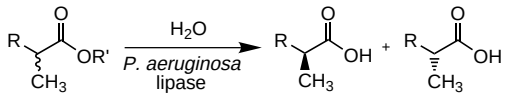
Directed Evolution of an Enantioselective Enzyme through Combinatorial Multiple-Cassette Mutagenesis

Manfred T. Reetz,* Stephanie Wilensek, Dongxing Zha, and Karl-Erich Jaeger

In 1997 we demonstrated that the methods of directed evolution of functional proteins[1–4] can be applied successfully in the quest to create enantioselective enzymes for use in organic synthesis.[5] The combination of suitable mutagenesis and gene-expression methods with high-throughput assays for determining the enantioselectivity of thousands of enzyme mutants forms the basis of a fundamentally new concept in the area of asymmetric catalysis. In only four cycles of random mutagenesis based on the error-prone polymerase chain reaction (epPCR) and high-throughput screening of about 8000 enzyme variants the selectivity factor E[6] of the lipase-catalyzed hydrolytic kinetic resolution of ester **1** was increased from 1.1 (with the wild-type enzyme) to 11 (with



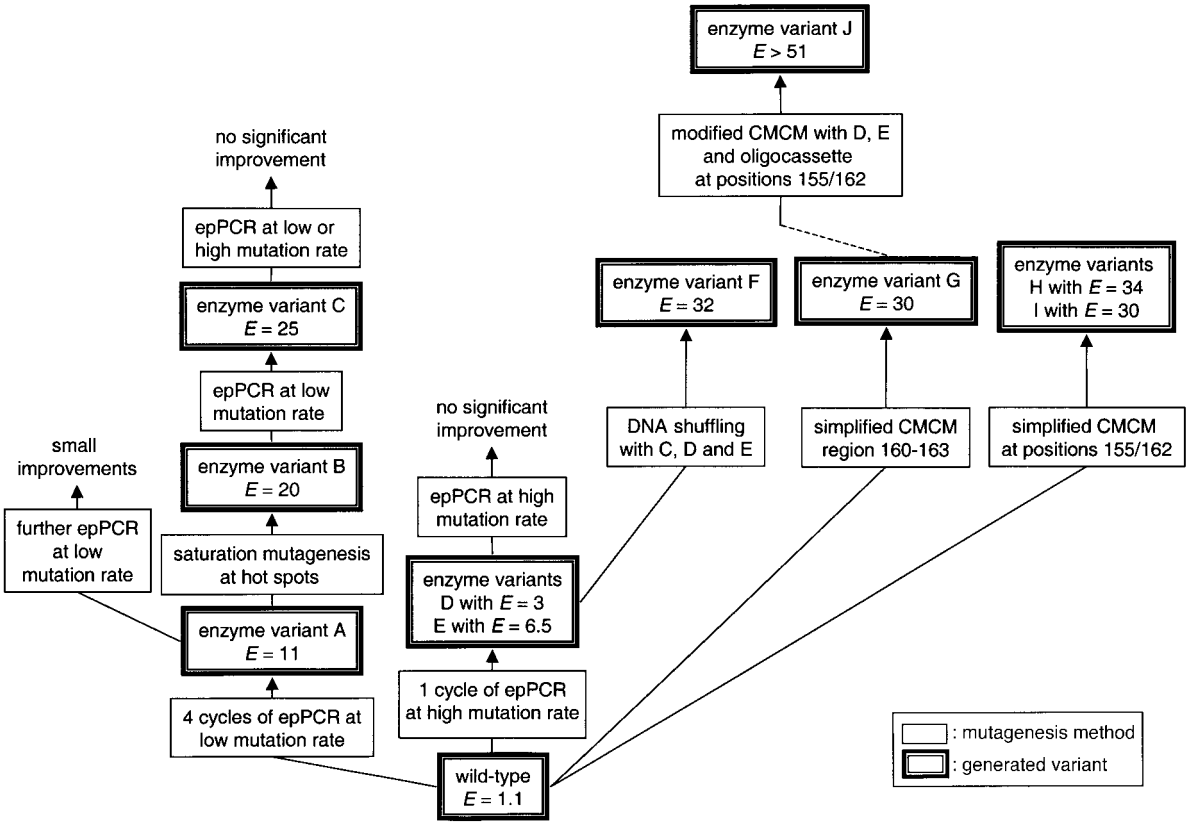
enzyme variant A; Scheme 1, bottom left) in favor of the *S* acid **2**. The epPCR experiments were conducted with the gene of the bacterial lipase from *Pseudomonas aeruginosa* at a low mutation rate corresponding to an average of one amino acid substitution per enzyme. This particular lipase is composed of 285 amino acids, a number that needs to be considered when calculating the theoretical protein-sequence space.[7] Since only small improvements in enantioselectivity were observed in further rounds of epPCR, saturation mutagenesis was applied at sensitive positions ("hot spots", identified by amino acid sequencing) of the improved enzyme variants, which led to the identification of a variant B displaying significantly increased enantioselectivity ($E = 20$; Scheme 1, middle left).[8] When another cycle of epPCR was added on top of this, variant C ($E = 25$) with five mutations (V47G, V55G, S149G, S155F, and S164G) was obtained (Scheme 1, left). However, the strategy of alternating epPCR and saturation mutagenesis seems to have limitations, since further experiments of this kind did not result in significant improvements. We now show for the first time that recombinant methods are viable tools in improving the enantioselectivity of an enzyme.

Since the hitherto best enzyme variant C was evolved on the basis of incremental improvements brought about sequentially by small changes (single amino acid exchanges), it might be argued that back-crossing with DNA shuffling[3] could improve the functional properties of the enzyme by eliminating mutations which might have a negative effect in concert with the rest of the mutations. However, upon applying DNA shuffling with the wild-type gene and the mutant gene encoding variant C, no significant improvements were observed. In order to create altered genetic material of higher diversity which is not a result of small incremental changes for subsequent utilization in DNA shuffling, the wild-type gene was subjected to epPCR at a considerably higher mutation rate, which corresponded to about three amino acid exchange events per enzyme molecule.[7] Upon screening 15000 variants several hits were found which had three mutations and displayed enhanced *S* selectivity relative to the wild-type, for example, variant D (with the mutations S53P, C180T, and G272A; $E = 3$) and variant E (D20N, S161P, and T234S; $E = 7$). The two relevant genes plus the gene encoding variant C were then subjected to DNA shuffling (fragment size 30–50 base pairs (bp)). This experiment turned out to be much more successful than the previous DNA shuffling because it provided a small pool of very enantioselective enzyme variants; the best one (variant F) had five mutations (V47G, S149G, S155F, S199G, and T234S) and showed an E value of 32 (Scheme 1, right).

Although further navigation in protein-sequence space[7] on the basis of epPCR, saturation mutagenesis, and/or DNA shuffling might well be successful, we decided to attempt other types of methodological developments by turning to cassette mutagenesis[4] with a simplified version of combinatorial multiple-cassette mutagenesis (CMCM).[9] The original form of CMCM as described by Crameri and Stemmer is a special type of DNA shuffling which was applied in the area of functional antibodies and which allows the generation of mutant-gene libraries in which the wild-type gene and cassettes composed of defined sequences are randomized.[9a] On the basis of our previous experiments concerning the identification of hot spots, the positions around amino acid number 160 appeared to constitute a hot region.[5, 8] Since cassette mutagenesis should not be applied to exceedingly large regions,[4] we restricted it to positions 160–163. Mutagenesis was performed using only the wild-type gene and a single oligocassette was synthesized by employing equimolar amounts of nucleotide mixtures (simplified CMCM). This ensures saturation at all four positions. Upon screening a library of 5000 variants, several enantioselective enzymes were found, with the best one (variant G with $E = 30$) being characterized by the mutations E160A, S161D, L162G, and N163F (Scheme 1, right). Since earlier work had shown that the introduction of glycine often has beneficial effects,[8] we focused on mutation L162G. Having previously identified position 155 as another important hot spot,[5b] we inserted a cassette (size: 69 bp) with double saturation at positions 155 and 162 into the wild-type gene. Accordingly, these two positions were saturated with equimolar nucleotide mixtures

[*] Prof. Dr. M. T. Reetz, S. Wilensek, Dr. D. Zha
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
Priv.-Doz. Dr. K.-E. Jaeger
Lehrstuhl für Biologie der Mikroorganismen
Ruhr-Universität Bochum
44780 Bochum (Germany)

  1433-7851/01/4019-3589 $ 17.50+.50/0



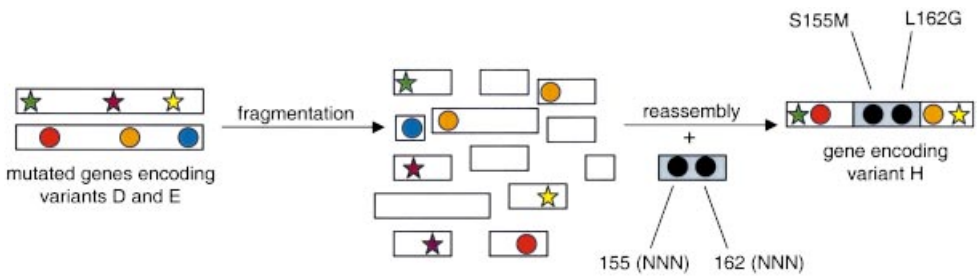
Scheme 1. Schematic representation of the directed evolution of enantioselective enzymes (lipase variants) which catalyze the hydrolytic kinetic resolution of ester **1**.

in a simplified CMCM process. Following expression and screening two very enantioselective variants were identified: Variants H (S155S, L162G) and I (S155V, L162G) which displayed E values of 34 and 30, respectively (Scheme 1, right). In both cases the mutation L162G is conserved, which substantiates earlier speculations concerning the positive role of this particular amino acid substitution.[8]

This type of search program restricts the accessible protein-sequence space considerably. We therefore decided to extend it by developing a strategy based on the extension of the CMCM principle. This was accomplished by performing DNA shuffling[3, 9a] with mutant genes encoding enzyme variants D and E and an oligocassette with simultaneous saturation at positions 155 and 162 (Scheme 2). This ensures the possibility of maximum recombination. Indeed, this approach turned out to be successful because several highly enantioselective enzymes were found, among them variant J with six exchanged amino acids (D20N, S53P, S155M, L162G, T180I, and T234S). The E value of >51 ($ee \geq 95\%$ at 24% conversion) means that the enantioselectivity has almost been doubled (Scheme 1, top right).

This work demonstrates that appropriate forms of recombinant methods are successful in the directed evolution of enantioselective enzymes. We have learned that protein-sequence space with respect to enantioselectivity is best explored by undertaking three steps: 1) Generation of mutants by epPCR at a high-mutation rate; 2) identification of hot regions and spots in the enzyme by epPCR and substantiation of them by simplified CMCM; and 3) extension of the process of CMCM[9] to cover a defined region of protein sequence space. Less than a total of 40000 enzyme variants were screened.[10] We are currently applying this protocol in the directed evolution of other enantioselective enzymes,[11] but also expect it to be useful in the evolutionary improvement of other enzyme properties.

Received: April 5, 2001 [Z16907]



Scheme 2. Extended CMCM in the evolution of an *S*-selective lipase variant J; see Scheme 1, right (green star: position 20; purple star: position 161; yellow star: position 234; red circle: position 53; orange circle: position 180; blue circle: position 272).

[1] a) D. W. Leung, E. Chen, D. V. Goeddel, *Technique (Philadelphia)* **1989**, *1*, 11–15; b) R. C.

  1433-7851/01/4019-3590 $ 17.50+.50/0

Cadwell, G. F. Joyce, *PCR Methods Appl.* **1994**, *3*, S136-S140; c) "Combinatorial Chemistry in Biology": B. Steipe, *Curr. Top. Microbiol. Immunol.* **1999**, *243*, 55–86.
[2] a) F. H. Arnold, *Acc. Chem. Res.* **1998**, *31*, 125–131; b) F. H. Arnold, *Nature* **2001**, *409*, 253–257.
[3] a) W. P. C. Stemmer, *Nature* **1994**, *370*, 389–391; b) A. Crameri, S.-A. Raillard, E. Bermudez, W. P. C. Stemmer, *Nature* **1998**, *391*, 288–291.
[4] M. Kammann, J. Laufs, J. Schell, B. Gronenborn, *Nucleic Acids Res.* **1989**, *17*, 5405.
[5] a) M. T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem.* **1997**, *109*, 2961–2963; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2830–2832; b) M. T. Reetz, K.-E. Jaeger, *Chem. Eur. J.* **2000**, *6*, 407–412; c) M. T. Reetz, *Pure Appl. Chem.* **2000**, *72*, 1615–1622.
[6] The selectivity factor E reflects the ratio of the relative rates of reaction of the two enantiomers.
[7] The number N of theoretically possible lipase variants can be calculated on the basis of the algorithm $N = 19^M \cdot 285!/[(285 - M)! \cdot M!]$, where M is the number of exchanged amino acids per enzyme molecule.[5] Thus, for $M = 1$: $N = 5415$; $M = 2$: $N \cong 15 \times 10^6$; $M = 3$: $N \cong 52 \times 10^9$. It can be speculated that in the vast majority of cases these variants will not catalyze the desired reaction at all.
[8] K. Liebeton, A. Zonta, K. Schimossek, M. Nardini, D. Lang, B. W. Dijkstra, M. T. Reetz, K.-E. Jaeger, *Chem. Biol.* **2000**, *7*, 709–718.
[9] a) A. Crameri, W. P. C. Stemmer, *BioTechniques* **1995**, *18*, 194–196; see also previous work on combinatorial cassette mutagenesis: b) J. F. Reidhaar-Olson, R. T. Sauer, *Science* **1988**, *241*, 53–57; c) J. D. Hermes, S. C. Blacklow, J. R. Knowles, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 696–700; d) A. P. Arkin, D. C. Youvan, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7811–7815.
[10] a) The original UV/Vis-based screening system[5a] was used to identify hits. In all cases these enzyme mutants were then studied with the racemic substrate (**1**) in separate experiments, and with analysis being carried out by conventional chiral GC. The *p*-nitrophenol ester **1** was employed in all cases because the hydrolysis product (*p*-nitrophenol) is easily detected by UV/Vis spectroscopy. The disadvantage is that the ester is activated and therefore undergoes a noncatalyzed background reaction to a small extent. This has been considered when reporting E values higher than 30. In control experiments the ethyl ester showed lower activity and selectivity, which is not surprising. Other substrates have not yet been studied. b) Review of high-throughput screening systems: M. T. Reetz, *Angew. Chem.* **2001**, *113*, 292–320; *Angew. Chem. Int. Ed.* **2001**, *40*, 284–310.
[11] Other recent publications concerning directed evolution of enantioselective enzymes: a) U. T. Bornscheuer, J. Altenbuchner, H. H. Meyer, *Biotechnol. Bioeng.* **1998**, *58*, 554–559; b) O. May, P. T. Nguyen, F. H. Arnold, *Nat. Biotechnol.* **2000**, *18*, 317–320; c) S. Fong, T. D. Machajewski, C. C. Mak, C. H. Wong, *Chem. Biol.* **2000**, *7*, 873–883.