A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEMBIOCHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: Comparing Different Strategies in Directed Evolution of Enzyme Stereoselectivity: Single versus Double Code Saturation Mutagenesis

Authors: Zhoutong Sun; Richard Lonsdale; Guangyue Li; Manfred T. Reetz

To be cited as: ChemBioChem 10.1002/cbic.201600296

Link to VoR: http://dx.doi.org/10.1002/cbic.201600296

A Journal of

ChemPubSoc Europe

www.chembiochem.org

WILEY-VCH

FULL PAPERS

Comparing Different Strategies in Directed Evolution of Enzyme Stereoselectivity: Single versus Double Code Saturation Mutagenesis

Zhoutong Sun[a, b], Richard Lonsdale[a, b], Guangyue Li[a, b] and Manfred T. Reetz*[a, b]

Abstract: Saturation mutagenesis at sites lining the binding pocket of enzymes constitutes a viable protein engineering technique for enhancing or inverting stereoselectivity. Statistical analysis shows that oversampling in the screening step (the bottleneck) increases astronomically as the number of residues in the randomization site increases, which is the reason why reduced amino acid alphabets have been employed in addition to splitting large sites into smaller ones. Limonene epoxide hydrolase (LEH) has previously served as the experimental platform in these methodology efforts, enabling comparisons between single code saturation mutagenesis (SCSM) and triple code saturation mutagenesis (TCSM) which employ only one versus three amino acids as building blocks, respectively. In the present study the comparative platform is extended by exploring the efficacy of double code saturation mutagenesis (DCSM) in which the reduced amino acid alphabet comprises two members, chosen rationally on the basis of structural information. The hydrolytic desymmetrization of cyclohexene oxide is used as the model reaction with formation of either (*R*,*R*)- or (*S*,*S*)-cyclohexane-1,2-diol. DCSM proves to be clearly superior to the likewise tested SCSM, affording both (*R*,*R*)- and (*S*,*S*)-selective mutants. The respective variants are also good catalysts in the reaction of other further substrates. Docking computations reveal the origin of enantioselectivity.

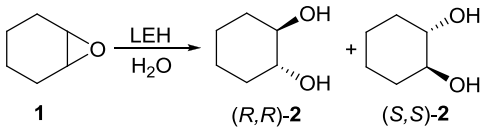
Introduction

Methodology development in directed evolution of stereoselective enzymes is an important current task, aiming to enhance the speed, efficacy and reliability of this widely practiced protein engineering technique.[1] Since screening for stereoselectivity, activity and/or substrate scope is the labor-intensive bottleneck, the goal is the creation of small and maximally "smart" mutant libraries. In this endeavor, a stochastic process called saturation mutagenesis (SM)[2] at sites lining the binding pocket has emerged as the method of choice (dubbed CASTing) for many enzymes.[1a,1c,3] When employing NNK codon degeneracy encoding all 20 canonical amino acids as building blocks, protein sequence space and concurrent oversampling increases astronomically as the size of a randomization site increases from one residue to, e.g., 10. Therefore, two "tricks" have been proposed which offer a way out of this dilemma: 1) Splitting up a large site into smaller ones, followed by iterative saturation mutagenesis (ISM) if necessary;[1g,4] and/or 2) Use of reduced amino acid alphabets.[1g,4] Reduced amino acid alphabets have been used previously for other purposes, e.g., studying whether correct folds[5a] and enzyme activity[5b] are maintained when employing less than the 20 canonical amino acids in protein construction, or when applying a binary code in the production of functional antibodies[6a] or monobodies.[6b] In the applications that we focus on, consideration of oversampling is essential. The necessary degree of oversampling for ensuring a certain degree of library coverage that the operator stipulates, e.g., 95%, is easily calculated using the CASTER computer aid (www.kofo.mpg.de/en/research/biocatalysis),[7] which is based on the Patrick/Firth algorithm.[8] Alternatively, the Nov-metric can be used to compute the n^{th} best mutant as a function of the number of screened transformants.[9] One and the same reduced amino acid alphabet as defined by the respective codon degeneracy can be applied when randomizing combinatorially the whole multi-residue site, or a different codon degeneracy can be chosen at each individual position of a multi-residue site, likewise in a single SM experiment.[10]

In the present study we compare the virtues and drawbacks of two different approaches to SM using appropriately chosen codon degeneracies for controlling stereoselectivity: Double code saturation mutagenesis (DCSM) based on two amino acids as building blocks versus single code saturation mutagenesis (SCSM) utilizing only one amino acid, in both cases for the same entire randomization site. In these and other reduced amino acid alphabets, wildtype (WT) amino acids are included.[10] In the model system we employ limonene epoxide hydrolase (LEH)[11] as the catalyst for the hydrolytic desymmetrization of cyclohexene oxide (**1**) with formation of cyclohexane-1,2-diols (*R*,*R*)- or (*S*,*S*)-**2** (Scheme 1). WT LEH shows minimal enantioselectivity in slight favor of (*S*,*S*)-**2** (4% ee). This experimental platform has already been utilized in several previous directed evolution studies, enabling further comparisons.[12] In one approach, 10 CAST residues were grouped into a single large site which was randomized using SCSM with valine, phenylalanine or tyrosine being tested as the sole building block, respectively.[12b] Valine proved to be the superior building block in this particular enzyme. When applying triple code saturation mutagenesis (TCSM) employing three amino acids as building blocks in randomization, even better results were obtained.[12c]

[a] *Dr. Z. Sun, Dr. R. Lonsdale, Dr. G. Li, Prof. Dr. M. T. Reetz*
Department of Synthetic Organic Chemistry
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr (Germany)
Fax: (+49) 208-306-2985
E-mail: reetz@mpi-muelheim.mpg.de
[b] *Dr. Z. Sun, Dr. R. Lonsdale, Dr. G. Li, Prof. Dr. M. T. Reetz*
Fachbereich Chemie
Philipps-Universität Marburg
Hans-Meerwein-Strasse, 35032 Marburg (Germany)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.



Scheme 1. Hydrolytic desymmetrization catalyzed by mutants of limonene epoxide hydrolase (LEH) generated by different saturation mutagenesis strategies.

Results and Discussion

Generation and screening of mutant libraries. As in previous directed evolution studies of LEH,[12b-c] the crystal structure of wildtype (WT) LEH[11b] served as a guide for choosing ten CAST residues. This time they were grouped into two 5-residue randomization sites, A5 (L74/F75/M78/I80/L103) and B5 (L114/I116/F134/F139/L147) (Fig. 1), thereby setting the platform for systematic comparison of SCSM and DCSM.

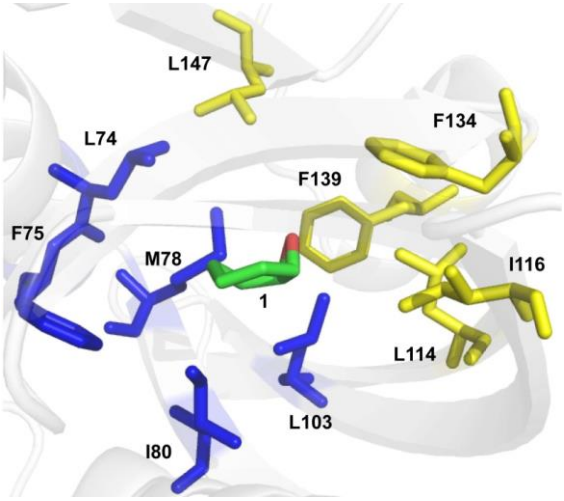


Figure 1. LEH binding pocket featuring substrate **1** surrounded by 10 CAST residues which are grouped into two 5-residue randomization sites A5 (blue) and B5 (yellow).

The crystal structure of WT LEH[11b] shows that almost all of the residues surrounding the binding pocket have hydrophobic character. Therefore, when designing a single or triple code, it was logical to choose amino acids with hydrophobic sidechains. In the case of the previous SCSM study[12b], valine, phenylalanine or tyrosine were chosen, valine leading to the best results in terms of enantioselectivity and activity, which was improved further by ISM using one of the other two amino acids as building blocks. In the TCSM study,[12c] V-F-Y was chosen as the triple code, which provided notably better results.

Based on these data, we chose valine for SCSM at randomization sites A5 and B5, respectively. In each case, 92 transformants corresponding to ~95% library coverage were assayed for activity, the hits (55) then being screened for enantioselectivity using chiral GC. The results are summarized in Table 1. The best (*R*,*R*)-selective mutant SZ539 (L74V/M78V/I80V) originated from library A5, showing 77% ee. It was then used as a template for SCSM at site B5, but this failed to provide any improved variants.

In case of site B5, mutant SZ544 (L114V/I116V/F139V) was discovered showing 75% ee in favor of (*S*,*S*)-**2**. This mutant formed the basis of SCSM at site A5, which provided variant SZ634 (L103V/L114V/I116V/F139V) with enhanced (*S*,*S*)-selectivity (89% ee). Further ISM was carried out using SZ639 (L74V/M78V/I80V) and SZ634 (L103V/L114V/I116V/F139V) as templates to visit the rest of the residues of L103/L114/I116/L147 and L74/M78/I80/L147 with phenylalanine as building block, respectively. The improved hits up to 92% ee were found in library AB5-V-F-2. The best results are summarized in Table 1.

Table 1. Results of applying single code saturation mutagenesis (SCSM) to LEH-catalyzed hydrolytic desymmetrization of substrate **1**.

Library (Code)	Template	Code	Mutations	ee%	%-conv.	Favored enantiomer	A/T*
A5-V (Val)	WT	WT		4	99	(*S*,*S*)	92/96
		SZ537	L74V/L103V	7	97	(*S*,*S*)	
		SZ538	L74V/I80V	61	98	(*R*,*R*)	
		SZ539	L74V/M78V/I80V	77	97	(*R*,*R*)	
		SZ541	I80V	75	97	(*R*,*R*)	
B5-V (Val)	WT	SZ543	L114V/I116V/L147V	54	95	(*S*,*S*)	92/96
		SZ544	L114V/I116V/F139V	75	98	(*S*,*S*)	
		SZ546	L114V/I116V	63	95	(*S*,*S*)	
A5-V-B5-V (Val)	SZ539	SZ639	L74V/M78V/I80V	77	97	(*R*,*R*)	92/96
B5-V-A5-V (Val)	SZ544	SZ633	F75V/L103V/L114V/I116V/F139V	88	97	(*S*,*S*)	92/96
		SZ634	L103V/L114V/I116V/F139V	89	99	(*S*,*S*)	
		SZ635	L114V/I116V/F139V	88	99	(*S*,*S*)	
AB5-V-F-1 (Phe)	SZ639		no improvement				46/48
AB5-V-F-2 (Phe)	SZ634	SZ662	L74F/M78F/L103V/L114V/I116V/F139V/L147F	90	89	(*S*,*S*)	46/48
		SZ663	L74F/M78F/L103V/L114V/I116V/F139V	90	95	(*S*,*S*)	
		SZ664	L74F/I80F/L103V/L114V/I116V/F139V	92	94	(*S*,*S*)	
		SZ665	L74F/M78F/I80F/L103V/L114V/I116V/F139V	92	96	(*S*,*S*)	

*A/T: Colonies picked/ computed colonies number for 95% coverage

Before presenting the data obtained by application of DCSM, the present SCSM strategy can be compared to the original SCSM approach in which the same 10 residues formed a single randomization site with valine also serving as the single code followed by phenylalanine-based ISM.[12b] The final results of the two approaches are comparable, but the screening effort differs considerably. The advantage of grouping 10 residues into two 5-residue compared with the previous study in which all residues formed a single randomization site using valine as the single code is the fact that the number of screened transformants has been reduced by a factor of eight (from 3220 to 368).

The main goal of the present study was to compare SCSM with DCSM within the same experimental framework as defined by the two 5-residue randomization sites. In this case valine and phenylalanine were chosen as the double code (V-F) for randomizing sites A5 and B5. In the A5-VF library using WT as template, mainly (*R*,*R*)-selective variants were found, the two best hits being M78F/I80V (80% ee) and M78V/I80V (75% ee) (Table 2). Enantioselectivity was then boosted by performing ISM. These two mutants were then used as template to visit the B5 site with generation of two libraries A5-B5-VF-1 and A5-B5-VF-2, respectively. The best hits SZ643 (M78F/I80V/L114F/L147V) and SZ655 (M78V/I80V/L114F) reached ≈92% ee. On the other hand, the initial B5-VF library contained the best (*S*,*S*)-selective variant SZ597 (L114V/I116V/F139V/L147F) (85% ee). ISM based on employing SZ597 as template and focusing randomization on site A5 provided several improved mutants showing enantioselectivity in the range 92-95% ee. The results are summarized in Table 2.

Table 2. Results of applying double code saturation mutagenesis (DCSM) based on valine-phenylalanine as building blocks to LEH-catalyzed hydrolytic desymmetrization of substrate **1**.

Library	Template	Code	Mutations	ee%	%-conv.	Favored enantiomer	A/T*
A5-VF	WT	WT		4	99	(*S*,*S*)	552/486
		SZ548	L74F	13	80	(*S*,*S*)	
		SZ549	L74V/M78V/I80V	59	99	(*R*,*R*)	
		SZ550	L74F/M78F/I80F	53	99	(*S*,*S*)	
		SZ551	F75V/M78F	54	92	(*R*,*R*)	
		SZ553	I80V	72	99	(*R*,*R*)	
		SZ574	M78F/I80V	80	99	(*R*,*R*)	
		SZ575	M78F/I80V/L103V	68	99	(*R*,*R*)	
		SZ576	F75V/M78F/I80V	71	98	(*R*,*R*)	
		SZ577	L74F/M78V	58	100	(*S*,*S*)	
		SZ580	M78V/I80V	75	99	(*R*,*R*)	
		SZ583	L74V/M78F/I80V	80	98	(*R*,*R*)	
B5-VF	WT	SZ558	L114F	21	71	(*R*,*R*)	368/324
		SZ562	L114V/I116V/F139V	83	97	(*S*,*S*)	
		SZ566	L114F/L147V	36	22	(*R*,*R*)	
		SZ569	L114V/I116V	72	86	(*S*,*S*)	
		SZ570	L114V/I116V/L147F	72	82	(*S*,*S*)	
		SZ597	L114V/I116V/F139V/L147F	85	88	(*S*,*S*)	
		SZ603	L114F/L147F	61	43	(*R*,*R*)	
		SZ605	L114V/I116V/F139V/L147V	82	99	(*S*,*S*)	
A5-B5-VF-1	SZ574	SZ640	M78F/I80V/L114F	90	98	(*R*,*R*)	368/324
		SZ641	M78F/I80V	84	97	(*R*,*R*)	
		SZ642	M78F/I80V/L114F/L147F	90	97	(*R*,*R*)	
		SZ643	M78F/I80V/L114F/L147V	91	96	(*R*,*R*)	
A5-B5-VF-2	SZ580	SZ655	M78V/I80V/L114F	92	99	(*R*,*R*)	368/324
B5-A5-VF-3	SZ597	SZ646	L74F/M78F/I80F/L114V/I116V/F139V/L147F	94	96	(*S*,*S*)	552/486
		SZ648	M78V/I80F/L114V/I116V/F139V/L147F	92	98	(*S*,*S*)	
		SZ649	I80F/L114V/I116V/F139V/L147F	93	92	(*S*,*S*)	
		SZ650	L74F/I80F/L114V/I116V/F139V/L147F	95	93	(*S*,*S*)	
		SZ651	M78F/I80F/L114V/I116V/F139V/L147F	93	94	(*S*,*S*)	
		SZ653	L74F/M78F/L114V/I116V/F139V/L147F	92	82	(*S*,*S*)	

*A/T: Colonies picked/ computed colonies number for 95% coverage

Even though the particular version of SCSM in this study involved the screening of only 460 transformants and provided excellent (*S*,*S*)-variants, it failed to generate optimal (*R*,*R*)-selective mutants. When comparing the present DCSM strategy with the original SCSM approach described in the previous study featuring best mutants (*S*,*S*)-selective (92%ee) and (*R*,*R*)-selective mutants (up to 96% ee),[12b] it becomes clear that similar enantioselectivities occur in both cases. However, the total screening effort in the original SCSM approach (4508 transformants) is about twice as high as in the present DCSM strategy (2208 transformants).

Finally, we were interested in the outcome of applying DCSM to the entire 10-residue randomization site in the absence of splitting it into smaller sites (Fig. 1). Here again the double code V-F was used, which means that ≈$5x10^4$ transformants would have to be screened for 95% coverage of the respective library C10-VF. However, in this case no effort was made to reach high library coverage. Instead, only 1840 transformants were screened

(3.5% library coverage). This is equal to the screening in the case of DCSM with split libraries, thereby allowing for comparison. The results show that several excellent (*S*,*S*)-selective variants were evolved, the best one being SZ724 (L74F/M78F/I80F/L114V/I116V/F139V) with 97% ee. Reversal of enantioselectivity in favor of (*R*,*R*)-**2** with 71% ee was also achieved, but truly high levels were not detected (Table 3). When aiming for both (*S*,*S*)- and (*R*,*R*)-selectivity, then splitting into two smaller randomization sites each comprising five residues is the superior strategy. Moreover, this approach allows for ISM using the same building block whenever the initial library fails to fulfill all requirements that the operator has defined. The results can also be compared to our original study in which an even smaller reduced amino acid alphabet (valine) was used to randomize the same 10-residue site in LEH.[12b] In that case full library coverage was easily achieved, leading to the discovery of both (*S*,*S*)- and (*R*,*R*)-selective variants (86% ee and 76% ee, respectively). Further optimization up to 92% ee (*S*,*S*) and 96% ee (*R*,*R*) was possible by ISM using a different amino acid as the single code (phenylalanine).

Table 3. Results of applying double code saturation mutagenesis (DCSM) based on valine-phenylalanine for ten residues in one library to LEH-catalyzed hydrolytic desymmetrization of substrate **1**.

Library	Code	Mutations	ee%	c%	Favored enantiomer	A/T*
C10-VF	WT		4	99	(*S*,*S*)	1840/52488 (~3.5% coverage)
	SZ669	L74F/M78F/I80V/L114V/I116V/F139V/L147V	88	99	(*S*,*S*)	
	SZ670	L74F/L103V/L114V/I116V/F139V	91	99	(*S*,*S*)	
	SZ672	L74V/M78F/I80V/L103V/I116V	16	94	(*R*,*R*)	
	SZ674	L74F/I80F/L103V/L114V/I116V/F139V/L147F	86	81	(*S*,*S*)	
	SZ675	L74F/M78V/I80V/L103V/L114V/L147V	19	94	(*R*,*R*)	
	SZ676	L74F/L103F/L114V/F139V	42	56	(*R*,*R*)	
	SZ677	L74F/L114F/I116V/L147F	49	65	(*R*,*R*)	
	SZ678	L74F/I80F/L114V/I116V/F139V	95	97	(*S*,*S*)	
	SZ679	L74V/M78F/I80V/L103V/I116V/F139V	38	97	(*R*,*R*)	
	SZ681	L74F/M78V/L103V/L114V/I116V/F139V	86	96	(*S*,*S*)	
	SZ682	L74F/L103V/L114F/L147F	40	92	(*R*,*R*)	
	SZ724	L74F/M78F/I80F/L114V/I116V/F139V	97	98	(*S*,*S*)	
	SZ806	L74F/M78F/I80F/L114V/I116V/L147F	91	97	(*S*,*S*)	
	SZ807	L74V/M78F/I80V/F139V	71	98	(*R*,*R*)	
	SZ808	L114V/I116V/F139V/L147V	82	97	(*S*,*S*)	
	SZ810	L74F/I80V	43	98	(*R*,*R*)	
	SZ811	L74F/M78F/I80V/L114V/I116V	83	98	(*S*,*S*)	

*A/T: Colonies picked/ computed colonies number for 95% coverage

The different strategies based on the new data originating from the present study are summarized in Scheme S1. The results show that DCSM is the superior strategy relative to SCSM. It is also more efficient than the previous SCSM approach.[12b] (see Table 4). As previously demonstrated,[12c] TCSM appears to be even better than DCSM based on the data shown here. However, when comparing TCSM directly with DCSM, it needs to be pointed out that different sets of residues were selected and different strategies for grouping single residues into larger randomization sites were chosen in each case. The kinetic parameters of the best mutants obtained from the DCSM strategy were also compared with the kinetic profiles of the best variants identified previously from application of SCSM and TCSM.[12b-c] Significant differences in activity were not found, except that variant SZ724 is four times as active as SZ348, both mutants favoring (*S*,*S*)-**2**. In general, activity of the evolved variants proved to be lower than that of WT LEH (Table 5).

Table 4. Comparative summary of different saturation mutagenesis strategies for the hydrolytic desymmetrization of substrate **1**.

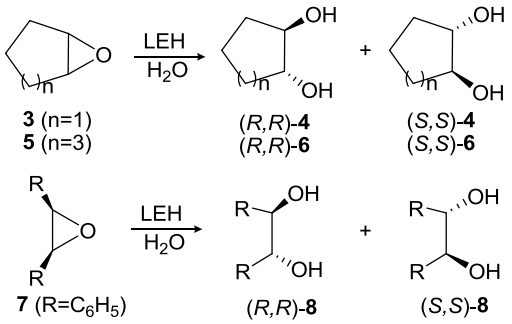
Strategy	Library (code usage)	Grouping and residues included	Total screening transformants (library coverage)	Best hits	Ref.
SCSM	Val	L74/F75/M78/I80/L103/ L114/I116/F134/F139/L147	3220 (>95%)	86% (*S*,*S*) 76% (*R*,*R*)	[12b]
SCSM	A5, B5, A5 → B5 B5 → A5 (Val)	A5: L74/F75/M78/I80/L103	368 (~95%)	89% (*S*,*S*) 77% (*R*,*R*)	This study
		B5: L114/I116/F134/F139/ L147			
DCSM	A5, B5, A5 → B5 B5 → A5 (Val/Phe)	A5: L74/F75/M78/I80/L103	1840 (>95%)	95% (*S*,*S*) 92% (*R*,*R*)	This study
		B5: L114/I116/F134/F139/ L147			
DCSM	Val/Phe	L74/F75/M78/I80/L103/ L114/I116/F134/F139/L147	1840 (~3.5%)	97% (*S*,*S*) 71% (*R*,*R*)	This study
TCSM	A, B, C A → B → C	A: I80/V83/L114/I116 B: L74/M78/L147	1344 (95%)	99% (*S*,*S*) 97% (*R*,*R*)	[12c]

	A → C → B (Val/Phe/Tyr)	C: M32/L35/L103			

Table 5. Kinetic parameters of best mutants from SCSM, DCSM and TCSM for hydrolytic desymmetrization of substrate **1**.

Strategy	Best mutants	Mutations	ee%	c%	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} M^{-1})	Ref.
	WT		4(*S*,*S*)	99	6.70±0.45	0.82±0.017	122.39	
SCSM	SZ92	L74F/M78F/L103V/L114V/I116V/F139V/L147V	92(*S*,*S*)	99	9.79±0.53	0.257±0.004	26.00	[12b]
	SZ338	L74F/M78V/I80V/L114F	96(*R*,*R*)	83	4.96±0.58	0.241±0.008	49.00	
DCSM	SZ655	M78V/I80V/L114F	92(*R*,*R*)	99	34.07±1.21	0.64±0.01	18.78	This study
	SZ724	L74F/M78F/I80F/L114V/I116V/F139V	97(*S*,*S*)	98	20.32±2.64	0.51±0.02	25.09	
TCSM	SZ348	I80Y/L114V/I116V	99(*S*,*S*)	97	19.11±1.93	0.11±0.005	5.75	[12c]
	SZ529	M32V/M78V/I80V/L114F	97(*R*,*R*)	99	10.12±0.81	0.18±0.005	17.78	

Exploring substrate scope of best variants. Some of the best hits obtained in the libraries described above were tested as catalysts in the reaction of other substrates, beginning with compounds **3**, **5** and **7** (Scheme 2). The results are summarized in Table 6. The most striking feature concerns the reaction of cyclopentene oxide (**3**), which reacts sluggishly if at all. This surprising phenomenon was also observed in an earlier study, in which case a crystal structure of the respective LEH mutant harboring **3** was obtained showing the substrate to be in a pose not amenable to smooth ring-opening attack by activated water.[12b] A similar explanation is likely to apply in the present case. In the case of the cyclic seven-membered epoxide **5**, excellent (*S*,*S*)-selectivity (>92-97% ee) at good conversion is possible, the best (*R*,*R*)-selective variant leading to 85% ee at moderate conversion.



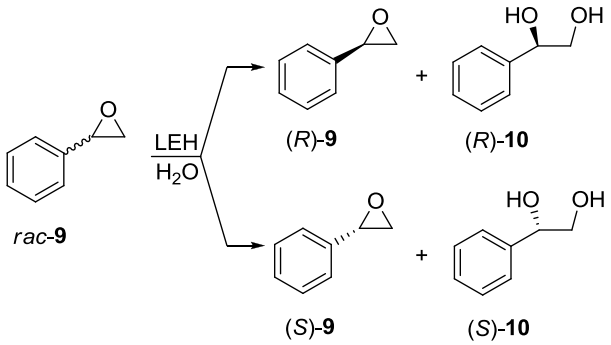
Scheme 2. Further *meso*-substrates for hydrolytic desymmetrization catalyzed by LEH variants evolved for substrate **1**.

Table 6. Results of testing the best LEH mutants evolved for epoxide **1** as catalysts in the hydrolytic desymmetrization of further *meso*-substrates.

	1		**3**		**5**		**7**	
Code	ee%	%-conv.	ee%	%-conv.	ee%	%-conv.	ee%	%-conv.
WT	4(*S*,*S*)	>99	13(*R*,*R*)	84	17(*S*,*S*)	97	93 (*R*,*R*)	>99
SZ640	90(*R*,*R*)	98	nd	<5	nd	<5	>99(*R*,*R*)	55
SZ642	90(*R*,*R*)	97	nd	<5	79(*R*,*R*)	27	>99(*R*,*R*)	41
SZ643	91(*R*,*R*)	96	nd	<5	75(*R*,*R*)	13	>99(*R*,*R*)	33
SZ646	96(*S*,*S*)	97	79(*S*,*S*)	6	95(*S*,*S*)	49	94(*R*,*R*)	13
SZ648	93(*S*,*S*)	94	81(*S*,*S*)	8	91(*S*,*S*)	72	71(*R*,*R*)	74
SZ649	94(*S*,*S*)	96	nd	<5	90(*S*,*S*)	39	76(*R*,*R*)	72
SZ650	95(*S*,*S*)	93	nd	<5	85(*S*,*S*)	25	>99(*R*,*R*)	15
SZ651	94(*S*,*S*)	94	81(*S*,*S*)	10	92(*S*,*S*)	83	74(*R*,*R*)	91
SZ653	95(*S*,*S*)	98	76(*S*,*S*)	10	93(*S*,*S*)	73	75(*R*,*R*)	19
SZ655	92(*R*,*R*)	99	nd	<5	85(*R*,*R*)	40	>99(*R*,*R*)	75
SZ663	91(*S*,*S*)	96	nd	<5	93(*S*,*S*)	74	44(*S*,*S*)	43
SZ664	93(*S*,*S*)	99	nd	<5	95(*S*,*S*)	60	8(*R*,*R*)	40
SZ665	93(*S*,*S*)	99	nd	<5	94(*S*,*S*)	56	6(*S*,*S*)	16
SZ670	90(*S*,*S*)	93	nd	<5	93(*S*,*S*)	85	23(*S*,*S*)	75
SZ678	96(*S*,*S*)	97	85(*S*,*S*)	8	96(*S*,*S*)	62	67(*R*,*R*)	36
SZ724	97(*S*,*S*)	98	85(*S*,*S*)	13	97(*S*,*S*)	73	90(*R*,*R*)	69

nd: not determined.

Finally, we studied the hydrolytic kinetic resolution of styrene oxide (*rac*-**9**) (Scheme 3). It can be seen that good (*S*)-selectivity can be achieved (*E* = 40.2), as opposed to (*R*)-selectivity which reaches a maximum of only *E* = 9.5 (Table 7).



Scheme 3. Hydrolytic kinetic resolution catalyzed by LEH variants evolved for substrate **1**.

Table 7. Results of testing the best LEH mutants evolved for epoxide **1** as catalysts in the hydrolytic kinetic resolution of styrene oxide (**9**).

	9		
Code	ee_p(%)	c%	*E*
WT	21 (*R*)	46	1.8
SZ640	nd	<5	-
SZ642	nd	<5	-
SZ643	nd	<5	-
SZ646	nd	<5	-
SZ648	92 (*S*)	36	40.2
SZ649	79 (*R*)	12	9.5
SZ650	nd	<5	-
SZ651	91 (*S*)	36	35.3
SZ653	nd	<5	-
SZ655	57 (*S*)	7	3.8
SZ663	nd	<5	-
SZ664	nd	<5	-
SZ665	nd	<5	-
SZ670	nd	<5	-
SZ678	82 (*S*)	15	11.6
SZ724	nd	<5	-

Docking analysis for explaining the source of evolved stereoselectivity. In order to shed some light on the possible origin of the observed evolved stereoselectivity, we analyzed the results of docking substrate **1** into homology models of the (*R*,*R*)-selective mutant SZ655 (92% ee) and the (*S*,*S*)-selective mutant SZ724 (97% ee) (Fig. 2). The mechanism of WT LEH is known to involve substrate binding and activation by hydrogen bonding between Asp101 and the epoxide O-atom, followed by S_N2 reaction initiated by the attack of a correctly positioned and activated water molecule.[11] In the case of the (*R*,*R*)-selective mutant, the binding mode of **1** is expected to position C1 closest to the activated water molecule, while the reshaped binding pocket of the (*S*,*S*)-selective mutant should position C2 closest to the activated water molecule. Indeed, the highest ranked docking pose for the SZ724 mutant places C2 closest to the nucleophilic water oxygen, C1 being further away by 0.5 Å. (Fig. 2a and Table S1). Out of the 15 docking poses found for this mutant, the pose shown in Figure 1a is the only one in which both the activating hydrogen bond to D101 is present, and the epoxide ring carbon atoms are close enough to the water molecule in order for nucleophilic attack to occur. For the SZ655 mutant, two poses were found in which both of these criteria were met, corresponding to the two highest ranked poses. Both poses, including the one displayed in Figure 2b, place the C1 atom closest to the water oxygen (Table S1), consistent with the experimentally observed preference for the formation of the (*R*,*R*)-**2**.

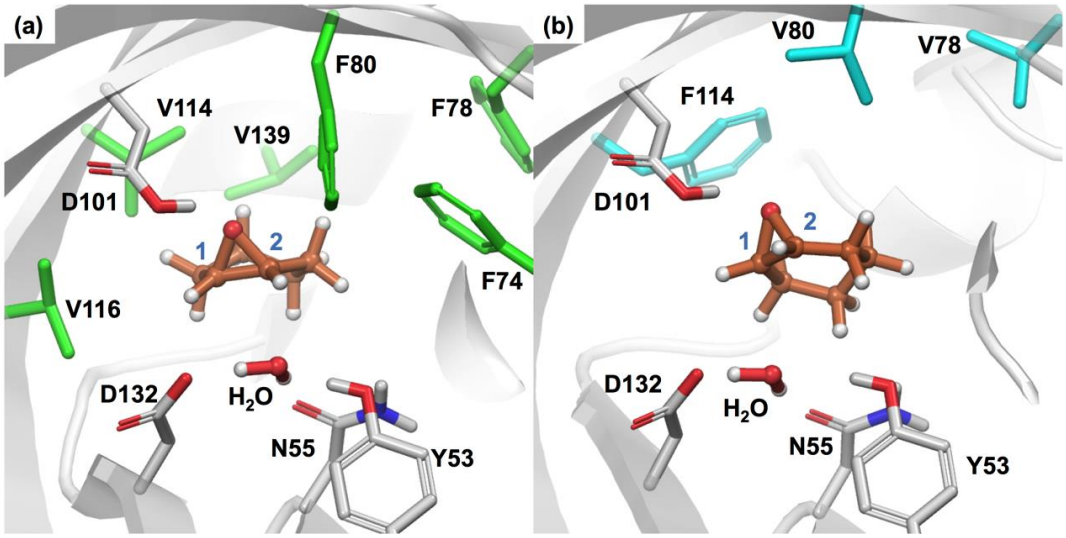


Figure 2. Highest ranked docking poses for cyclohexene oxide (**1**) in the (a) SZ724 and (b) SZ655 mutant models of LEH. The catalytic residues are shown in white stick representation. The mutated residues are shown in green and blue for the SZ724 and SZ655 mutants, respectively. Geometric parameters for the displayed docking poses are provided in Table S1.

Conclusions

In previous studies,[12] the hydrolytic desymmetrization of cyclohexene oxide (**1**) catalyzed by limonene epoxide hydrolase (LEH) with formation of (*R*,*R*)- and (*S*,*S*)-cyclohexane-1,2-diols (**2**) served as a model system for testing different directed evolution strategies. In the present study we have extended these efforts by applying and comparing single code saturation mutagenesis (SCSM) with double code saturation mutagenesis (DCSM). Both approaches focus combinatorial randomization on sites lining the binding pocket (CASTing),[1g,3] and both proved to be successful. However, the experimental data indicates that DCSM is the superior approach. It should be mentioned that in directed evolution of LEH as catalyst in the desymmetrization of **1**, triple code saturation mutagenesis (TCSM) using three appropriately chosen amino acids is also more efficient than SCSM and surpasses all alternative approaches.[12c]

The best (*R*,*R*)- and (*S*,*S*)-selective variants are characterized by sequences that have not been observed in previous CAST-based directed evolution studies of LEH,[12] but some of the point mutations do in fact occur in earlier mutants. On the basis of docking computations it was possible to develop simple models for explaining the source of enhanced and inverted stereoselectivity of the new variants. Some variants are also viable catalysts in asymmetric reactions of several other substrates. When focusing only on synthetic aspects, other epoxide hydrolases should also be considered.[13] Finally, we emphasize once more that the appropriate choice of highly reduced amino acid alphabets for beating the numbers problem in directed evolution of stereoselective enzymes needs to be guided by X-ray structures, consensus data or exploratory NNK-based saturation mutagenesis at individual CAST sites.[7,10,12] The optimal strategy will depend upon the particular enzyme under study.

Experimental Section

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 $MgSO_4$, 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™ 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 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

Primer design and library construction depend upon different grouping strategies and the particular amino acid chosen, and in the case of LEH this involves ten residues which were divided into two groups (Scheme 2): 1) Amplification of the short fragments of LEH using mixed primers F1/R1 for A5, F2/R2 for B5, F1/R1 and F2/R2 for C10, respectively; 2) using mixed primers F1/R2 by over-lap PCR[14] to create the full fragment for C10, using the purified fragment from step 1 as template; 3) Amplification of the whole plasmid pET22bLEHwt[12b-c] using the products of step 1 (A5, B5) or step 2 (C10) as megaprimers, leading to the final variety plasmids for library generation. Primers are listed in Tables S2-S4. The PCR products were digested by *Dpn* I and transformed into electro-competent *E. coli* BL21(DE3) to create the library for screening.

Screening Procedures

The same screening procedures described previously were used to assay all libraries.[12b-c]

Homology modelling

Models of the SZ724 and SZ655 mutants of LEH were constructed by homology modeling, using the previously determined crystal structures of the SZ92 (L74F/M78F/L103V/L114V/I116V/F139V/L147V) and SZ338 (L74F/M78V/I80V/L114F) mutants as templates, respectively.[12b] Homology modeling was performed using the Structure Prediction Wizard in Prime,[15] using the knowledge-based approach.[12b] The resulting structures were refined and prepared for docking using the Protein Preparation Wizard.[16] Hydrogen atoms were added according to the protonation states determined by PROPKA.[17] The protonation state of D101 was manually selected to be in the protonated (neutral) form. The coordinates of the nucleophilic water oxygen atom were copied from the WT LEH crystal structure (PDB 1NU3)[11b] following structural alignment with the mutant structures, ensuring that the known activating hydrogen bonds to Y53, N55 and D132 were formed. The positions of all hydrogen atoms were energy minimized using the Impact program[18] and the OPLS2005 forcefield.

Docking of cyclohexene oxide.

The structure of cyclohexene oxide was built in the Maestro program and prepared for docking using LigPrep.[19] Docking was performed to the prepared models of the SZ724 and SZ655 mutants using Glide[20] with standard precision (SP) settings. A maximum of ten docking poses was requested for each model. A total of 15 and 9 docking poses was obtained for the SZ724 and SZ655 models, respectively. Activation of the epoxide ring oxygen by D101 is a prerequisite for ring opening in LEH; therefore, only the docking poses that contained this interaction were considered during analysis.

Acknowledgements

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

Keywords: reduced amino acid alphabet • directed evolution • epoxide hydrolases • saturation mutagenesis • stereoselectivity

[1] Recent reviews of directed evolution: a) A. S. Bommarius, *Annu. Rev. Chem. Biomol. Eng*. **2015**, *6*, 319-345; b) A. Currin, N. Swainston, P. J. Day, D. B. Kell, *Chem. Soc. Rev*. **2015**, *44*, 1172-1239; c) C. A. Denard, H. Ren, H. Zhao, *Curr. Opin. Chem. Biol*. **2015**, *25*, 55-64; d) E. M. J. Gillam, J. N. Copp, D. F. Ackerley, in *Methods in Molecular Biology*., *Vol 1179*, Humana Press, Totowa, NJ, **2014**; e) M. Goldsmith, D. S. Tawfik, *Methods Enzymol*. **2013**, *523*, 257-283; f) E. M. Brustad, F. H. Arnold, *Curr. Opin. Chem. Biol*. **2011**, *15*, 201-210; g) M. T. Reetz, *Angew. Chem. Int. Ed*. **2011**, *50*, 138-174; h) C. Jäckel, D. Hilvert, *Curr. Opin. Biotechnol*. **2010**, *21*, 753-759; i) N. J. Turner, *Nat. Chem. Biol*. **2009**, *5*, 568-574; j) S. Lutz, U. T. Bornscheuer, *Protein Engineering Handbook*, Wiley−VCH, Weinheim, **2009**.

[2] Recent reviews of saturation mutagenesis: a) E. Williams, J. N. Copp, D. F. Ackerley, in *Directed Evolution Library Creation*, (Eds.: E. M. J. Gillam, J. N. Copp, F. Ackerley), *Methods in Molecular Biology*, *Vol. 1179*, Humana Press, Totowa, NJ, **2014**, pp. 83-101; b) R. M. P. Siloto, R. J. Weselake, *Biocatal. Agric. Biotechnol*. **2012**, *1*, 181-189.

[3] M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem.* **2005**, *117*, 4264-4268; *Angew. Chem. Int. Ed.* **2005**, *44*, 4192-4196.

[4] a) M. T. Reetz, L.-W. Wang, in part M. Bocola, *Angew. Chem.* **2006**, *118*, 1258-1263; *Erratum*, 2556; *Angew. Chem. Int. Ed.* **2006**, *45*, 1236-1241; *Erratum*, 2494; b) C. G. Acevedo-Rocha, S. Kille, M. T. Reetz, In *Directed Evolution Library Creation*: Methods and Protocols, 2nd Edition, Vol 1179 (Eds.: D. Ackerley, J. Copp, E. Gillam) *Methods in Molecular Biology,* Humana Press, Totowa, **2014**, pp 103-128.

[5] a) A. R. Davidson, K. J. Lumb, R. T. Sauer, *Nat. Struct. Biol*. **1995**, *2*, 856-864; b) K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem*. **2005**, *280*, 37742-37746.

[6] a) F. A. Fellouse, B. Li, D. M. Compaan, A. A. Peden, S. G. Hymowitz, S. S. Sidhu, *J. Mol. Biol*. **2005**, *348*, 1153-1162; b) A. Koide, R. N. Gilbreth, K. Esaki, V. Tereshko, S. Koide, *Proc. Natl. Acad. Sci. U S A.* **2007**, *104*, 6632-6637.

[7] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891-903.

[8] W. M. Patrick, A. E. Firth, *Biomol. Eng*. **2005**, *22*, 105−112.

[9] Y. Nov, *Appl. Environ. Microbiol*. **2012**, *78*, 258−262.

[10] Z. Sun, Y. Wikmark, J.-E. Bäckvall, M. T. Reetz, *Chem. Eur. J.* **2016**, *22*, 5046-5054.

[11] a) M. J. van der Werf, K. M. Overkamp, J. A. M. de Bont, *J. Bacteriol.* **1998**, *180*, 5052–5057; b) M. Arand, B. M. Hallberg, J. Zou, T. Bergfors, F. Oesch, M. J. van der Werf, J. A. M. de Bont, T. A. Jones, S. L. Mowbray, *EMBO J.* **2003**, *22*, 2583–2592.

[12] a) H. Zheng, M. T. Reetz, *J. Am. Chem. Soc.* **2010**, *132*, 15744-15751; b) Z. Sun, R. Lonsdale, X.-D. Kong, J.-H. Xu, J. Zhou, M. T. Reetz, *Angew. Chem.* **2015**, *127*, 12587-12592. *Angew. Chem. Int. Ed*. **2015**, *54*, 12410-12415; c) Z. Sun, R. Lonsdale, L. Wu, G. Li, A. Li, J. Wang, J. Zhou, M. T. Reetz, *ACS Catal*. **2016**, *6*, 1590-1597.

[13] Reviews of epoxide hydrolases: a) M. Kotik, A. Archelas, R. Wohlgemuth, *Curr. Org. Chem*. **2012**, *16*, 451-482; b) J. H. L. Spelberg, E. J. de Vries, In *Enzyme Catalysis in Organic Synthesis*, 3rd Edition. (Eds.: K. Drauz, H. Gröger, O. May) Wiley-VCH, Weinheim, **2012**, pp 363-416; c) S. Hwang, C. Y. Choi, E. Y. Lee, *J. Ind. Eng. Chem.* **2010**, *16*, 1-6; d) M. Widersten, A. Gurell, D. Lindberg, *Biochim. Biophys. Acta*. **2010**, *1800*, 316–326.

[14] A. V. Bryksin, I. Matsumura, *Biotechniques*. **2010**, *48*, 463-465.

[15] Prime, version 3.9, Schrödinger, LLC, New York, NY, **2015**.

[16] Schrödinger Suite 2015-1 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, NY, **2015**.

[17] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525.

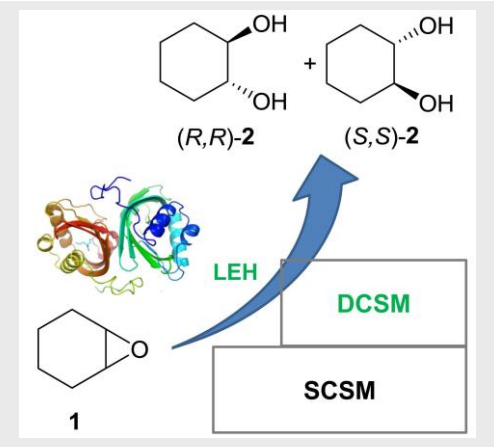
[18] Impact, version 6.5, Schrödinger, LLC, New York, NY, **2014**.

[19] LigPrep, version 3.2, Schrödinger, LLC, New York, NY, **2014**.

[20] Glide, version 6.5; Schrödinger, LLC: New York, NY, **2014**.

FULL PAPERS

Two is better than one: Double code saturation mutagenesis (DCSM) based on two amino acids as building blocks is more efficient than single code saturation mutagenesis (SCSM) in the directed evolution of enzyme stereoselectivity, as demonstrated by the hydrolytic desymmetrization of cyclohexene oxide catalyzed by limonene epoxide hydrolase.



*Z. Sun, R. Lonsdale, G. Li, M. T. Reetz**

Page No. 1– Page No. 8
Comparing Different Strategies in Directed Evolution of Enzyme Stereoselectivity: Single versus Double Code Saturation Mutagenesis