, 3287-3305; (b) Acevedo-Rocha, C. G.; Kille, S.; Reetz, M. T. In *Methods in Molecular Biology*, Gillam, E. M. J.; Copp, J. N.; Ackerley, D. F. Eds.; Humana Press, Totowa, **2014**, Vol, *1179*, pp. 103-128.

(29) Otwinowski, Z.; Minor, W. *Methods in Enzymology*, Carter Jr, C. W.; Sweet, R. M. Eds.; Academic Press, New York, **1997**, Vol. 276, pp 307-326.

(30) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, 40, 658-674.

(31) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Crystallogr. D.* **2002**, 58, 1948-1954.

(32) Emsley, P.; Cowtan, K. Acta Crystallogr. D. 2004, 60, 2126-2132.

(33) Discovery Studio Modeling Environment, v. 4.0 (Accelrys Software Inc., San Diego, 2007).

(34) (a) Protein Preparation Wizard, Schrödinger Suite v. 2014-4 (Schrödinger, LLC, New York, 2014); (b) Epik v. 3.0 (Schrödinger, LLC, New York, 2014); (c) Impact v. 6.5 (Schrödinger, LLC, New York, 2014); (d) Prime v. 3.8 (Schrödinger, LLC, New York, 2014). (35) LigPrep, v. 3.2 (Schrödinger, LLC, New York, 2014).

(36) (a) Glide, v. 6.5 (Schrödinger, LLC, New York, 2014).
(36) (a) Glide, v. 6.5 (Schrödinger, LLC, New York, 2014); (b) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. J. Med. Chem. 2006, 49, 6177–6196; (c) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. J. Med. Chem. 2004, 47, 1739-1749.

