Beating Bias in Directed Evolution of Biocatalysts by Using High Fidelity Solid Phase Library Synthesis

INTRODUCTION

Directed evolution is a powerful technique for the creation of novel biocatalysts that have many applications in organic synthesis and biotechnology. It promises to eliminate limited substrate scope, insufficient activity, and poor regioselectivity or stereoselectivity of natural biocatalysts.

This approach has not been utilized to its full potential because of shortcomings in common PCR-based mutagenesis techniques, such as error prone PCR (epPCR) and saturation mutagenesis (SM). The main weaknesses of these techniques are codon bias and unwanted insertions and deletions which reduce the quality of the library and increase the screening required.

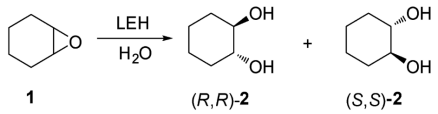
Here we show how these issues can be mitigated by using high-fidelity solid-phase chemical gene synthesis on silicon chips, followed by efficient gene assembly, for combinatorial library construction. To assess the efficiency of the method, we use an epoxide hydrolase (the limonene epoxide hydrolase, LEH), the catalyst in the enantioselective hydrolytic desymmetrization of cyclohexene oxide with stereoselective formation of (*R,R*)- and (*S,S*)-cyclohexane-1,2-diol, as an experimental model.

In this study, we present a side-by-side comparison of the results obtained with a traditional combinatorial PCR-based SM and a synthetic library produced using Twist Bioscience's on-chip solid-phase DNA synthesis technology. This comprehensive evaluation of the two mutagenesis methodologies demonstrates that the synthetic DNA approach is more efficient at obtaining the desired mutants than the traditional PCR-based approach, opening up new perspectives for directed evolution.

EXPERIMENTAL DESIGN

Experimental Model

The experimental model used in this work is the limonene epoxide hydrolase (LEH) from *Rhodococcus erythropolis* (van der Werf et al., 1998) , a catalyst of the hydrolytic desymmetrization of cyclohexene oxide (**1**) with formation of (*R,R*)-**2** and (*S,S*)-**2** (Scheme 1).



Scheme 1. LEH-catalyzed hydrolytic desymmetrization of epoxide 1.

Wild type LEH leads to poor enantioselectivity with a enantiomeric ratio of 48:52 in slight favor of (*S,S*)-**2**. Structural analysis of LEH (Arand et al., 2003) and SM-based directed evolution studies (Sun et al., 2016; Zheng and Reetz, 2010) identified four residues (M78/I80/L114/I116) lining the substrate binding pocket that are critical for enantioselectivity. These residues were targeted for combinatorial PCR-based mutagenesis. Three amino acids, valine, phenylalanine, and tyrosine (V-F-Y) were introduced at these four positions using "triple code saturation mutagenesis" (TCSM) instead of the usual NNK codon degeneracy. Theoretically, 256 DNA mutants can be expected from this approach (Fig. 1).

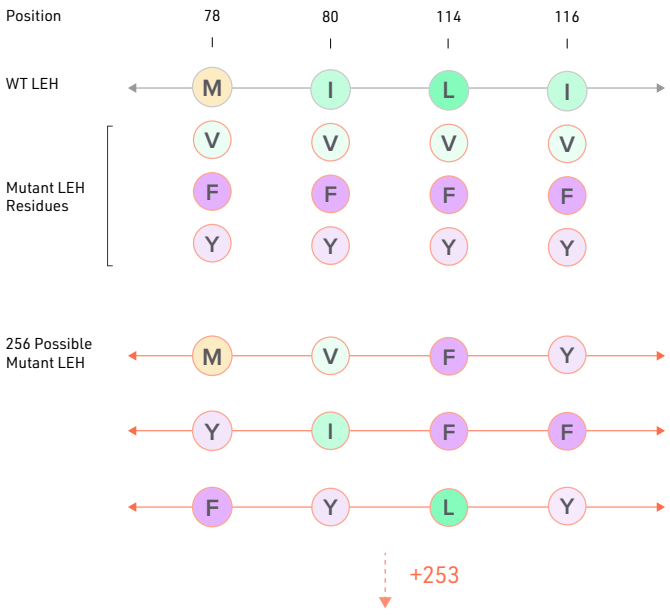


Figure 1. Mutagenic strategy to produce all combinations of mutations at four positions.

PCR-Based Construction of LEH SM library

Construction of the LEH library using standard combinatorial PCR base SM mutagenesis is summarized in Fig. 2A. It is comprised of a PCR step using a mixture of mutant primers to incorporate the various combinations of the mutations in the DNA of interest, followed by a whole plasmid amplification using the amplified fragments as megaprimers. After removal of the WT plasmid by restriction enzyme digestion, the PCR amplified plasmids harboring the different mutations are transformed in *E. coli* BL21 (DE3) to make the library.

Solid-Phase Chemical Synthesis of LEH Mutant Library

Oligonucleotide primers with predefined composition and combinatorial complexity were synthesized on a silicon plate in equimolar ratios (Banyai et al., 2015) and extracted from the plate. The primers were then mixed and assembled into full-length genes by overlap extension PCR to generate the desired variants (Fig. 2B). The fragments were digested with *Nde I* and *Hind* III and ligated into the plasmid pET 22b. The recombinant plasmids containing the *LEH* gene mutants were transformed into *E. coli* BL21 (DE3).

Quality Assessment of Libraries

The two libraries were compared in terms of quality at the DNA level by massively parallel sequencing analysis using three oversampling factors (1, 2, and 3). Four criteria were considered for quality assessment:

- **Genetic diversity:** the ratio of the number of mutants obtained to the number of theoretically possible mutants (256 mutants in this case)
- **Sequence integrity:** the proportion of incomplete sequences, and sequences with deletions and insertions
- **Cloning efficiency:** the proportion of clones bearing a mutant gene sequence vs. wild type sequence
- **Codon bias:** the frequency of target residues at each mutated position.

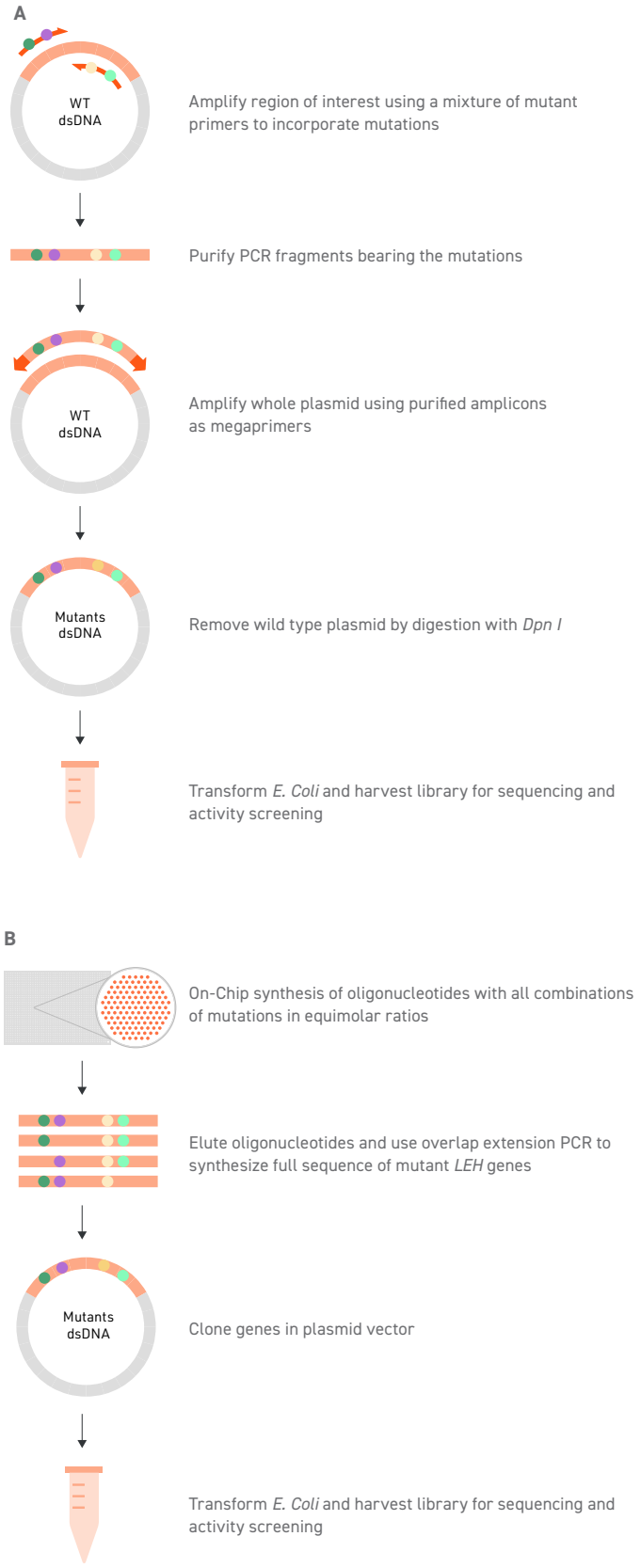


Figure 2. (A) LEH library construction using the standard combinatorial PCR- based SM-approach. (B) LEH library construction using solid phase DNA synthesis approach.

RESULTS AND DISCUSSION

TABLE 1

OVERSAMPLING FACTOR	NUMBER OF COLONIES SEQUENCED	LIBRARIES	NUMBER OF DIFFERENT MUTANTS FOUND	NUMBER OF MUTANTS WITH INCOMPLETE SEQUENCE*	NUMBER OF WILD TYPE CLONES	GENETIC DIVERSITY (%)**
1	276	PCR Based	109	36	4	42.5
		Solid phase gene synthesis	164	18	2	64.1
2	552	PCR Based	135	88	8	52.7
		Solid phase gene synthesis	218	47	2	85.2
3	828	PCR Based	144	130	11	56.3
		Solid phase gene synthesis	249	65	4	97.3

* Number of colonies containing deletions/insertions for which sequencing was not possible for various reasons.
** Ratio of the number of mutants obtained after sequencing to the number of theoretically possible mutants (256 mutants) expressed in percentage.

Genetic Diversity

When using an oversampling factor of 3, 97% of the designed genetic diversity was achieved with the solid-phase DNA synthesis approach: 249 out of the 256 theoretically possible mutants were found in the library (Table 1). In contrast, when using the same oversampling factor of 3, the combinatorial PCR-based SM-library yielded a maximum genetic diversity of 56%, and only 144 out of 256 mutants were formed (Table 1). With only 56% genetic diversity, it is unlikely that the high activity mutants found using the synthetic DNA approach would have been found in the PCR based library, even if screening a much larger number of clones. The synthetic DNA approach makes it possible to obtain most of the mutants expected from a saturation mutagenesis experiment and identify proteins with improved properties that would not be discovered in a PCR-based combinatorial library.

Genetic diversity was also assessed by looking at the number of times a given mutant appears in the library (at oversampling factor of 3). As shown in Fig. 3, for the combinatorial PCR-based SM library, over 40% of the 144 mutants identified are found in 5 to 8 instances and some mutants are found 10 and even 17 times in the library. In sharp contrast, in the synthetic library, the mutants present in 2 to 4 copies account for 60% of the total number of clones. Moreover, the maximum instances of the mutants in the synthetic library is only 8.

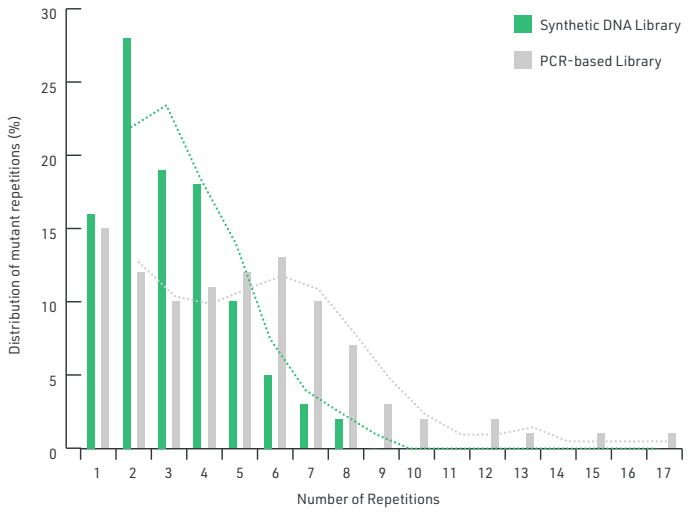


Figure 3. Distribution of mutants with different repetitions in both PCR-based SM library and on-chip synthesized DNA-based library at oversampling factor of 3. Repetition means the number of appearances of a given mutant in the respective libraries. Y axis represents the percentage of repetitions relative to the total number of mutants.

Sequence Integrity

Table 1 shows that the number of mutants with an incomplete sequence is more than twice as high in the PCR-based SM library as in the synthetic library. This may be caused by errors arising from the two PCR steps, but it should be noted that a few deletions and insertions were also observed in the synthetic library. These errors may be introduced during the assembly step.

Mutant Cloning Efficiency

The number of wild type transformants occurring in the conventional PCR-based library (1.3% of the library clones at an oversampling factor of 3) is almost three times higher than that of the synthetic gene library (0.48% of the library clones) (Table 1).

Distribution of Residues at Each Position (Codon Bias)

In the PCR-based SM library the frequency of different residues at each mutated position varies widely when an oversampling factor of 3 is used. The wild type amino acid at each position shows the highest frequency (Fig. 4), suggesting a codon bias. In contrast, in the synthetic library, the frequency of the different residues at each position is almost identical, with a nearly even distribution (Fig. 4). Similar trends are also observed with oversampling factor of 1 and 2 (data not shown). These results show that the synthetic library does not suffer a significant codon bias, which results in higher genetic diversity.

Figure 4. Distribution of different residues (M, F, V and Y) at positions 78, 80, 114 and 116 of LEH in PCR-based SM library (left) and in on-chip gene synthesis based library (right) at oversampling factor of 3.

Functional Screening of the Libraries

A sample of 276 clones of each library were individually cultivated and cell extracts were used in the hydrolytic desymmetrization of cyclohexene oxide (**1**) with formation of (*S,S*)- and (*R,R*)-**2** (Table 2). In the case of (*S,S*)-selectivity, only variants showing enantiomeric ratios ≥ 95 : 5 are defined as positive hits, while the threshold for (*R,R*)-selectivity (which appears to be more difficult to evolve) is lowered to 10:90. These results show that the synthetic library produces more highly effective mutants with respect to enantioselectivity than the PCR-based SM library, while all of the mutants obtained with the PCR-based approach are also found in the synthetic library.

Screening results for PCR-based SM library and on-chip gene synthesis library considering only (S,S)-variants displaying enantiomeric ratios of ≥ 95:5 for (S,S)-selectivity and at least 10:90 for (R,R)-selectivity are shown. Bold-marked variants in the synthetic library refer to those that are also found in the combinatorial PCR-based SM library.

TABLE 2: FUNCTIONAL SCREENING OF MUTANT LIBRARIES

LIBRARY	CLONE #	MUTANTS	CONVERSION (%)*	PRODUCT ENANTIOMERIC RATIO**
PCR-based SM library	PCR1-18	M78F/I80F/L114V/I116V	>99	98 : 2 *(S,S)*
	PCR1-92	I80Y/L114V/I116V	>99	>99 : 1 *(S,S)*
	PCR2-48	M78F/I80Y/L114V	98	97 : 3 *(S,S)*
	PCR2-87	M78F/I80Y/L114V/I116V	>99	>99 : 1 *(S,S)*
	PCR2-55	M78Y/I80Y/I116V	98	98 : 2 *(S,S)*
	PCR1-75	I80Y/I116V	98	98 : 2 *(S,S)*
	PCR2-45	I80V/L114F	80	92 : 8 *(R,R)*
	PCR1-83	M78V/I80V/L114V	99	90 : 10 *(R,R)*
Synthetic library	TW1-08	**M78F/I80F/L114V/I116V**	>99	98 : 2 *(S,S)*
	TW1-20	**I80Y/L114V/I116V**	>99	>99 : 1 *(S,S)*
	TW1-27	M78Y/I80F/L114V/I116V	>99	96 : 4 *(S,S)*
	TW1-41	M78Y/I80Y/L114V/I116V	94	98: 2 *(S,S)*
	TW1-56	M78V/I80Y/L114V/I116V	98	99 : 1 *(S,S)*
	TW1-81	**M78F/I80Y/L114V**	98	97 : 3 *(S,S)*
	TW1-71	**M78F/I80Y/L114V/I116V**	>99	>99 : 1 *(S,S)*
	TW3-35	M78V/I80F/L114V/I116V	98	96 : 4 *(S,S)*
	TW3-52	M78F/I80Y/I116V	97	97 : 3 *(S,S)*
	TW3-62	**M78Y/I80Y/I116V**	98	98 : 2 *(S,S)*
	TW3-67	**I80Y/I116V**	98	98 : 2 *(S,S)*
	TW1-45	**I80V/L114F**	80	92 : 8 *(R,R)*
	TW2-74	M78Y/I80V/L114F	60	95 : 5 *(R,R)*
	TW3-32	**M78V/I80V/L114V**	99	90 : 10 *(R,R)*
	TW3-37	M78F/I80V/L114F	98	96 : 4 *(R,R)*
	TW3-45	M78V/I80V/L114F/I116V	99	90 : 10 *(R,R)*
	TW3-76	M78Y/I80V	98	92 : 8 *(R,R)*

*Determined by GC, ** determined by chiral GC analysis

CONCLUSIONS AND PERSPECTIVES

Saturation mutagenesis (SM) has been applied successfully in directed evolution for enhancing enzyme properties such as activity, regio- and stereoselectivity as well as stability (Reetz, 2016; Sullivan et al., 2013; Sun et al., 2016; Zheng and Reetz, 2010). In most cases, screening the library is the labor-intensive step in the process and constitutes the bottleneck of directed evolution (Acevedo-Rocha et al., 2014). The development of more effective methodologies that reduce sequence errors and codon biases, that are inherent in traditional PCR-based approaches, is critical to take full advantage of directed evolution for protein engineering.

In this study, we demonstrate that a novel approach using solid phase synthesis of oligonucleotides, in which codon bias and sequence errors are greatly reduced, allows the identification of novel mutants with improved properties. The synthetic combinatorial library has shown to be superior to the PCR-based method with respect to all four criteria that are considered for quality assessment. Two of these criteria, codon bias and sequence integrity, are essential for removing the screening bottleneck of directed evolution.

While the use of synthetic libraries increases the efficiency of directed evolution, success is contingent to developing a smart diversity design. A detailed knowledge of the protein structure and function helps narrow down on the essential elements of the active site of the protein to be engineered and allows the design of an effective strategy. The high fidelity synthetic DNA approach to generate saturated mutant libraries has great potential to be applicable other biocatalysts and proteins and accelerate discovery of proteins with enhanced properties and functionalities.

EXPERIMENTAL PROCEDURES

Products used in this work: KOD Hot Start DNA Polymerase (Novagen), restriction enzymes (*Dpn* I, *Nde* I and *Hind* III) and T4 DNA ligase (New England Biolabs), oligonucleotides (Life Technologies), plasmid preparation kit (Zymo Research), PCR purification kit (QIAGEN), Lysozyme and DNase I (AppliChem). All commercial chemicals were purchased from Sigma-Aldrich, Tokyo Chemical Industry (TCI) or Alfa Aesar. DNA sequencing was conducted by GATC Biotech.

PCR-based SM combinatorial library construction: The library was constructed using the standard megaprimer approach with KOD Hot Start polymerase as shown in Fig. 1A and as described previously (Sullivan et al., 2013; Zheng and Reetz, 2010). Amplification of the *LEH* gene fragments (megaprimers) was done using mixed primers (100 µM) and pET22bLEHwt as template (50–100 ng) using the following conditions: 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. Amplification of the whole plasmid using the megaprimers (100 ng) and pET22bLEHwt (50–100 ng) as template was performed using the following conditions: 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. 3). *Dpn* I digestion was carried out at 37 °C for 6 h after addition of NEB CutSmart™ Buffer and 2 µL *Dpn* I to the 50 µl PCR reaction. PCR products (1 to 2 µL) were then directly transformed into *E.coli* BL21(DE3) by electroporation to create the PCR-based SM combinatorial library.

Synthetic combinatorial library construction: The synthetic library was created with an on-chip silicon-based DNA writing platform as developed by Twist Bioscience as illustrated in Fig. 1B. The full-length gene fragments obtained by overlap extension PCR (1 µg) were digested with *Nde* I and *Hind* III at 37 °C for 6 h and ligated into the plasmid pET22b cut with the same enzymes. The recombinant plasmids carrying the LEH gene mutants (1–2 µL) were transformed into *E. coli* BL21(DE3) to create the synthetic combinatorial library.

Cyclohexene oxide activity: Cyclohexene oxide activity was measured by chiral GC analysis in bacterial cell extracts as described previously (Sun et al., 2016).

ACKNOWLEDGEMENTS

Twist Bioscience would like to thank Manfred T. Reetz, Aitao Li, Carlos G. Acevedo-Rocha, Zhoutong Sun, Tony Cox, and Jia (Lucy) Xu for allowing us to write about their research and findings. The work presented here was published in ChemBioChem 19, 221–228 in February 2018 with a follow-up concept article published in ChemBioChem 10.1002/cbic.201800339 in July 2018.

REFERENCES

Acevedo-Rocha, C.G., Agudo, R., and Reetz, M.T. (2014). Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems. J. Biotechnol. *191*, 3–10.

Arand, M., Hallberg, Bm., Zou, J., Bergfors, T., Oesch, F., van der Werf, M.J., de Bont, J.A., Jones, Ta., and Mowbray, S.L. (2003). Structure of *Rhodococcus erythropolis* limonene-1, 2-epoxide hydrolase reveals a novel active site. EMBO J. *22*, 2583–2592.

Banyai, W., Peck, B.J., Fernandez, A., Chen, S., and Indermuhle, P. (2015). De novo synthesized gene libraries (Google Patents).

Reetz, M.T. (2016). Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology (Weinheim: Wiley-VCH).

Sullivan, B., Walton, A.Z., and Stewart, J.D. (2013). Library construction and evaluation for site saturation mutagenesis. Enzyme Microb. Technol. *53*, 70–77.

Sun, Z., Lonsdale, R., Wu, L., Li, G., Li, A., Wang, J., Zhou, J., and Reetz, M.T. (2016). Structure-Guided Triple-Code Saturation Mutagenesis: Efficient Tuning of the Stereoselectivity of an Epoxide Hydrolase. ACS Catal. *6*, 1590–1597.

van der Werf, M.J., Overkamp, K.M., and de Bont, J.A. (1998). Limonene-1, 2-epoxide hydrolase from *Rhodococcus erythropolis* DCL14 belongs to a novel class of epoxide hydrolases. J. Bacteriol. *180*, 5052-5057.

Zheng, H., and Reetz, M.T. (2010). Manipulating the Stereoselectivity of Limonene Epoxide Hydrolase by Directed Evolution Based on Iterative Saturation Mutagenesis. J. Am. Chem. Soc. *132*, 15744–15751.