Directed Evolution

Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test**

Manfred T. Reetz, Marco Bocola, José Daniel Carballeira, Dongxing Zha, and Andreas Vogel*

Directed evolution of proteins has emerged as a powerful tool to enhance the stability, activity, and selectivity of enzymes.[1] It involves the proper combination of random gene mutagenesis, expression of thousands of mutant enzymes, and high-

[*] Prof. Dr. M. T. Reetz, Dr. M. Bocola, Dr. J. D. Carballeira, Dr. D. Zha, Dr. A. Vogel
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim/Ruhr (Germany)
Fax: (+49) 208-306-2985
E-mail: reetz@mpi-muelheim.mpg.de

[**] This work was supported by the Fonds der Chemischen Industrie and the Deutsche Forschungsgemeinschaft (Schwerpunkt 1170).

  DOI: 10.1002/ange.200500767

throughput screening.[2] Since structural and mechanistic information is not necessary, the process is fundamentally different from so-called rational design based on site-specific mutagenesis.[3] We have previously applied directed evolution to the creation of enantioselective lipases,[4] epoxide hydrolases,[5] and monooxygenases,[6] and other groups have contributed in this field as well.[1,7] The combination of rational design and random mutagenesis at defined positions in the form of focused libraries[1] has also been employed with the aim of enhancing ligand binding,[8] increasing catalytic activity,[1,9] or influencing enantioselectivity.[4b,10]

A major limitation in exploiting enzymes as catalysts in synthetic organic chemistry is the often encountered limited degree of substrate acceptance.[11] Especially in industrial applications, "methods are desired that have a broad substrate spectrum".[11a] In addition to the use of libraries focusing on a predefined region[1,9] of an enzyme, directed evolution through such standard techniques as the error-prone polymerase chain reaction (epPCR), saturation mutagenesis, and/or DNA shuffling can in principle be employed to solve this problem,[1] and successful cases with regard to nonaccepted substrates have been reported.[1,9,12]

Herein, we describe a method for expanding the scope of substrate acceptance of a given enzyme with the aim of including a wide range of structurally different compounds. Two straightforward steps are required: the design and the generation of relatively small focused libraries of enzyme mutants produced by randomization at several sets of two spatially close amino acid positions around the active site. The choice of two amino acids which are spatially close to one another allows for potential synergistic conformational effects arising from side-chain orientations, an unpredictable phenomenon which cannot be brought about by single-site saturation mutagenesis.[13] The optimal choice of the respective pairs of amino acids is guided by the 3D structure of the wild-type (WT) enzyme with a bound substrate. Geometric inspection allows the definition of sites at which the side chains of the individual amino acids in each pair point toward the binding site of the WT enzyme. On the basis of the secondary structure of an enzyme,[14] it is clear that, for a given pair of spatially as close as possible amino acids in which the side chains point in the direction of the binding site, the following geometric relation pertains: If one member of the pair occurs in the protein sequence at position n, then the second one is found at a sequence of (n+1) in a loop, (n+2) in a β sheet, (n+3) in a 3_{10} helix, and (n+4) in an α helix (Figure 1).[15]

Complete randomization at each pair is then performed in a process that we call the combinatorial active-site saturation test (CAST). Our strategy constitutes a practical compromise between conventional saturation mutagenesis at single sites and simultaneous randomization at multiple sites by using large cassettes. It ensures an appreciable degree of structural diversity and therefore a reasonable chance of finding useful catalysts arising from the $20^2 = 400$ theoretically different mutants in each library, a number which corresponds to $32^2 = 1024$ different codons when using NNK degeneracy (N: any nucleotide; K: G or T).[1] This is in contrast to the limited structural variation of the 20 expected single mutants

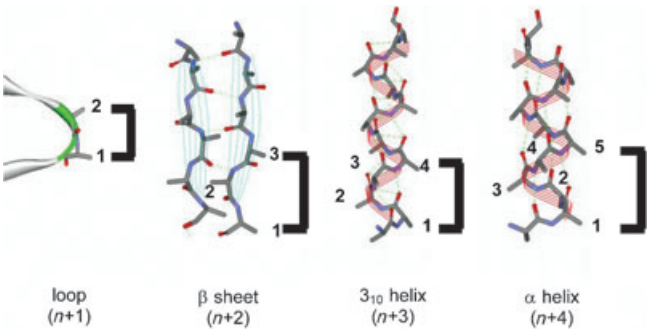


Figure 1. Structural guides in designing libraries of mutant enzymes for CASTing according to the secondary structure of proteins.

(32 codons) in the case of randomization at only single sites,[1,16] a value which may limit the probability of discovering hits. The second important feature of the CAST approach is the limited screening effort. To achieve 95% probability of mutant coverage in the case of randomization at two amino acid positions, about 3000 clones need to be screened (oversampling).[17] In contrast, a focused cassette involving simultaneous randomization at four amino acid sites,[4b] for example, produces theoretically a library of 160000 different mutants, a number requiring $32^4 = 1048576$ codons. This would require a huge screening effort comprising $> 3 \times 10^6$ clones for 95% coverage.[17] Although screening only a small fraction of such a focused library can actually lead to the discovery of positive hits,[4b,10a,10c] the overall process entails uncertainty with respect to the statistical possibility of missing positive mutants.

In a model study, we applied CASTing to expand the substrate acceptance of the lipase from *Pseudomonas aeruginosa*[18] as a catalyst in the hydrolysis of carboxylic acid esters.[4,19] In doing this, one can focus either on the carboxylic acid part or on the alcohol part of the ester (or both). The WT enzyme catalyzes the hydrolysis of triglycerides or fatty acid esters such as palmitic acid *p*-nitrophenyl ester but not of more sterically demanding substrates; not even benzoic acid esters are accepted. Since this constitutes a particularly difficult problem, we focused CASTing on this part of the enzyme. The lipase consists of 285 amino acids, with Ser 82 being the site at which the rate-determining formation of the intermediate oxyanion occurs.[18]

Upon applying the above guidelines (Figure 1) to the crystal structure of the lipase from *Pseudomonas aeruginosa*, solved by Dijkstra et al. in 2000[20] and partially reproduced in Figure 2, the following amino acid pairs were defined: Met 16/Leu 17, Leu 118/Ile 121, Leu 131/Val 135, Leu 159/Leu 162, and Leu 231/Val 232. Five libraries, A–E, respectively, were then created separately by complete saturation, that is, simultaneous radomization at each pair.

About 3000 bacterial colonies from each of the five libraries were harvested on agar plates by a colony picker and deposited in the deep wells of microtiter plates containing Luria–Bertani broth. Following growth of the bacteria and expression of the enzyme, aliquot portions of the lipase-containing supernatants were used to catalyze the hydrolysis of a set of 11 structurally different esters **1**–**11**. In all cases the

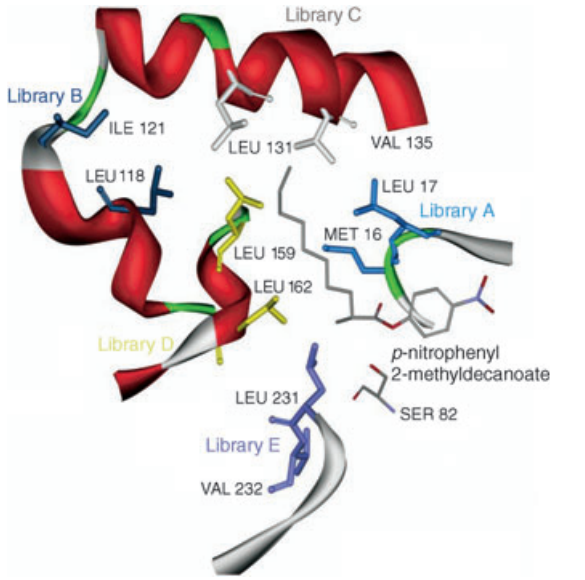
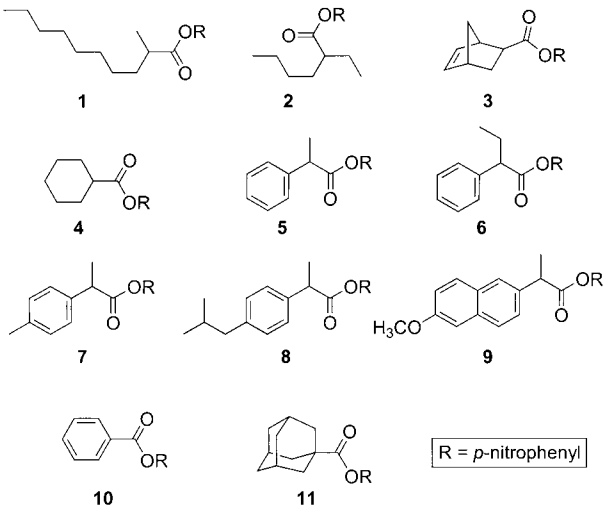


Figure 2. CASTing of the lipase from *Pseudomonas aeruginosa* leading to the construction of five libraries of mutants (A–E) produced by simultaneous randomization at two amino acid sites. (For illustrative purposes, the binding of substrate **1** is shown.)



p-nitrophenyl ester was chosen because the hydrolysis would liberate *p*-nitrophenol which can be monitored easily by UV/Vis spectroscopy.[2,4] A standard UV/Vis plate reader was used to record each reaction rate as measured by the time-resolved appearance of the *p*-nitrophenolate absorption peak at 405 nm (8 min per 96- or 384-well microtiter plate). Parts of the data were checked by GC, and the agreement was found to be excellent. The overall effort in studying the required 165 000 reactions (5 libraries × 3000 clones × 11 substrates) is actually limited, because the process of gene mutagenesis, gene expression, and harvesting of mutants needs to be performed only once.

The results of this multisubstrate screening are remarkable in several respects. Only a part of the data is shown here. Mutants were identified that catalyze the hydrolysis of substrates **1**–**11**; these substrates are either not accepted at all by the WT enzyme or the mutants were found to be considerably more active than the WT enzyme. Figure 3 summarizes the relative reaction rates of WT- and mutant-catalyzed hydrolyses.[21] The eight most important hits (mutants ACA5, A16D8, A2D3, D1A12, D1B10, D1C4, D1F8, and D1E1) originate from libraries A or D, with the latter providing the majority of improved mutants.

The data show that for each substrate, **1**–**11**, one or more mutants were found which lead to a pronounced increase in hydrolysis rate. This means that broad substrate acceptance

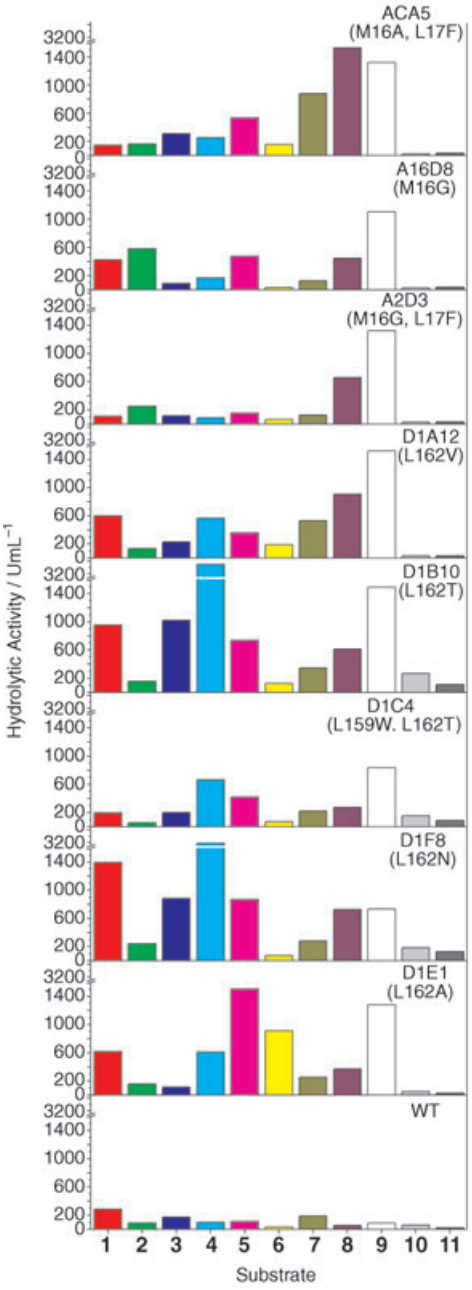


Figure 3. Substrate profiles of lipase variants produced by CASTing. Hydrolytic activities of selected mutants and the WT enzyme were measured photometrically in a continuous assay by following the absorbance at 405 nm for 8 min in a microtiter plate.

has been achieved. Even such bulky substrates as the Naproxen ester **9** are hydrolyzed rapidly, in this case by at least six different mutants. In contrast, the lipase from *Candida rugosa* shows much lower low activity as do other commercial lipases.[22] The adamantyl carboxylic acid ester **11** is a special case. It is not accepted by the WT enzyme, a result that contrasts with the catalytic effect of mutants D1B10, D1C4, and D1F8, which lead to clearly detectable, although low, turnover. In this case further improvement is necessary.

The eight mutants identified by the screening process were sequenced.[23] Again, the results are remarkable. Firstly, in three cases the mutants are characterized by two amino acid substitutions, whereas in five cases single mutations are involved (Table 1). Some of the largest effects are observed in

Table 1: Active mutants of the lipase from *Pseudomonas aeruginosa* created by CASTing.

Mutant	Library[a]	Mutations
ACA5	A	Met16Ala, Leu17Phe
A16D8	A	Met16Gly
A2D3	A	Met16Gly, Leu17Phe
D1A12	D	Leu162Val
D1B10	D	Leu162Thr
D1C4	D	Leu159Trp, Leu162Thr
D1F8	D	Leu162Asn
D1E1	D	Leu162Ala

[a] Library A: simultaneous randomization at Met16 and Leu17; Library D: simultaneous randomization at Leu159 and Leu162; see Figure 2.

the double mutants, which would not have been discovered by conventional saturation mutagenesis. Moreover, the simple expectation that the introduction of a sterically smaller amino acid creates more space for a bulky substrate to fit properly is not fulfilled in all cases, a fact making a detailed interpretation of such subtle effects difficult at this time. The mutation Leu17Phe, which occurs in two different mutants, is an example. It is also interesting to note that amino acid 162 is a particularly sensitive hot spot (Table 1), in line with the results of earlier directed-evolution studies.[4] We conclude that CASTing is a practical first step in a directed-evolution project with the aim of expanding the range of substrate acceptance. Further improvements can be envisioned in an evolutionary sense by recasting, by combining the mutations of two hits, or by applying the usual mutagenesis methods, such as epPCR or DNA shuffling, which cover the whole gene.[1]

Finally, we note that several of the substrates are chiral, which means that kinetic resolution of the racemates is relevant. Although we have not (yet) performed systematic studies regarding this point, it appears that CASTing can indeed be used to evolve enantioselectivity as well, which is not surprising. For example, mutant D1A12 catalyzes the hydrolytic kinetic resolution of ester **6** with a selectivity factor of $E = 20 \pm 5$ and mutant ACA5 performs the same reaction with $E = 25 \pm 5$, both in favor of the *S* substrate. The WT enzyme does not accept this substrate to any appreciable extent, which means that enantioselectivity could not be measured reliably.

In summary, a practical method (CAST) for solving the long-standing problem of wide substrate acceptance of enzymes has been developed. CASTing combines the features of rational design and combinatorial amino acid randomization at focused sites. It is different from previous focused libraries[1,9] in that a complete set of pairs of amino acids around the bound substrate is considered. This enables the systematic creation of relatively small libraries of mutants, the screening of which by currently available assays leads to positive hits and consequently to the identification of the critical regions around the active site that are crucial for substrate acceptance. We believe that the method can also be considered when attempting to improve the catalytic profile of enzymes in general, for example, in the enhancement of activity and perhaps stability, as well as regio- and stereoselectivity. It is also relevant when attempting to convert a given enzyme into a different type of catalyst.[12] Therefore, CASTing may well constitute a useful alternative to the usual epPCR as the initial step in directed-evolution studies.

Received: March 2, 2005
Revised: April 15, 2005
Published online: June 1, 2005

Keywords: directed evolution · enzymes · lipases · protein engineering · substrate acceptance

[1] a) K. A. Powell, S. W. Ramer, S. B. del Cardayré, W. P. C. Stemmer, M. B. Tobin, P. F. Longchamp, G. W. Huisman, *Angew. Chem.* **2001**, *113*, 4068–4080; *Angew. Chem. Int. Ed.* **2001**, *40*, 3948–3959; b) F. H. Arnold, G. Georgiou, *Directed Enzyme Evolution: Screening and Selection Methods*, *Vol. 230*, Humana Press, Totowa, NJ, **2003**; c) S. Brakmann, K. Johnsson, *Directed Molecular Evolution of Proteins (or How to Improve Enzymes for Biocatalysis)*, Wiley-VCH, Weinheim, **2002**; d) A. Svendsen, *Enzyme Functionality—Design, Engineering, and Screening*, Marcel Dekker, New York, **2004**; e) S. Brakmann, A. Schwienhorst, *Evolutionary Methods in Biotechnology (Clever Tricks for Directed Evolution)*, Wiley-VCH, Weinheim, **2004**.

[2] a) M. T. Reetz, *Angew. Chem.* **2001**, *113*, 292–320; *Angew. Chem. Int. Ed.* **2001**, *40*, 284–310; b) M. T. Reetz in reference [1d], pp. 559–598; c) D. Wahler, J.-L. Reymond, *Curr. Opin. Biotechnol.* **2001**, *12*, 535–544.

[3] Review: F. Cedrone, A. Menez, E. Quemeneur, *Curr. Opin. Struct. Biol.* **2000**, *10*, 405–410.

[4] 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, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem.* **2001**, *113*, 3701–3703; *Angew. Chem. Int. Ed.* **2001**, *40*, 3589–3591; c) M. T. Reetz, *Tetrahedron* **2002**, *58*, 6595–6602; d) M. Bocola, N. Otte, K.-E. Jaeger, M. T. Reetz, W. Thiel, *ChemBioChem* **2004**, *5*, 214–223; e) M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5716–5722.

[5] a) M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand, A. Cronin, Y. Genzel, A. Archelas, R. Furstoss, *Org. Lett.* **2004**, *6*, 177–180; b) F. Cedrone, S. Niel, S. Roca, T. Bhatnagar, N. Ait-Abdelkader, C. Torre, H. Krumm, A. Maichele, M. T. Reetz, J. C. Baratti, *Biocatal. Biotransform.* **2003**, *21*, 357–364.

[6] a) M. T. Reetz, B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier, M. M. Kayser, *Angew. Chem.* **2004**, *116*, 4167 – 4170; *Angew. Chem. Int. Ed.* **2004**, *43*, 4075 – 4078; b) M. T. Reetz, F. Daligault, B. Brunner, H. Hinrichs, A. Deege, *Angew. Chem.* **2004**, *116*, 4170 – 4173; *Angew. Chem. Int. Ed.* **2004**, *43*, 4078 – 4081.

[7] Reviews: references [4c,e]; "Changing the Enantioselectivity of Enzymes by Directed Evolution": M. T. Reetz, *Methods Enzymol.* **2004**, *388*, 238 – 256; N. J. Turner, *Trends Biotechnol.* **2003**, *21*, 474 – 478.

[8] a) S. Schlehuber, A. Skerra, *Biophys. Chem.* **2002**, *96*, 213 – 228; b) A. Knappik, L. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellnhofer, A. Hoess, J. Wölle, A. Plückthun, B. Virnekäs, *J. Mol. Biol.* **2000**, *296*, 57 – 86.

[9] Examples: a) L. B. Evnin, J. R. Vásquez, C. S. Craik, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6659 – 6663; b) J. Yang, Y. Koga, H. Nakano, T. Yamane, *Protein Eng.* **2002**, *15*, 147 – 152; c) E. Whittle, J. Shanklin, *J. Biol. Chem.* **2001**, *276*, 21500 – 21505; d) A. R. Schmitzer, F. Lépine, J. N. Pelletier, *Protein Eng. Des. Sel.* **2004**, *17*, 809 – 819; e) C. M. Hill, W.-S. Li, J. B. Thoden, H. M. Holden, F. M. Raushel, *J. Am. Chem. Soc.* **2003**, *125*, 8990 – 8991; f) F. Brühlmann, W. Chen, *Biotechnol. Bioeng.* **1999**, *63*, 544 – 551; g) E. M. Gabor, D. B. Janssen, *Protein Eng. Des. Sel.* **2004**, *17*, 571 – 579; h) J. Yang, Y. Koga, H. Nakano, T. Yamane, *Protein Eng.* **2002**, *15*, 147 – 152; i) S. V. Taylor, P. Kast, D. Hilvert, *Angew. Chem.* **2001**, *113*, 3408 – 3436; *Angew. Chem. Int. Ed.* **2001**, *40*, 3310 – 3335.

[10] a) M. Chen-Goodspeed, M. A. Sogorb, F. Wu, S.-B. Hong, F. M. Raushel, *Biochemistry* **2001**, *40*, 1325 – 1331; b) "Directed Evolution of Enantioselective Enzymes as Catalysts in the Production of Chiral Pharmaceuticals": M. T. Reetz in *Pharmacochemistry Library*, *Vol. 32 (Trends in Drug Research III)* (Ed.: H. van der Goot), Elsevier, Amsterdam, **2002**, pp. 27 – 37; c) Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585 – 592; d) S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45 – 54.

[11] a) M. Breuer, K. Ditrich, T. Habicher, B. Hauer, M. Keßeler, R. Stürmer, T. Zelinski, *Angew. Chem.* **2004**, *116*, 806 – 843; *Angew. Chem. Int. Ed.* **2004**, *43*, 788 – 824; b) K. Drauz, H. Waldmann, *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook*, *Vol. I–III*, 2nd ed., Wiley-VCH, Weinheim, **2002**; c) K. Faber, *Biotransformations in Organic Chemistry*, 4th ed., Springer, Berlin, **2000**; d) A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258 – 268.

[12] Discussions of enzyme promiscuity, plasticity, and so-called moonlighting: a) A. Aharoni, L. Gaidukov, O. Khersonsky, S. McQ Gould, C. Roodveldt, D. S. Tawfik, *Nat. Genet.* **2005**, *37*, 73 – 76; b) S. Raillard, A. Krebber, Y. Chen, J. E. Ness, E. Bermudez, R. Trinidad, R. Fullem, C. Davis, M. Welch, J. Seffernick, L. P. Wackett, W. P. C. Stemmer, J. Minshull, *Chem. Biol.* **2001**, *8*, 891 – 898; c) P. J. O'Brien, D. Herschlag, *Chem. Biol.* **1999**, *6*, R91 – R105; d) S. D. Copley, *Curr. Opin. Chem. Biol.* **2003**, *7*, 265 – 272; e) U. T. Bornscheuer, R. J. Kazlauskas, *Angew. Chem.* **2004**, *116*, 6156 – 6165; *Angew. Chem. Int. Ed.* ***2004***, *43*, 6032 – 6040.

[13] The randomization at two amino acids which are not spatially close to each other yet have side chains pointing toward the active site (for example, in two different libraries) may of course also be considered.

[14] a) A. Fersht, *Structure and Mechanism in Protein Science*, W. H. Freeman, New York, **1999**; b) R. A. Copeland, *Enzymes*, Wiley-VCH, Weinheim, **2000**.

[15] These guidelines are derived from ideal secondary structural units. If in a given case the latter are not ideal, as, for example, in a kinked helix, then additional positions need to be considered. In the case of a kinked helix, for example, it may be a sequentially neighboring amino acid (n+1). It is mandatory to include such structural aspects when designing combinatorial active-site saturation test libraries, that is, the shape of the binding pocket serves as a guide.

[16] Examples: a) G. DeSantis, K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz, M. J. Burk, *J. Am. Chem. Soc.* **2003**, *125*, 11476 – 11477; b) S. A. Funke, A. Eipper, M. T. Reetz, N. Otte, W. Thiel, G. van Pouderoyen, B. W. Dijkstra, K.-E. Jaeger, T. Eggert, *Biocatal. Biotransform.* **2003**, *21*, 67 – 73; c) R. Georgescu, G. Bandara, L. Sun in *Directed Evolution Library Creation* (Eds.: F. H. Arnold, G. Georgiou), Humana Press, Totowa, NJ, **2003**, pp. 75 – 89.

[17] For this calculation we used the traditional statistical method; see, for example: M. S. Warren, S. J. Benkovic, *Protein Eng.* **1997**, *10*, 63 – 68, and references therein.

[18] a) K.-E. Jaeger, B. Schneidinger, K. Liebeton, D. Haas, M. T. Reetz, S. Philippou, G. Gerritse, S. Ransac, B. W. Dijkstra in *Molecular Biology and Biotechnological Application* (Eds.: T. Nakazawa, K. Furukawa, D. Haas, S. Silver), ASM Press, Washington, **1996**, pp. 319 – 339; b) K.-E. Jaeger, K. Liebeton, A. Zonta, K. Schimossek, M. T. Reetz, *Appl. Microbiol. Biotechnol.* **1996**, *46*, 99 – 105.

[19] We thank K.-E. Jaeger for providing the lipase-expression system.[18]

[20] M. Nardini, D. A. Lang, K. Liebeton, K.-E. Jaeger, B. W. Dijkstra, *J. Biol. Chem.* **2000**, *275*, 31219 – 31225.

[21] Because of the instability of the lipase from *Pseudomonas aeruginosa*,[18] it was not possible to purify mutants and to perform detailed kinetic studies. A previous kinetic study was performed by using the supernatants and assuming similar amounts of enzyme expressed and secreted (M. T. Reetz, S. Wilensek, unpublished results; see also: M. T. Reetz, *Tetrahedron* **2002**, *58*, 6595 – 6602). To lend support to this assumption, we measured the lipase activity towards the "natural" compound palmitic acid *p*-nitrophenyl ester as a reference substrate. All of the mutants of the present study showed similar activities towards this reference substrate but slightly lower activities relative to the WT enzyme. Therefore, the low activities of the WT enzyme towards the different substrates in Figure 3 are not because of a lower expression rate. This was confirmed by sodium dodecylsulfate PAGE analysis. This means that the profiles of the mutants reflect inherent activity enhancements.

[22] A comparative study encompassing various substrates and lipases is underway, as are CASTing experiments on commercially available lipases that have higher thermal stability than the lipase from *Pseudomonas aeruginosa*.

[23] None of the variants contain additional mutations because of the PCR.