

## RESEARCH

## Open Access



# Exploring productive sequence space in directed evolution using binary patterning versus conventional mutagenesis strategies

Zhoutong Sun<sup>1†</sup>, Pamela Torres Salas<sup>1†</sup>, Elina Siirola<sup>1†</sup>, Richard Lonsdale<sup>1†</sup> and Manfred T. Reetz<sup>1,2\*</sup>**Abstract**

**Background:** Recent methodology development in directed evolution of stereoselective enzymes has shown that various mutagenesis strategies based on saturation mutagenesis at sites lining the binding pocket enable the generation of small and smart mutant libraries requiring minimal screening.

**Methods:** In this endeavor, limonene epoxide hydrolase (LEH) has served as an experimental platform, the hydrolytic desymmetrization of cyclohexene oxide being the model reaction with formation of (*R,R*)- and (*S,S*)-cyclohexane-1,2-diol. This system has now been employed for testing reduced amino acid alphabets based on the Hecht concept of binary patterning, with and without additional hydrophobic amino acids.

**Results and Conclusions:** It turns out that in binary pattern based saturation mutagenesis as applied to LEH, polar amino acids are seldom introduced. When applying binary patterning in combination with additional hydrophobic amino acids as building blocks in iterative saturation mutagenesis, excellent LEH variants were evolved for the production of both (*R,R*)- and (*S,S*)-diols (80–97 % ee), but again the introduction of polar amino acids occurs rarely. Docking computations explain the source of enhanced and inverted stereoselectivity. Some of the best variants are also excellent catalysts in the hydrolytic desymmetrization of other meso-epoxides, although both enantiomeric diols are not always accessible.

**Keywords:** Directed evolution, Reduced amino acid alphabet, Epoxide hydrolases, Saturation mutagenesis, Stereoselectivity

**Background**

With the current goal of enhancing the speed, efficacy and reliability of directed evolution of stereoselective enzymes, methodology development plays a central role (Bommarius 2015; Denard et al. 2015; Reetz 2011; Lutz and Bornscheuer 2009). The purpose is the generation of small mutant libraries which require minimal screening (bottleneck of directed evolution). In this research, saturation mutagenesis (SM), which is a stochastic process, occupies a leading position. Various forms of SM

have been developed over the years (Reetz 2011; Williams et al. 2014), one of the earliest examples involving the enhancement of oxidative stability (Estell et al. 1985). The first example of SM at sites lining the binding pocket of enzymes in the quest to increase enantioselectivity was not reported until 2001, but at that time only a single arbitrarily chosen randomization site was studied (Reetz et al. 2001). Later this process was generalized by considering many sites at or near the binding pocket (Reetz et al. 2005). It was dubbed Combinatorial Active-site Saturation Test (CAST) (Reetz et al. 2006; Acevedo-Rocha et al. 2014), a convenient acronym to distinguish it from SM at other sites remote from the binding pocket (Reetz 2011). We have pointed out that in the case of the usual NNK codon degeneracy encoding all 20 canonical amino acids as building blocks, protein sequence space

\*Correspondence: reetz@mpi-muelheim.mpg.de

<sup>†</sup>Zhoutong Sun, Pamela Torres Salas, Elina Siirola and Richard Lonsdale contributed equally to this work

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

Full list of author information is available at the end of the article

and the respective oversampling increase astronomically as the size of a randomization site increases from one residue to larger ones, e.g., 10. Consequently, two strategies have been proposed: (1) Splitting up a large site into several smaller ones, if necessary to be followed by iterative saturation mutagenesis (ISM) and/or (2) Utilization of reduced amino acid alphabets (Reetz 2011; Acevedo-Rocha et al. 2014).

Reduced amino acid alphabets have been used for other purposes, e.g., testing whether proper folds (Li et al. 2003) and enzyme activity (Walter et al. 2005; Akamura et al. 2002) are maintained when using less than the usual 20 canonical amino acids in protein construction, or when applying binary codes in the production of functional antibodies (Fellouse et al. 2005) or monoclonal antibodies (Koide et al. 2007). When applying SM in the directed evolution of stereo- and/or regioselectivity, consideration of oversampling is mandatory (Reetz 2011; Acevedo-Rocha et al. 2014). The required degree of oversampling for ensuring a defined library coverage that the researcher chooses, e.g., 95 %, can be calculated using the CASTER computer aid ([www.kofo.mpg.de/en/research/biocatalysis](http://www.kofo.mpg.de/en/research/biocatalysis)) (Reetz and Carballeira 2007), which is based on the Patrick/Firth algorithm (Patrick and Firth 2005). Alternatively, the Nov-metric allows the calculation of the *n*th best mutant as a function of the number of screened transformants (Nov 2012). It also needs to be pointed out that two ways for applying SM are possible: (1) Use of one and the same codon degeneracy for randomizing combinatorially the entire multi-residue site; or (2) Use of a different codon degeneracy at each individual position of a multi-residue site. Both types of SM are performed in a single experiment (Sun et al. 2016a).

For several years we have employed limonene epoxide hydrolase (LEH) (van der Werf et al. 1998; Arand et al. 2003) as the model reaction for methodology development in directed evolution, the hydrolytic desymmetrization of cyclohexene oxide (**1**) with formation of (*R,R*)- and (*S,S*)-**2** serving as the model reaction (Scheme 1) (Zheng and Reetz 2010; Sun et al. 2015, 2016b). Various reduced amino acid alphabets were tested, triple code saturation mutagenesis (TCSM) encoding three properly chosen

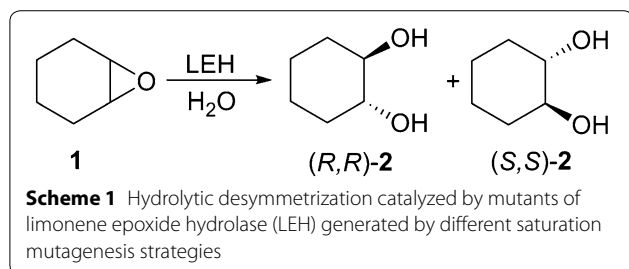
amino acids as building blocks (in addition to wildtype, WT) appearing to be a reasonable compromise between structural diversity and degree of screening (Sun et al. 2016b). In the present study we considered binary patterning in designing SM libraries and compared them to conventional SM libraries. Hecht and coworkers introduced the concept of binary patterning in the de novo design of proteins by devising a simple binary code of polar and nonpolar residues arranged in an appropriate order, thereby inducing the formation of globular  $\alpha$ -helical folds (Kamtekar et al. 1993; Roy and Hecht 2000). Specifically, we compared two different strategies for using reduced amino acid alphabets in library design for protein engineering of LEH: (1) Use of binary patterning based on choosing hydrophobic and hydrophilic amino acids (Strategy 1: libraries A and B); (2) Relying solely on the natural hydrophobicity of LEH's active site (Strategy 2: libraries C and D). In order to make proper comparisons with the previously reported strategies, we employed the same model reaction (Scheme 1).

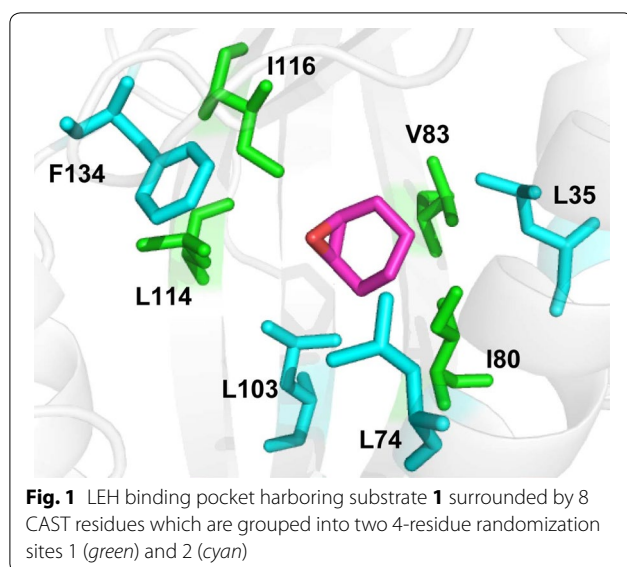
## Results and discussion

### Residue selection and library design

Similar to previous directed evolution studies of LEH (Sun et al. 2015, 2016b), the crystal structure of wildtype (WT) LEH (Arand et al. 2003) served as a guide for choosing CAST residues. In the present study, all together eight residues in close proximity of the active site of LEH (PDB 1NU3) were selected for SM, which were grouped in two randomization locations, sites 1 (I80, V83, L114 and I116) and site 2 (L35, L74, L103 and F134). Grouping the eight residues into two sites was performed arbitrarily, but residues located very close to each other in the primary structure of the enzyme (I80 and V83; I114 and I116, site 1) were included in the same site for easy library generation by PCR (Fig. 1).

Following strategy 1 for the construction of libraries based on sites 1 and 2, degenerate primers were designed to introduce high electronic variety (e.g., acidic, basic, and aliphatic residues): SDT (Leu, Val, His, Asp, Arg and Gly), and VDT (Leu, Ile, Val, His, Asn, Asp, Arg, Ser and Gly), and BWT (Phe, Tyr, Leu, Val, His and Asp). A single degenerate codon was planned for each position of the 4-residue randomization sites using these degenerate primers, which keeps the screening effort at a reasonable level. Consequently, SDT degeneracy was chosen for all of the targeted leucines (L35, L74, L103, L114) and one valine (V83), VDT for two isoleucines (I80, I116), and BWT for phenylalanine (F134). The details of all degenerate codons used at each position for different libraries are summarized in Table 1. Library A consists of the residues of site 1 and library B consists of the residues of site 2. To





ensure 95 % library coverage, the screening of  $2912 \times 3$  variants of library A and  $1294 \times 3$  variants of library B would be required. Experimentally, we assayed only 2912 and 1294, respectively, using the adrenaline on-plate pre-test for activity (Wahler and Reymond 2002) followed by chiral GC analysis of the active variants.

In the case of strategy 2, degenerate codons were selected for further libraries C and D, based on the natural hydrophobic character of the residues. Since seven out of eight residues to be randomized are aliphatic (Leu, Val or Ile), aliphatic residues Ala, Leu, Val, Ile and Gly were selected. Additionally, phenylalanine was included. In other words, all residues already existing in the selected positions were chosen, with Ala and Gly being added. The codons of choice facilitating the desired randomization were designed as GBT (B = G + C + T, for Val, Ala and Gly) and HTT (H = A + C + T, for Phe, Leu and Ile). In this way, the required screening effort for 95 % coverage of each library was calculated to be  $1294 \times 3$  variants for 95 % library coverage. Library C consisted of the residues of site 1 and library D of the residues of site 2, respectively.

### Characterization of mutants in libraries A and B

WT LEH catalyzes the hydrolysis of substrate **1** with full conversion, but the product proved to be essentially racemic, in slight favor of (*S,S*)-**2** (2 % ee). The best active catalyst in library A showing notably enhanced (*S,S*)-selectivity turned out to be L114H/I116 V (80 % ee; Table 2). Several active variants with reversed enantioselectivity were also identified. Out of these, the best was the double variant I80 V/V83L, which induces a relatively high level of selectivity for (*R,R*)-**2** (77 % ee). Interestingly, the (*S,S*)-selective variants have modifications at position 114 alone or at residues 114 and 116, whereas the (*R,R*)-selective variants are characterized by a mutation at position 80 alone or two mutations at the 80 and 83 positions (Table 2). It can be seen that a polar amino acid was utilized by application of the binary code only once, specifically in variant SZ211 (L114H/I116 V) which leads to an enantioselectivity of 80 % ee in favor of (*S,S*)-**2**.

Library B provided only variants with very low stereoselectivity in the range 22–37 % ee in slight favor of (*S,S*)-**2**. In hope of further optimization, two variants (I80 V/V83L and L114H/I116 V) were selected from library A for generating two A  $\rightarrow$  B libraries as part of ISM experiments. Out of 1472 variants of L114H/I116 V  $\rightarrow$  B library screened, 60 showed activity in the adrenaline assay (substrate concentration 10 mM). However, none of these showed improved stereoselectivity compared to L114H/I116 V. The same number of variants was screened in the case of the I80 V/V83L  $\rightarrow$  B library, but improvements in selectivity compared to the double mutant I80 V/V83L could not be detected.

### Characterization of mutants in libraries C and D

Several variants for both (*R,R*)- and (*S,S*)-**2** were identified in library C (Table 3). The best (*R,R*)-selective variant turned out to be I80 V/V83I/L114 V (ee = 80 %), and the best (*S,S*)-variant was identified as I80F/V83I/L114 V/I116 V (ee = 98 %). Interestingly, I80F/V83I constitutes an important couple, since it occurs in several good variants. When combined with additional L114 V, the enzyme favors the formation of (*R,R*)-**2**. When I116 V

**Table 1** List of degeneration codon used in each position of different libraries

Library	Degenerate codon	Alphabets	Library	Degenerate codon	Alphabets
Library A (Site 1)	I80: VDT V83: SDT L114: SDT I116: VDT	VDT: Leu, Ile, Val, His, Asn, Asp, Arg, Ser, Gly; SDT: Leu, Val, His, Asp, Arg, Gly BWT: Phe, Tyr, Leu, Val, His, Asp	Library C (site 1)	I80: GBT/HTT V83: GBT/HTT L114: GBT/HTT I116: GBT/HTT	GBT: Val, Ala, Gly HTT: Phe, Leu, Ile
Library B (Site 2)	L35: SDT L74: SDT L103: SDT F134: BWT		Library D (Site 2)	L35: GBT/HTT L74: GBT/HTT L103: GBT/HTT F134: GBT/HTT	

**Table 2 Best LEH variants of libraries A and B as catalysts in the hydrolytic desymmetrization of epoxide 1 with formation of (*R,R*)- or (*S,S*)-2**

Library	Code	Mutations	Favored enantiomer	ee %	c %
A	WT		( <i>R,R</i> )	2	99
	SZ133	I80L	( <i>R,R</i> )	56	81
	SZ197	I80V	( <i>R,R</i> )	72	93
	SZ210	I80V/V83L	( <i>R,R</i> )	77	96
	SZ211	L114H/I116V	( <i>S,S</i> )	80	95
	SZ212	L114G/I116V	( <i>S,S</i> )	74	77
	SZ546	L114V/I116V	( <i>S,S</i> )	70	95
B	SZ717	L35V	( <i>S,S</i> )	22	86
	SZ213	L35V/L74V	( <i>S,S</i> )	37	28

**Table 3 Best LEH variants of libraries C and D as catalysts in the hydrolytic desymmetrization of epoxide 1 with formation of (*R,R*)- or (*S,S*)-2**

Library	Code	Mutations	Favored enantiomer	ee %	c %
C	WT		( <i>S,S</i> )	2	99
	SZ57	I80V/L114V	( <i>R,R</i> )	75	99
	SZ197	I80V	( <i>R,R</i> )	72	93
	SZ503	I80V/V83I/L114V	( <i>R,R</i> )	80	97
	SIF01	I80F/V83I/L114V	( <i>R,R</i> )	72	94
	SZ501	I80F/V83I/L114V/I116G	( <i>S,S</i> )	97	98
	SZ502	I80F/V83I/L114V/I116V	( <i>S,S</i> )	98	95
	SZ504	I80F/V83I/L114V/I116V	( <i>S,S</i> )	95	98
	SZ546	L114V/I116V	( <i>S,S</i> )	68	95
	D	TOQ-PA-197-04	L103I/F134L	( <i>R,R</i> )	40
TOQ-PA-197-06		L74F/L103I/F134L	( <i>R,R</i> )	55	35
TOQ-PA-197-01		L35V/L74F/L103V	( <i>S,S</i> )	36	92
TOQ-PA-197-05		L35V/L74F	( <i>S,S</i> )	38	67
TOQ-PA-197-07		L35A/L74F/L103I	( <i>S,S</i> )	38	62

or I116G is added, enantioselectivity reverses with preferential formation of (*S,S*)-2 (up to 98 % ee). In further experiments, I80V/V83I/L114V was selected for an ISM round to improve stereoselectivity. In this case, 1288 clones were submitted to the adrenaline assay. Some of the variants showed similar activity, but none proved to be more efficient, although very high enantioselectivity was detected. The best variants are listed in Table 4.

The results originating from library D are disappointing (Table 3), 1288 clones being submitted to the adrenaline assay (288 active). The best (*S,S*)-variant was identified as L35V/L74F (ee = 38 %), while the best (*R,R*)-selective variant proved to be L74F/L103I/F134L (ee = 55 %). Upon analyzing these results more closely, it seems that positions L103/L134 are an important couple for obtaining (*R,R*)-2 variants. In contrast, mutations at residue L35 appear to be important for obtaining good (*S,S*)-2 variants. Subsequently, mutant L74F/L103I/F134L, as the best (*R,R*)-2 performer, was chosen as starting point for an ISM round with creation of library DC1. In this case 1288 clones were submitted to the adrenaline assay. Although 232 of them were active, just one showed a slight improvement in enantioselectivity (ee = 61 %). Similarly, L35V/L74F was used as a template for preparing library DC2. Gratifyingly, out of 1288 clones submitted to the adrenaline assay, 664 were active, which is a very high score. Some of improved hits showed enantioselectivities in the range 80–90 % ee (Table 4). It is apparent that only hydrophobic amino acids (and in some cases additionally glycine) were introduced. Mutant SZ288 having five point mutations (L35V/L74F/V83T/L114V/I116V) shows high (*S,S*)-selectivity (90 % ee), but the only polar amino acid (V83T) was introduced randomly in the PCR process, and is not part of the library design.

Finally, it was of interest to explore the crossover between two different strategies regarding sets of degenerate codon usage. Accordingly, libraries C to B (library CB) and A to D (library AD) were tested based on the common variant I80V found in both libraries C and A, respectively. However, no improved hits were discovered in either library (Additional file 1: Table S1).

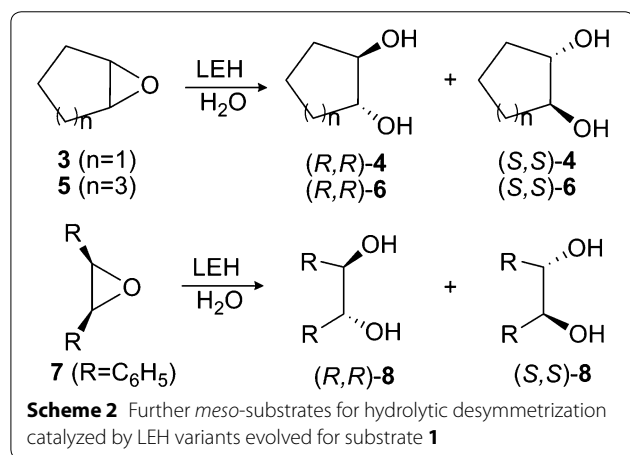
#### Exploring substrate scope of best variants

Some of the best hits obtained in the libraries described above were also employed as catalysts in the reaction of other substrates, compounds 3, 5 and 7 being tested first (Scheme 2). The results are summarized in Table 5. The most striking feature concerns the reaction of cyclopentene oxide (3), which reacts sluggishly, as in our previous report in which case an X-ray structure of the respective LEH mutant harboring epoxide 3 was obtained showing the substrate to be in a pose not suited for smooth ring-opening attack by activated water (Sun et al. 2015). However, most of the variants obtained in the present work still maintain activity with conversions of 3 amounting to >30 % for both 5 and 10 mM substrate concentrations. In the case of the cyclic seven-membered epoxide 5, excellent (*S,S*)-selectivity (>94 to 97 % ee) at good conversion is possible. In contrast, the best (*R,R*)-selective variant leads to only 53 % ee at good conversion. In the case

**Table 4 Best LEH variants of library CD and DC for catalytic desymmetrization of epoxide 1**

Library	Code	Mutations	Favored enantiomer	ee %	c %
CD	SZ503	I80 V/V83I/L114 V (template)	( <i>R,R</i> )	80	97
	TOQ-PA-218-08	L35G/I80 V/V83I/L103I/L114 V	( <i>R,R</i> )	80	90
	TOQ-PA-218-24	L74 V/I80 V/V83I/L114 V	( <i>R,R</i> )	67	91
	TOQ-PA-218-32	L74I/I80 V/V83I/L114 V	( <i>R,R</i> )	79	97
DC1	TOQ-PA-197-06	L74F/L103I/F134L (template)	( <i>R,R</i> )	55	35
	TOQ-PB-280-07	L74F/V83I/L103I/L114A/I116 V/F134L	( <i>R,R</i> )	40	55
	TOQ-PB-280-09	L74F/I80 V/L103I/F134L	( <i>R,R</i> )	61	41
DC2	TOQ-PA-197-05	L35 V/L74F (template)	( <i>R,R</i> )	38	67
	SZ274	L35 V/L74F/I80 V/V83I/L114 V	( <i>R,R</i> )	38	63
	SZ277	L35 V/L74F/I80G/V83I/L114 V	( <i>R,R</i> )	35	46
	SZ278	L35 V/L74F/I80F/V83I/I116 V	( <i>S,S</i> )	83	52
	SZ280	L35 V/L74F/I80F/I116 V	( <i>S,S</i> )	80	70
	SZ283	L35 V/L74F/L114 V/I116 V	( <i>S,S</i> )	87	80
	SZ284	L35 V/L74F/I80F/I116 V	( <i>S,S</i> )	81	65
	SZ285	L35 V/L74F/L114 V/I116 V	( <i>S,S</i> )	86	83
	SZ288 <sup>a</sup>	L35 V/L74F/V83T/L114 V/I116 V	( <i>S,S</i> )	90	66
	SZ290	L35 V/L74F/V83I/L114 V/I116 V	( <i>S,S</i> )	84	67

<sup>a</sup> The mutation of V83T of SZ288 was introduced randomly in the PCR procedure, which was not designed in library DC2



of sterically bulky epoxide **7**, notable improvements were not achieved, but the best (*R,R*)-selective mutant still shows higher ee than WT LEH (Table 5).

#### Docking analysis for explaining the source of evolved stereoselectivity

The general features of the mechanism of LEH-catalyzed reactions of epoxides are well known (Arand et al. 2003). The substrate needs to be positioned in a way that the activated water, properly positioned by Asp132, Asn55 and Tyr53, can induce a smooth  $S_N2$  reaction, activation by Asp101 by way of an H-bond to the epoxide O-atom also being necessary. In order to gain insight into the possible origin of the evolved stereoselectivity, substrate **1** was

docked into homology models of the (*S,S*)-selective mutant SZ502 (98 % ee) and the (*R,R*)-selective mutant SZ503 (80 % ee) (Table 5). The highest ranked docking poses for **1** in the SZ502 and SZ503 mutants that fulfill the H-bonding criteria to D101 described above are displayed in Additional file 1: Figure S1, Table S2 displaying the interatomic distances. A short discussion of the docking results, which are consistent with the observed (*R,R*)- and (*S,S*)-selectivities, is featured in the Supporting Information.

#### Conclusions

In previous LEH studies describing the hydrolytic desymmetrization of cyclohexene oxide (**1**) with formation of (*R,R*)- and (*S,S*)-**2** (Zheng and Reetz 2010; Sun et al. 2015, 2016b), saturation mutagenesis was focused on sites lining the binding pocket by employing several different approaches, including single code saturation mutagenesis (SCSM) (Sun et al. 2015), double code saturation mutagenesis (DCSM) (Sun et al. 2016c) and triple code saturation mutagenesis (TCSM) (Sun et al. 2016b) as extensions of the CASTing tool box. New mutants were discovered independently, displaying excellent enantioselectivity. Although the hypothesis regarding the choice of hydrophobic amino acids as building blocks proved to be successful in most cases, we were curious to know whether the appropriate combination of hydrophobic and hydrophilic amino acids in a binary way would also constitute a viable strategy in directed evolution, systematically exploring the LEH sequence space for active and stereoselective mutants.

**Table 5 Results of testing the best LEH mutants evolved for epoxide 1 as catalysts in the hydrolytic desymmetrization of further meso-substrates**

Code	1		3		5		7		Substrate concentration (mM)
	ee %	c %	ee %	c %	ee %	c %	ee %	c %	
WT	3 (S,S)	87	7 (R,R)	69	22 (S,S)	99	93 (R,R)	98	5
SZ501	97 (S,S)	96	83 (S,S)	30	89 (S,S)	97	19 (R,R)	93	
SZ502	97 (S,S)	93	94 (S,S)	34	97 (S,S)	96	19 (R,R)	93	
SZ503	79 (R,R)	97	51 (R,R)	48	53 (R,R)	99	94 (R,R)	99	
SZ504	96 (S,S)	93	87 (S,S)	31	94 (S,S)	97	24 (R,R)	95	
WT	2 (S,S)	99	13 (R,R)	64	18 (S,S)	99	92 (R,R)	99	10
SZ501	97 (S,S)	98	94 (S,S)	18	97 (S,S)	91	21 (R,R)	62	
SZ502	98 (S,S)	95	94 (S,S)	39	97 (S,S)	96	22 (R,R)	85	
SZ503	80 (R,R)	97	51 (R,R)	64	54 (R,R)	99	95 (R,R)	99	
SZ504	95 (S,S)	98	90 (S,S)	26	94 (S,S)	91	27 (R,R)	55	

In this study, we extended methodology development by applying and comparing different binary degenerate codon usages, including selected acidic, basic as well as aliphatic amino acids. Crossover experiments were also performed by ISM to explore the sequence space more extensively. Although the binary codon usage alone led to active variants showing enantioselectivities of up to 80 % ee, the overwhelming incorporation of hydrophobic amino acids proved to be essential for best results. Given that most of the residues lining the LEH binding pocket are hydrophobic (van der Werf et al. 1998; Arand et al. 2003), this may not be surprising. Indeed, it supports our previous conjectures and choices of reduced amino acids in the directed evolution of LEH (Sun et al. 2015, 2016b). However, choosing only hydrophobic amino acids as components of reduced amino acid alphabets cannot be viewed as a general guideline. Therefore, we suggest that binary codes may lead to good results in enzymes in which both polar and non-polar residues surround the active site. This hypothesis needs to be tested for other enzymes in future work.

Interestingly, the sequences of the best variants reveal novel point mutations not previously seen in earlier saturation mutagenesis approaches (Zheng and Reetz 2010; Sun et al. 2015, 2016b), are necessary for the evolution of high enantioselectivity and activity. Some of the best variants are also excellent catalysts in the hydrolytic desymmetrization of several other substrates. Finally, the docking analysis sheds light on the source of enhanced and inverted stereoselectivity.

## Experimental section

### Materials

KOD Hot Start DNA Polymerase, dNTPs and MgSO<sub>4</sub> for PCR were obtained from Novagen. Lysozyme and

DNase I were purchased from AppliChem. Dpn I was from New England Biolabs. The oligonucleotides were synthesized by Invitrogen and were used in the standard, desalted form without further purification. PCR purification kit and miniprep kit were from Qiagen or from Zymo Research. All the chemicals were purchased from Sigma-Aldrich, Tokyo Chemical Industry (TCI) or Alfa Aesar.

### 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 MgSO<sub>4</sub>, 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 s, 56 °C 30 s, 68 °C 40 s) × 32 cycles, 68 °C 120 s, 16 °C 30 min. For mega-PCR: 95 °C 3 min, (95 °C 30 s, 60 °C 30 s, 68 °C 5 min 30 s) × 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™ Buffer and 2 µL Dpn I were added in 50 µL PCR mixture and the digestion was carried out at 37 °C for more than 3 hours. After Dpn I digestion, the PCR products (1 µL) were directly transformed into electro-competent *E. coli* BL21(DE3) to create the final library for quick quality control and screening.

### Primer design and library creation

Library A and C were constructed as the following procedures: (1) Amplification of the short fragments of LEH using (mixed) primers F1/R1 for library A and F4/R4 for library B, respectively; (2) using the products of step 1 as

megaprimers amplify the whole plasmid of LEH leading to the final variety plasmids for library generation. The PCR products were digested by *Dpn* I and transformed into electro-competent *E. coli* BL21(DE3) to create the library for screening.

Library C and D were created in two cycles as the procedures described for library A and C: (1) The short fragments of LEH was amplified by primers F2/R2 and F5/R5, respectively; (2) Amplification of the whole plasmid of WT LEH using the products of step 1 as megaprimers to generate the diversity plasmid libraries; (3) The PCR products of step 2 were digested by *Dpn* I and transformed into electro-competent *E. coli* BL21(DE3) to create the diversity clones, then pooling these clones and extract the plasmid as further template; (4) using primers F3/R3 and F6/R6 to amplify the short fragments based on the plasmids template obtained in step 3, respectively; (5) Amplification of the whole plasmid again using the products of step 4 as megaprimers to generate the final diversity plasmid libraries; (6) The PCR products of last step were digested by *Dpn* I and transformed into electro-competent *E. coli* BL21(DE3) to create the library for screening. The primers were summarized in Additional file 1: Table S3.

#### Expression of the libraries

The colonies were picked to inoculate LB (supplied with carbenicillin [50 µg/mL] 300 µL) on 96 deep well plates and the precultures were incubated at 37 °C, 220 rpm, overnight. These cultures (120 µL) were used to prepare glycerol stocks, and to the rest of the precultures, TB [with carbenicillin (50 µg/mL) and lactose (0.5 % w/v), 600 µL] was added. The main cultures were incubated at 28 °C, 220 rpm, 7 hours. The cells were harvested (4000 rpm, 10 min), washed once with potassium phosphate buffer (100 mM, pH 7.4, 400 µL) and lysed using lysis buffer (the same buffer 400 µL with 1 mg lysozyme and 6 U DNase I) at 30 °C, 800 rpm, 1 hour. The lysates were clarified by centrifugation (4000 rpm, 30 minutes) and the clarified lysates were used as enzyme preparation in activity assay.

#### Activity assay and determination of the enantiomeric ratio

Activity of the variants was first assayed on microtiter plates using an adapted adrenaline assay, as previously described, using clarified lysate as the enzyme preparation. Test buffer for 8 plates consisted of acetonitrile (4.4 mL), substrate **1** (121 µL, final concentration in the reactions 10 mM) and potassium phosphate buffer (100 mM, pH 7.4, 83.6 mL). The test buffer (110 µL/well) was applied onto the micro titer plates and the enzyme lysates (40 µL) were added. The plates were incubated at 30 °C, 500 rpm, 3 hours. Afterwards, NaIO<sub>4</sub> solution

(20 µL, 77 mg in 24 mL of water) was added and the plates were further incubated at 30 °C, 500 rpm, 10 minutes. Subsequently, adrenaline solution [20 µL, (epinephrine 132 mg, water 24 mL, conc HCl 5 drops for solubilizing the adrenaline)] was added, which caused the immediate formation of red color in inactive reactions. Active variants gave colourless wells. Based on the activity test the selected lysates (300 µL) were transformed into a new 96 deep well plate with reaction buffer [100 µL, potassium phosphate buffer, 100 mM, pH 7.4 containing acetonitrile (20 % v/v) and substrate **1** (40 mM)]. In this way the final substrate concentration in the reactions was 10 mM and acetonitrile 5 % v/v. The reactions were incubated at 30 °C, 800 rpm, 16 hours. The reactions were extracted with EtOAc (3 × 150 µL) and the organic phases were analysed by chiral GC (see Additional file 1: Table S4).

#### Homology modelling

Homology models of the SZ502 and SZ503 mutants of LEH were prepared by using the Structure Prediction Wizard in Prime (2015), and the Knowledge-based approach. The previously determined crystal structures of the *R,R*- and *S,S*-selective SZ338 (L74F/M78 V/I180 V/L114F) and SZ92 (L74F/M78F/L103 V/L114 V/I116 V/F139 V/L147 V) mutants as templates for SZ503 and SZ502, respectively (Sun et al. 2015). Structural refinement was performed on the homology models using the Protein Preparation Wizard (Schrödinger Suite 2015). Hydrogen atoms were added according to the protonation states determined by PROPKA (Olsson et al. 2011). Activation of the epoxide ring oxygen by protonated D101 is a prerequisite for ring opening in LEH; therefore D101 was manually assigned the protonated (neutral) form. The positions of all hydrogen atoms were energy minimized using the Impact program (2014) and the OPLS2005 force field. The coordinates of the nucleophilic water molecule were copied from the WT LEH crystal structure (PDB 1NU3) following structural alignment with the mutant homology model.

#### Docking of cyclohexene oxide

The structure of cyclohexene oxide (**1**) was built in the Maestro program (2014) and prepared for docking using LigPrep (2014). Docking was performed to the homology model of the SZ502 and SZ503 homology models using Glide (2014) with the standard precision (SP) settings. A maximum of 20 docking poses for each model was requested and 19 poses were found for each of the two models. Previous QM calculations have revealed that the substrate should be positioned such that the D101 proton points towards the front face of the epoxide ring (Hopmann et al. 2005). Hence, only the docking poses that contained this interaction were considered during analysis.

## Additional file

**Additional file 1.** Additional figures and tables including: **Table S1.** Best LEH variants of library CB and AD for desymmetrization of epoxide 1. **Table S2.** Distances [in Å] and angles [degrees] calculated for the highest ranked docking poses calculated for cyclohexene oxide in the WT, SZ502 and SZ503 homology models of LEH. **Table S3.** List of primers used in this study. **Table S4.** Analytic conditions of GC. **Figure S1.** Highest ranked docking pose for cyclohexene oxide (1) in the (a) SZ502 and (b) SZ503 mutant homology models of LEH. **Figure S2.** Docking of substrate 1 to WT LEH.

### Authors' contributions

ZS, PTS, ES and RL contributed equally to this work. All authors read and approved the final manuscript.

### Author details

<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.

### Acknowledgements

Support from the Max-Planck-Society and the LOEWE Research cluster SynChemBio is gratefully acknowledged.

### Competing interests

The authors declare that they have no competing interests.

Received: 10 August 2016 Accepted: 10 September 2016

Published online: 15 September 2016

### References

- Acevedo-Rocha CG, Hoebenreich S, Reetz MT (2014) Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. In: Gillam EMJ, Copp JN, Ackerley DF (eds) Directed evolution library creation, methods and protocols, vol 1179, 2nd edn. Humana Press, Totowa, pp 103–128
- Akanuma S, Kigawa T, Yokoyama S (2002) Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc Natl Acad Sci USA* 99(21):13549–13553
- Arand M, Hallberg BM, Zou J, Bergfors T, Oesch F, van der Werf MJ, de Bont JAM, Jones TA, Mowbray SL (2003) Structure of Rhodococcus erythropolis limonene-1,2-epoxide hydrolase reveals a novel active site. *EMBO J* 22(11):2583–2592
- Bommarius AS (2015) Biocatalysis: a status report. *Annu Rev Chem Biomol Eng* 6(1):319–345
- Denard CA, Ren H, Zhao H (2015) Improving and repurposing biocatalysts via directed evolution. *Curr Opin Chem Biol* 25:55–64
- Estell DA, Graycar TP, Wells JA (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J Biol Chem* 260(11):6518–6521
- Fellouse FA, Li B, Compaan DM, Peden AA, Hymowitz SG, Sidhu SS (2005) Molecular Recognition by a Binary Code. *J Mol Biol* 348(5):1153–1162
- Glide (2014) Version 6.5. Schrödinger, LLC, New York
- Hopmann KH, Hallberg BM, Himo F (2005) Catalytic mechanism of limonene epoxide hydrolase, a theoretical study. *J Am Chem Soc* 127(41):14339–14347
- Impact (2014) Version 6.5. Schrödinger, LLC, New York
- Kamtekar S, Schiffer JM, Xiong HY, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino-acids. *Science* 262(5140):1680–1685
- Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S (2007) High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci USA* 104(16):6632–6637
- Li T, Fan K, Wang J, Wang W (2003) Reduction of protein sequence complexity by residue grouping. *Protein Eng Des Sel* 16(5):323–330
- LigPrep (2014) Version 3.2. Schrödinger, LLC, New York
- Lutz S, Bornscheuer UT (2009) Protein Engineering Handbook. Wiley-VCH, Weinheim
- Maestro (2014) V10.0. Schrödinger, LLC, New York
- Nov Y (2012) When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl Environ Microbiol* 78(1):258–262
- Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537
- Patrick WM, Firth AE (2005) Strategies and computational tools for improving randomized protein libraries. *Biomol Eng* 22(4):105–112
- Prime (2015) Version 3.9. Schrödinger, LLC, New York
- Reetz MT (2011) Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew Chem Int Ed* 50(1):138–174
- Reetz MT, Carballeira JD (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protocols* 2(4):891–903
- Reetz MT, Wilensek S, Zha D, Jaeger KE (2001) Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angew Chem Int Ed* 40:3589–3591
- Reetz MT, Bocola M, Carballeira JD, Zha D, Vogel A (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew Chem Int Ed* 44(27):4192–4196
- Reetz MT, Wang LW, Bocola M (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed* 45(8):1236–1241 (**Erratum, 2494**)
- Roy S, Hecht MH (2000) Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry* 39:4603–4607
- Schrödinger Suite 2015-1 (2015) Protein Preparation Wizard; Epik version 3.1, Schrödinger, LLC, New York, NY, 2015; Impact version 6.6, Schrödinger, LLC, New York, NY, 2015; Prime version 3.9, Schrödinger, LLC, New York
- Sun Z, Lonsdale R, Kong XD, Xu JH, Zhou J, Reetz MT (2015) Reshaping an enzyme binding pocket for enhanced and inverted stereoselectivity: use of smallest amino acid alphabets in directed evolution. *Angew Chem Int Ed* 54(42):12410–12415
- Sun Z, Wikmark Y, Bäckvall J-E, Reetz MT (2016a) New concepts for increasing the efficiency in directed evolution of stereoselective enzymes. *Chem Eur J* 22:5046–5054
- Sun Z, Lonsdale R, Wu L, Li G, Li A, Wang J, Zhou J, Reetz MT (2016b) Structure-guided triple-code saturation mutagenesis: efficient tuning of the stereoselectivity of an epoxide hydrolase. *ACS Catalysis* 6:1590–1597
- Sun Z, Lonsdale R, Li G, Reetz MT (2016c) Comparing different strategies in directed evolution of enzyme stereoselectivity: single- versus double-code saturation mutagenesis. *ChemBioChem*. doi:10.1002/cbic.201600296
- 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(19):5052–5057
- Wahler D, Reymond JL (2002) The adrenaline test for enzymes. *Angew Chem Int Ed* 41(7):1229–1232
- Walter KU, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* 280(45):37742–37746
- Williams E, Copp JN, Ackerley DF (2014) Site-saturation mutagenesis by overlap extension PCR. In: Gillam EMJ, Copp JN, Ackerley DF (eds) Directed evolution library creation, methods and protocols, vol 1179, 2nd edn. Humana Press, Totowa, pp 83–101
- 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(44):15744–15751