CHEMBIOCHEM

DOI: 10.1002/cbic.200800371

Constructing and Analyzing the Fitness Landscape of an Experimental Evolutionary Process

Manfred T. Reetz* and Joaquin Sanchis[a]

Iterative saturation mutagenesis (ISM) is a promising approach to more efficient directed evolution, especially for enhancing the enantioselectivity and/or thermostability of enzymes. This was demonstrated previously for an epoxide hydrolase (EH), after five sets of mutations led to a stepwise increase in enantioselectivity. This study utilizes these results to illuminate the nature of ISM, and identify the reasons for its operational efficacy. By applying a deconvolution strategy to the five sets of mutations and measuring the enantioselectivity factors (E) of the EH variants, $\Delta\Delta G^{\ddagger}$ values become accessible. With these values, the construction of the complete fitness-pathway landscape is possible. The free energy profiles of the 5! = 120 evolutionary pathways leading from the wild-type to the best mutant show that 55 trajectories are energetically favored, one of which is the originally observed route. This particular pathway was analyzed in terms of epistatic effects operating between the sets of mutations at all evolutionary stages. The degree of synergism increases as the stepwise evolutionary process proceeds. When encountering a local minimum in a disfavored pathway, that is, in the case of a dead end, choosing another set of mutations at a previous stage puts the evolutionary process back on an energetically favored trajectory. The type of analysis presented here might be useful when evaluating other mutagenesis methods and strategies in directed evolution.

Introduction

Directed evolution is a useful way to engineer the properties of proteins, including the stability, activity, substrate scope, and enantioselectivity of enzymes as catalysts in synthetic organic chemistry and biotechnology.[1] It involves the combination of gene mutagenesis, expression and screening (or selection). The most common mutagenesis methods are error-prone polymerase chain reaction (epPCR),[2] saturation mutagenesis,[3] and DNA shuffling,[4] or variations and combinations thereof.[1] An important focus of current research is methodology development to make this type of protein engineering more efficient through the generation of higher-quality libraries.[5] Our recent contribution to this endeavor is iterative saturation mutagenesis (ISM) in its two embodiments, namely the combinatorial active-site saturation test (CAST) for influencing substrate acceptance and/or enantioselectivity,[6,7] and the B-factor iterative test (B-FIT) for increasing protein thermostability.[8] CASTing involves two steps. On the basis of crystallographic data or homology models of an enzyme to be engineered, sites A, B, C, D, etc. around the complete binding pocket hosting the substrate are first identified. Following this analysis, all sites are randomized through application of saturation mutagenesis. A set of focused libraries of mutants is created and screened for substrate scope (activity) and/or enantioselectivity. This initial step constitutes a systematization of earlier uses of focused libraries.[9] Thereafter, the mutant gene encoding a given improved enzyme variant (hit) is used as a template for performing saturation mutagenesis at the other sites, and the process is repeated as often as needed. A schematic illustration of ISM is shown in Figure 1, which involves four sites A, B, C and D, each site being "visited" only once in a given pathway.

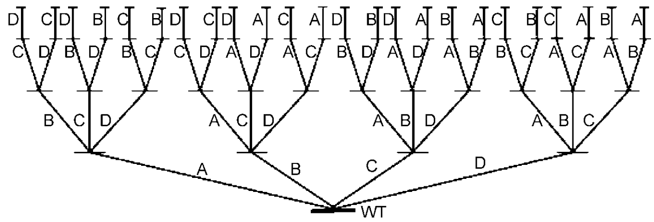


Figure 1. General scheme illustrating ISM as in iterative CASTing for the case of four sites, A, B, C, and D, at which saturation mutagenesis is performed. (Extension by visiting a given site more than once in an upward pathway is possible.)[7]

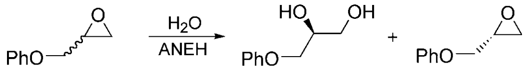
This strategy is quite different from the approach in which mutations originating from two or more libraries are simply combined,[10] which does not allow for new and potentially highly beneficial amino acid substitutions to occur. In contrast, ISM maximizes the probability of positive epistatic effects in terms of additivity and/or synergism between sets of mutations at each branching point (Figure 1). Moreover, in the case of a local minimum, the ISM strategy allows for a different pathway to be chosen when a randomization library fails to generate any improved variants.

Iterative CASTing was first applied in a study regarding the directed evolution of the epoxide hydrolase from *Aspergillus niger* (ANEH) as a biocatalyst for the enantioselective hydrolytic kinetic resolution of glycidyl phenyl ether (GPE, *rac*-**1**).[7] The

[a] *Prof. Dr. M. T. Reetz, Dr. J. Sanchis*
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

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

wild-type (WT) leads to a selectivity factor of only $E=4.6$ in slight favor of (*S*)-**2**.



rac-**1** (*S*)-**2** (*R*)-**1**

CAST analysis based on the crystal structure of the WT ANEH[11] suggested six possible sites for saturation mutagenesis, namely A (amino acid positions 193/195/196), B (215/217/219), C (329/330), D (349/350), E (317/318) and F (244/245/249; Figure 2). The essential results of the saturation mutagenesis experiments from this initial study are summarized in Figure 3.[7] The chosen pathway B→C→D→F→E leads to a highly enantioselective variant LW202, which has a selectivity factor of $E=115$ (site A was not considered as part of the chosen pathway in this study). The final enzyme variant LW202 is characterized by five *sets* of mutations that were accumulated in a stepwise fashion during the evolutionary process, namely Leu215Phe/Ala217Asn/Arg219Ser at B, Met329Pro/Leu330Tyr at C, Cys350Val at D, Leu249Tyr at F, and Thr317Trp/Thr318Val at E, which is a total of nine amino acid substitutions. In all saturation experiments, NNK codon degeneracy was chosen (N: adenine/cytosine/guanine/thymine; K: guanine/thymine), which means that all 20 proteinogenic amino acids were used as building blocks. However, no attempt was made to ensure full coverage of the relevant protein sequence space by appropriate oversampling. In fact, only relatively small mutant libraries were considered. Moreover, since variant LW202 is already highly enantioselective, the initial study did not include the exploration of other pathways. It is possible or even likely that a more systematic search based on exploration of other pathways would have provided a number of additional hits characterized by different sequences, but with similarly high or even higher enantioselectivity. The overall effort required the screening of 20000 clones, which is about the same number evaluated in an earlier study based on epPCR at various mutation rates, leading to a selectivity factor of only $E=11$.[12] Thus, compared to the use of epPCR, knowledge-guided targeted mutagenesis appears to be a superior methodology, generating "smarter" libraries that are characterized by a dramatically higher *frequency* and *quality* of hits.[7]

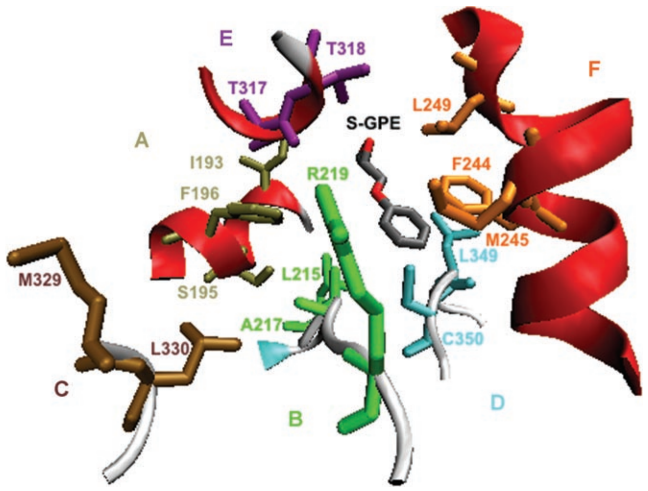


Figure 2. Proposed CAST sites A, B, C, D, E and F in the enzyme ANEH harboring (*S*)-**1** (*S*-GPE).

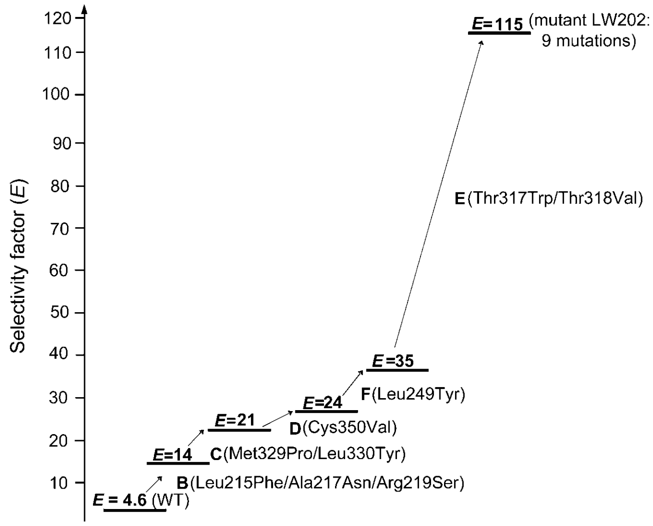


Figure 3. Iterative CASTing pathway B→C→D→F→E that leads to the highly enantioselective mutant LW202 as a catalyst in the kinetic resolution of *rac*-**1**.[7]

In this study we have devised a general strategy for evaluating mutagenesis methods in laboratory evolution, and applied it specifically in order to determine the reasons for the apparent efficacy of iterative CASTing. Two questions are addressed: 1) When considering solely the experimentally obtained *sets* of mutations (Figure 3), is the transversed pathway B→C→D→F→E the only one that leads to variant LW202, or do other pathways based on permutational combinations of these specific sets of mutations likewise provide that particular variant? 2) What is the nature of the epistatic interactions that occur between the five sets of mutations along the trajectory B→C→D→F→E, that is, to what extent do additive, synergistic and/or antagonistic interactions operate at each step of the evolutionary process?

Results and Discussion

Construction of all pathways in the fitness landscape

We first had to construct the 30 mutants that are relevant when considering all combinations of the five sets of mutants obtained successively in the experimentally transversed sequence B→C→D→F→E. Following the evaluation of these 30 mutants as catalysts in the model reaction *rac*-**1**→(*S*)-**2**+(*R*)-**1**, the selectivity factors *E* and respective free energy values $\Delta\Delta G^{\neq}$ allow the construction of the $5!=120$ possible direct trajectories leading from the WT-ANEH to enzyme variant LW202 (Figure 4).

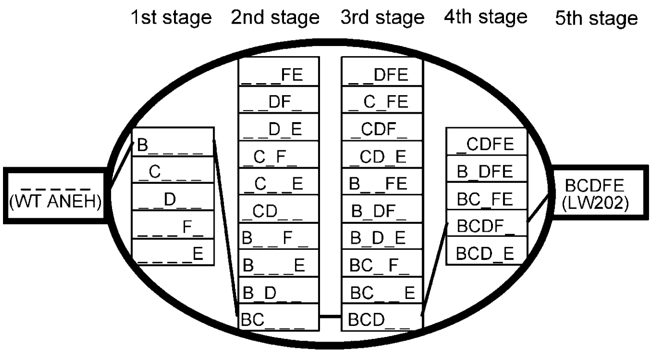


Figure 4. The 30 possible mutants as intermediate stages between WT-ANEH and enzyme variant LW202 based on the use of five sets of mutations. The connecting lines indicate the original pathway B→C→D→F→E, which is one of 120 possible trajectories (see text).

By using site-specific mutagenesis, 26 new mutants were prepared and tested in the model reaction. Mutants corresponding to B, BC, BCD and BCDF were already available from the original study. The fitness landscape of this system constitutes a six-dimensional surface (mutation sets B, C, D, E and F are independent vectors, and $\Delta\Delta G^{\neq}$ is the dependent variable), which is difficult to depict graphically. We present data of this kind in the form of a "fitness-pathway landscape" as shown in Figure 5. All 120 possible pathways linking the WT

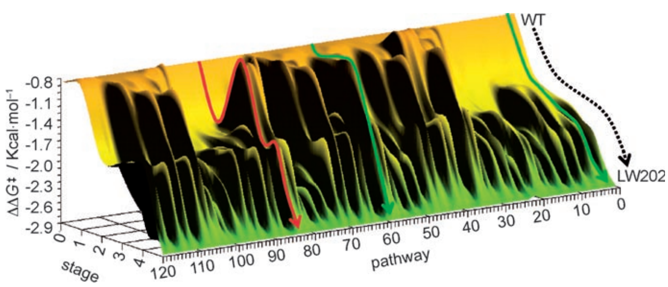


Figure 5. Energy profile of the two types of pathways leading from the WT to the mutant LW202: Energetically favored (green) as in the original B→C→D→F→E (pathway 2) or D→C→F→E→B (pathway 60) and disfavored (red) as in E→C→F→D→B (pathway 84).

(top) with the target variant LW202 (bottom), as specified by the experimental results, are stacked. Upon mapping the free energy profiles of these 120 mutational pathways, we discovered that two different types of trajectories exist. The first type shows a continuous decrease in the free energy, as in the original pathway B→C→D→F→E or in the case of D→C→F→E→B (Figure 5, green pathways). In sharp contrast, the second type of pathway is characterized by the presence of a turning point followed by a peak along the trajectory that is higher in energy than at the previous evolutionary stage, indicating a local minimum. A typical example is the case of E→C→F→D→B (Figure 5, red pathway).

Upon exploring the complete fitness-pathway landscape, some remarkable features were unveiled (Figure 6). Most importantly, a total of 55 different pathways were found to be energetically favorable, which corresponds to about 46% of all possibilities in going from the WT to LW202 by using the five sets of mutations. Thus, our data show that laboratory evolution based on iterative forms of saturation mutagenesis can follow many energetically favored pathways to improved enzymes. Direct comparison with other systems that reach different conclusions, such as natural evolution, cannot be made due to the difference in experimental platforms.[13]

We also note that when a local minimum in a disfavored trajectory (trajectory 84 or other red pathways, Figure 6) is encountered and no energetically favored forward step for further catalyst improvement can be found, the experimenter can escape from such a "dead end". This is possible by returning to the previous stage and choosing another combination from the remaining sets of mutations. For all 65 disfavored pathways, we did not find a single trajectory in which this strategy would fail to put the evolutionary process back on a positive track (Figure 6).

We also considered the first derivative of the fitness-pathway landscape. The result of this analysis clearly supports our previous conclusion regarding the relatively high number of energetically favored trajectories (Figure 7). Much of the landscape surface is below the XY planes (green), which means that these stages are energetically favored.

Analysis of cumulative mutational effects

Figures 6 and 7 do not reveal the details of the epistasis in any of the pathways. Therefore, in order to understand the operational efficiency of the iterative process, complete analysis of the interactions among the different sets of mutations is necessary. We focus here on the original pathway B→C→D→F→E as an example of an energetically favored trajectory, and pathway 84 (Figure 5) as an example of a disfavored trajectory.

The experimental data at hand allow the calculation of the free energy of interaction ($\Delta G^{\neq}_{ij}$) between any two sets of mutations i and j [Equation (1)], in analogy to the case regarding the interaction of two point mutations.[14]

$$\Delta G^{\neq}_{ij} = \Delta\Delta G^{\neq}_{\text{exp}} - (\Delta\Delta G^{\neq}_{i} + \Delta\Delta G^{\neq}_{j}) \quad (1)$$

where $\Delta\Delta G^{\neq}_{\text{exp}}$ is the difference in the activation energy between both enantiomers experimentally obtained for the binary combination, and $\Delta\Delta G^{\neq}_{i}$ and $\Delta\Delta G^{\neq}_{j}$ are the experimental energies obtained for each set of mutants separately. The values of $\Delta G^{\neq}_{ij}$ are either a measure of synergistic effects ($\Delta G^{\neq}_{ij} < 0$), of additive effects ($\Delta G^{\neq}_{ij} = 0$, no interaction), of partially additive effects ($\Delta G^{\neq}_{ij} > 0$ and $((|\Delta\Delta G^{\neq}_{i}|$ and $|\Delta\Delta G^{\neq}_{j}|) < |\Delta\Delta G^{\neq}_{\text{exp}}|))$, or they denote antagonistic effects ($\Delta G^{\neq}_{ij} > 0$ and $((|\Delta\Delta G^{\neq}_{i}|$ or $|\Delta\Delta G^{\neq}_{j}|) > |\Delta\Delta G^{\neq}_{\text{exp}}|))$ as schematized in Figure 8. All of the thermodynamic cycles were calculated from Equation (1) for the binary combination of sets of mutations (i+j), and extensively for the ternary (i+j+k), the quaternary (i+j+k+l) and the quinary (i+j+k+l+m) combinations. Specifically, set i (in blue) contributes $\Delta\Delta G^{\neq}_{i}$, set j (orange) contributes $\Delta\Delta G^{\neq}_{j}$, and the expected additive increment column (striped) reflects the sum ($\Delta\Delta G^{\neq}_{i} + \Delta\Delta G^{\neq}_{j}$). The difference be-

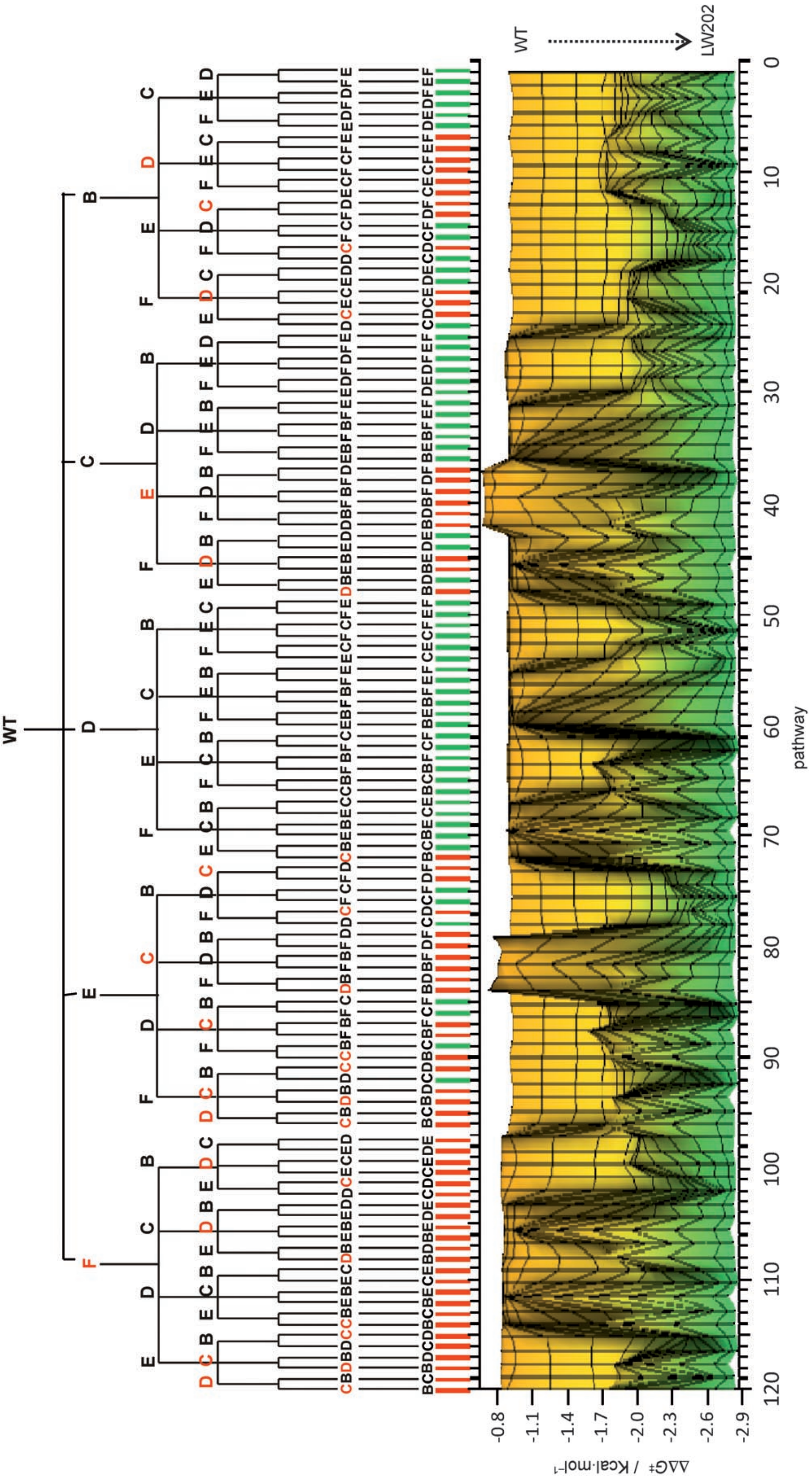


Figure 6. Energy profiles in the frontal view of Figure 5 of all 120 trajectories leading from the WT-ANEH to enzyme variant LW202. Green notations indicate energetically favored pathways, whereas red notations stand for disfavored trajectories.

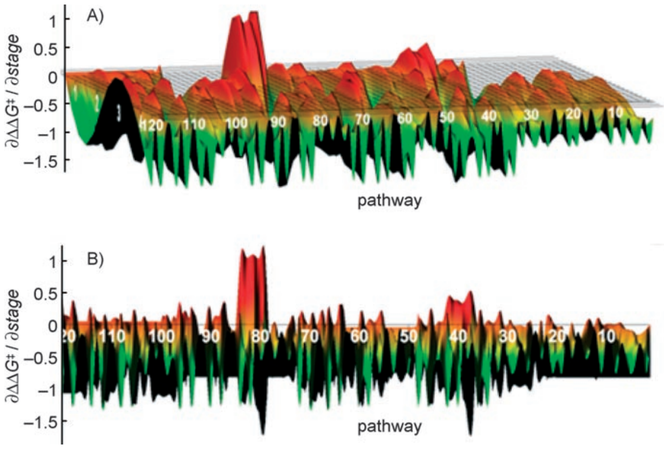


Figure 7. First derivative of $\Delta\Delta G^{\neq}$ at every stage of each pathway. Red areas (above the *XY* plane) indicate positive slopes, that is, energetically disfavored pathways. Yellow/orange and green areas (below *XY* plane) indicate negative slopes and define steps involved in energetically favored stages. A) Side view of surface. B) Frontal view of surface (plane *XZ*). $(x,y,z) = (\text{pathway}, (stage_1+stage_2)/2, (\Delta\Delta G^{\neq}_1-\Delta\Delta G^{\neq}_2)/(stage_1-stage_2))$.

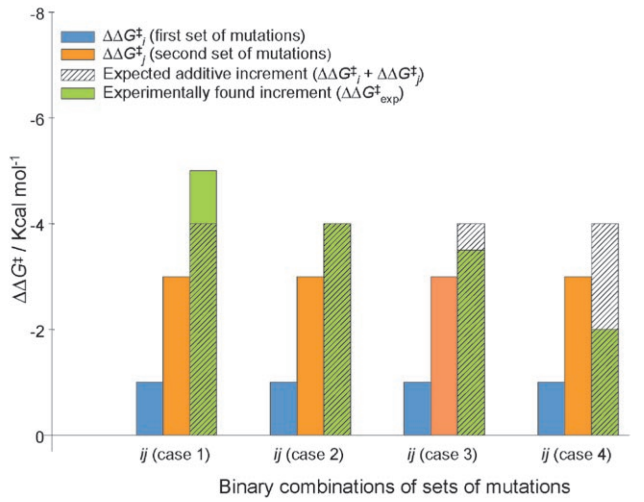


Figure 8. Explanation of different cases that can be found in the interaction between two sets of mutations (*i* and *j*) when the thermodynamic cycle described in Equation (1) is calculated. Case 1) Experimentally found increment is bigger than the expected additive increment (synergistic effect). Case 2) Experimentally found increment is equal to the expected additive increment (additive effect, no interaction). Case 3) Experimentally found increment is smaller than the expected additive increment but bigger than the energy of the single sets (partially additive effect). Case 4) Experimentally found increment is smaller than either one of the sets (antagonistic effect).

tween the experimentally found increment ($\Delta\Delta G^{\neq}_{exp}$, in green) and the expected additive increment corresponds to the energy of interaction ($\Delta G^{\neq}_{ij}$). The four different situations according to the value of $\Delta G^{\neq}_{ij}$ are illustrated in Figure 8.

In order to visualize the epistatic effects of the favoured pathway B→C→D→F→E at each stage, we processed the data as shown in Figure 9. The final epistatic result at each stage is not characterized by additivity, but by synergy, and the magnitude of the synergistic effects increases more or less continuously in the process of reaching LW202. Any antagonistic effects, as in the case of the fourth set of mutations alone,

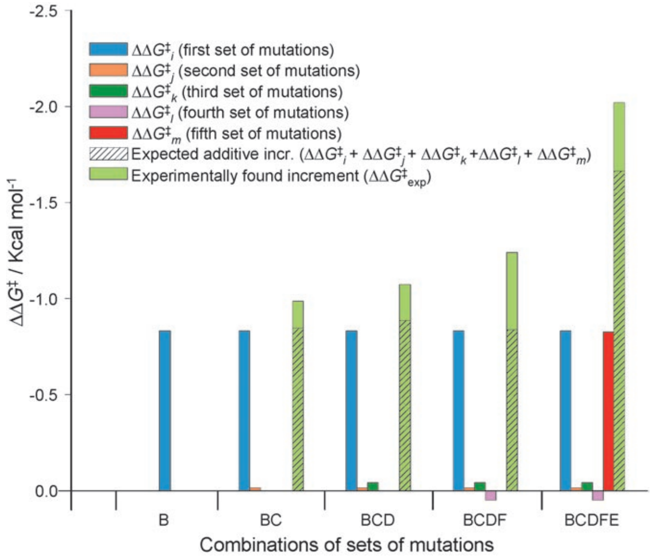


Figure 9. Thermodynamic cycle [Eq. (1)] regarding the interaction of the sets of mutations involved at every stage along the energetically favored pathway B→C→D→F→E.

are purged by new combinations. These effects also become visible when considering the selectivity factors *E*. For example, in the last experimental stage of the original study involving step E (Figure 3), the *E* value increases dramatically from 35 to 115, but when the set of mutations (Thr317Trp/Thr318Val) introduced there is tested alone as a mutant, $E=16$ (Table S2 in the Supporting Information). Similar effects were found in the case of other energetically favored pathways.

None of the mutants, including those arising from the various combinations of the respective sets, leads to enantioselectivities higher than that of the originally evolved variant LW202 (Table S2). In agreement with the previous findings (see above), this observation sheds light on the nature of the iterative process: none of the sets of mutations that were generated during the five-step evolutionary process defined by B→C→D→F→E is superfluous. This stands in contrast to the results of a different study based on epPCR and DNA shuffling, in which only two of the accumulated six mutations were necessary for high enantioselectivity.[15]

We gained further insight into the favored pathway B→C→D→F→E by adopting the so-called "inverse thinking" approach regarding mutational effects occurring in proteins. This type of analysis was first proposed by Mildvan.[16] Accordingly, we compared the free energy of stabilization resulting from the introduction of every set of mutations into the WT and the free energy of destabilization resulting from the deletion of this set in a specific mutant. This analysis was recursively performed at each stage of the original pathway B→C→D→F→E. On average, the weight of contribution to the total fitness of every set increases in magnitude as the evolutionary process passes to higher levels, as measured by enhanced enantioselectivity (Figure 10). Clearly, the degree of cooperativity (synergism) among the accumulated sets of mutations increases with each evolutionary step, in line with our previous conclusion.

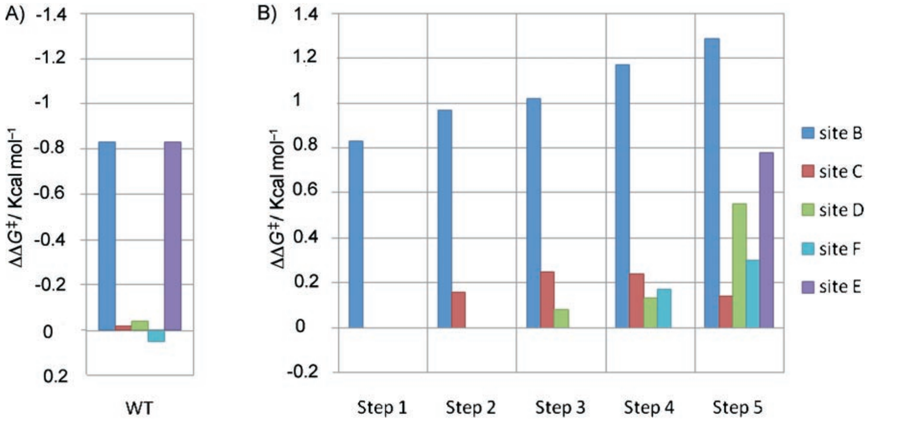


Figure 10. Analysis by the Mildvan approach.[16] A) Stabilization $\Delta\Delta G^{\neq}$ of the mutant relative to the WT upon the introduction of a set of mutations. B) Destabilization $\Delta\Delta G^{\neq}$ observed upon replacement of each set of mutations by the WT amino acids at every step in the experimentally transversed pathway B→C→D→F→E.

ly outperforming the use of traditional epPCR employed in an earlier study.[12] The selectivity factor *E* in the model kinetic resolution of a chiral epoxide increased from 4.6 (WT) to 115 (best variant LW202) as a result of five cycles of saturation mutagenesis at the sites around the binding pocket of the enzyme. This study has led to the identification of the factors that contribute to the efficacy of iterative CASTing. A deconvolution strategy was applied so that the five sets of mutations originally generated in the stepwise evolutionary process toward highly enhanced enantioselectivity were reconstructed in all theoretically possible permutations. Thus, all 120 pathways from the WT to the best mutant LW202 were mapped on the basis of the free energy values $\Delta\Delta G^{\neq}$ obtained from the respective selectivity factors *E*. Two types of trajectories exist in ISM, namely energetically favored and disfavored pathways. About 46% of the trajectories proved to be energetically favored; this is a high score. In the case of a disfavored pathway, which involves at least one local minimum, it is possible to escape from such a dead end by returning to the previous stage and choosing a different set of mutations, thereby putting the evolutionary process back on a positive track.

The information accessible by the deconvolution approach also sheds light on another facet of the nature of ISM, namely the way accumulated sets of mutations interact with one another in terms of epistasis. It is possible to analyze epistatic effects between the five sets of mutations in any of the 120 pathways, and we have demonstrated this for a favored and a disfavored trajectory. In the case of the original favored pathway B→C→D→F→E, synergistic effects dominated, whereas negative epistatic influences were found to play a minor role. Rare antagonistic effects are completely purged by the appearance of new combinations, so local minima are avoided. Moreover, synergism increases as the evolutionary process proceeds. In the case of a typical energetically disfavored pathway, we uncovered the reason for the existence of local minima, namely the dominance of antagonistic effects. On a molecular level, the physical reasons for positive or negative epistatic effects in the case of the best mutant LW202 have yet to be identified.

Statistically, the improved mutants provided by ISM, as in iterative CASTing, are not theoretically accessible by repeating cycles of epPCR, irrespective of the mutation rate. We have demonstrated this experimentally in two comparative directed evolution studies involving ANEH.[7,12] Random mutagenesis based on epPCR,[2] which is still the most commonly employed technique in directed evolution,[1] is highly unlikely to provide mutants with two or three amino acid substitutions in a defined locus in the enzyme. Yet the accumulation of several

Using our first approach, we then analyzed one of the energetically disfavored pathways in Figure 6, namely number 84. Only the essential features are highlighted here as summarized in Figure 11. Negative epistatic effects operate at several stages. The interaction between sites E and C constitutes a particularly strong antagonistic effect that causes a prominent local minimum along trajectory 84 (Figure 4). The second, less-prominent local minimum occurs in the combination ECFD, as full additivity is not quite achieved. In the final combination, ECFDB, a strong synergistic effect dominates as expected.

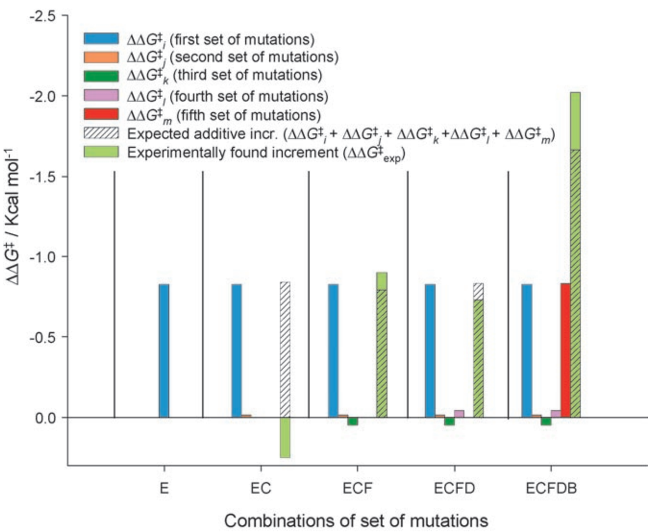


Figure 11. Thermodynamic cycle [Eq. (1)] regarding the interaction of the sets of mutations involved at every stage along the energetically disfavored pathway number 84 (E→C→F→D→B).

Conclusions

We have previously demonstrated that ISM in the embodiment of iterative CASTing is an efficient method for the directed evolution of an enantioselective epoxide hydrolase (ANEH),[7] clearmutations in a spatially focused manner effectively influences enantioselectivity or substrate acceptance (rate) of enzymes. This is due to the increased probability of synergistic effects when performing saturation mutagenesis iteratively.

CASTing excludes the possibility of discovering distal mutations when attempting to influence catalytic properties such as substrate acceptance and/or enantioselectivity. Remote effects are certainly an intriguing phenomenon in such endeavors.[9c,17] However, from a practical point of view, such a restriction does not necessarily have negative consequences. The opposite actually pertains, because all experience in this study and in earlier reports[7,9] shows that knowledge-guided targeted randomization is beneficial when attempting to maximize the quality of enzyme libraries. Nevertheless, epPCR can be expected to continue to play an important role in directed evolution. For example, in cases in which structural information is lacking, epPCR might well be a logical choice. The "hot spots" identified by random mutagenesis can then be used as sites for possible ISM exploration if needed.

We point out that the results of this study call for further research in at least two areas. The scheme illustrating ISM (Figure 1) needs to be tested more systematically in a different model system by exploring at least a few of the possible alternative pathways. In contrast to the experimental platform of this study, exploration of a more extensive protein sequence space would allow new amino acid substitutions and enable the discovery of sequences different from LW202,[7] which could lead to high enantioselectivity in the model reaction. In view of the present results, this appears likely. Finally, the type of analysis presented herein can be used to evaluate other mutagenesis methods and strategies for probing protein sequence space in laboratory evolution.[1–5]

Experimental Section

Site-directed mutagenesis using the Stratagene QuikChange protocol[3a] was applied in order to generate all of the required mutants. pQEEH-based plasmids (4631 bp)[18] containing the proper mutations for each case were employed as templates. The sequences of the primers used in the generation of the mutants are summarized in the Supporting Information (Table S1). Successful amplifications were digested with DpnI (1 μL, 37 °C, 1 h, twice). Purified plasmids were sequenced by Medigenomix (Martinsried, Germany) using standard primers (QEF, QEFIII or QER).

All mutants were streaked out on LB agar plates from a glycerol stock. After incubation (30 °C, 60 h), single colonies were placed into LB media (3.5 mL), supplemented with carbenicillin (100 μg mL^{-1}). After incubation (30 °C, 12 h), the latter precultures (2 mL) were inoculated in fresh LB (18 mL) containing carbenicillin and incubated until OD_{600} was around three. Second-generation bacteria culture (2.4 mL) was added to a solution of sodium phosphate buffer (20 mM, pH 7.2, 16.1 mL) followed by the addition of racemic glycidyl phenyl ether, *rac*-(**1**) (83.3 mM in acetonitrile, 1.5 mL). After incubation (30 °C, 250 rpm), samples were withdrawn (after 60 min) and centrifuged (13 000 rpm, 20 min). The samples (100 μL) were mixed with the same volume of the internal standard solution, (*R*)-(+)-1-phenylbutan-1-ol, (6.66 mM in methanol). Enantioselectivity of the model reaction induced by the protein variants was measured by HPLC using a Chiralcel OD-R chiral column from Daicel Chemical Industries (Tokyo, Japan) with methanol/water (7:3) (flow rate 0.5 mL min^{-1}). Retention time: (*R*)-**2**, $t_{R(R)}$ = 8.6 min; (*S*)-**2**, $t_{R(S)}$ = 9.8 min; (*R*)-**1**, $t_{R(R)}$ = 20.9 min; (*S*)-**1**, $t_{R(S)}$ = 24.4 min.

Values for E and $\Delta\Delta G^{\ddagger}$ were calculated from equations (2)[19] and (3), respectively. Experimentally obtained (ee_p, ee_s, conversion) and calculated values (E, $\Delta\Delta G^{\ddagger}$) are listed in the Table S2. The average uncertainty of $\Delta\Delta G^{\ddagger}$ values is ±0.03 kcal mol^{-1}.

$$E = \frac{\ln\left[\frac{1-ee_S}{1+ee_S/ee_P}\right]}{\ln\left[\frac{1+ee_S}{1+ee_S/ee_P}\right]} = \frac{(k_{cat}/K_M)_{fast}}{(k_{cat}/K_M)_{slow}} \tag{2}$$

$$\Delta\Delta G^{\ddagger} = -RT \ln E \tag{3}$$

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (Schwerpunktprogramm 1170 "Directed Evolution to Optimize and Understand Molecular Biocatalysis"; Project RE 359/13–1) and the Fonds der Chemischen Industrie. We thank especially L.-W. Wang and J. D. Carballeira, as well as J. Peyralans and A. Taglieber for inspiring discussions.

Keywords: directed evolution · enantioselectivity · evolutionary pathways · fitness landscape · saturation mutagenesis

[1] Reviews of directed evolution: a) K. M. Arndt, K. M. Müller, *Methods in Molecular Biology, Vol. 352: Protein Engineering Protocols*, Humana, Totowa, **2007**; b) F. H. Arnold, G. Georgiou, *Screening and Selection Methods, Vol. 230: Directed Enzyme Evolution*, Humana, Totowa, **2003**; c) N. J. Turner, *Trends Biotechnol.* **2003**, *21*, 474–478; d) S. V. Taylor, P. Kast, D. Hilvert, *Angew. Chem.* **2001**, *113*, 3408–3436; *Angew. Chem. Int. Ed.* **2001**, *40*, 3310–3335; e) S. Brakmann, A. Schwienhorst, *Evolutionary Methods in Biotechnology: Clever Tricks for Directed Evolution*, Wiley-VCH, Weinheim, **2004**; f) E. G. Hibbert, F. Baganz, H. C. Hailes, J. M. Ward, G. J. Lye, J. M. Woodley, P. A. Dalby, *Biomol. Eng.* **2005**, *22*, 11–19; g) S. B. Rubin-Pitel, H. Zhao, *Comb. Chem. High Throughput Screening* **2006**, *9*, 247–257; h) J. Kaur, R. Sharma, *Crit. Rev. Biotechnol.* **2006**, *26*, 165–199; i) S. Bershtein, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2008**, *12*, 151–158; j) M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5716–5722; k) M. T. Reetz in *Advances in Catalysis, Vol. 49* (Eds.: B. C. Gates, H. Knözinger), Elsevier, San Diego, **2006**, pp. 1–69; l) M. T. Reetz in *Asymmetric Organic Synthesis with Enzymes* (Eds.: V. Gotor, I. Alfonso, E. García-Urdiales), Wiley-VCH, Weinheim, **2008**, pp. 21–63.

[2] a) D. W. Leung, E. Chen, D. V. Goeddel, *Technique (Philadelphia)* **1989**, *1*, 11–15; b) R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1992**, *2*, 28–33.

[3] a) H. H. Hogrefe, J. Cline, G. L. Youngblood, R. M. Allen, *BioTechniques* **2002**, *33*, 1158–1165; b) C. N. Dominy, D. W. Andrews in *Methods Molecular Biology, Vol. 235: E. coli Plasmid Vectors: Methods and Applications* (Eds.: N. Casali, A. Preston), Humana, Totowa, **2003**, pp. 209–223; c) R. Georgescu, G. Bandara, L. Sun in *Directed Evolution Library Creation* (Eds.: F. H. Arnold, G. Georgiou) Humana, Totowa, **2003**, pp. 75–84; d) R. D. Kirsch, E. Joly, *Nucleic Acids Res.* **1998**, *26*, 1848–1850.

[4] a) W. P. C. Stemmer, *Nature* **1994**, *370*, 389–391; b) 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.

[5] a) S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291–297; b) R. J. Fox, G. W. Huisman, *Trends Biotechnol.* **2008**, *26*, 132–138; c) A. Herman, D. S. Tawfik, *Protein Eng. Des. Sel.* **2007**, *20*, 219–226; d) J. D. Bloom,

M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, F. H. Arnold, *Curr. Opin. Struct. Biol.* **2005**, *15*, 447–452.
[6] 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.
[7] M. T. Reetz, L.-W. Wang, M. Bocola, *Angew. Chem.* **2006**, *118*, 1258–1263; *Angew. Chem. Int. Ed.* **2006**, *45*, 1236–1241.
[8] a) M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem.* **2006**, *118*, 7909–7915; *Angew. Chem. Int. Ed.* **2006**, *45*, 7745–7751; b) M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891–903.
[9] Examples of focused libraries:[6–8,15a] a) M. S. Warren, S. J. Benkovic, *Protein Eng.* **1997**, *10*, 63–68; b) Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585–592; c) G. P. Horsman, A. M. F. Liu, E. Henke, U. T. Bornscheuer, R. J. Kazlauskas, *Chem. Eur. J.* **2003**, *9*, 1933–1939; d) C. Nowlan, Y. Li, J. C. Hermann, T. Evans, J. Carpenter, E. Ghanem, B. K. Shoichet, F. M. Raushel, *J. Am. Chem. Soc.* **2006**, *128*, 15892–15902; e) K. Miyazaki, F. H. Arnold, *J. Mol. Evol.* **1999**, *49*, 716–720; f) L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *J. Biol. Chem.* **2004**, *279*, 46810–46817; g) S. Bartsch, R. Kourist, U. Bornscheuer, *Angew. Chem.* **2008**, *120*, 1531–1534; *Angew. Chem. Int. Ed.* **2008**, *47*, 1508–1511; h) L. Liang, J. Zhang, Z. Lin, *Microb. Cell Fact.* **2007**, *6*, 36; i) E. M. Gabor, D. B. Janssen, *Protein Eng. Des. Sel.* **2004**, *17*, 571–579; j) A. Juillerat, T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick, H. Vogel, K. Johnsson, *Chem. Biol.* **2003**, *10*, 313–317; k) N. U. Nair, H. Zhao, *ChemBioChem* **2008**, *9*, 1213–1215.
[10] M. T. Reetz, J. D. Carballeira, J. J.-P. Peyralans, H. Höbenreich, A. Maichele, A. Vogel, *Chem. Eur. J.* **2006**, *12*, 6031–6038.
[11] Y. Zou, B. M. Hallberg, T. Bergfors, F. Oesch, M. Arand, S. L. Mowbray, T. A. Jones, *Structure* **2000**, *8*, 111–122.
[12] 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.
[13] a) D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, *Science* **2006**, *312*, 111–114; b) F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, S. J. Tans, *Nature* **2007**, *445*, 383–386.
[14] a) D. C. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, *Cell* **1984**, *38*, 835–840; b) J. A. Wells, *Biochemistry* **1990**, *29*, 8509–8517; c) A. Horovitz, *Curr. Biol.* **1996**, *6*, R121-R126.
[15] a) M. T. Reetz, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem.* **2001**, *113*, 3701–3703; *Angew. Chem. Int. Ed.* **2001**, *40*, 3589–3591; b) M. Bocola, N. Otte, K.-E. Jaeger, M. T. Reetz, W. Thiel, *ChemBioChem* **2004**, *5*, 214–223; c) M. T. Reetz, M. Puls, J. D. Carballeira, A. Vogel, K.-E. Jaeger, T. Eggert, W. Thiel, M. Bocola, N. Otte, *ChemBioChem* **2007**, *8*, 106–112.
[16] A. S. Mildvan, *Biochemistry* **2004**, *43*, 14517–14520.
[17] Examples of remote effects regarding enantioselectivity/substrate acceptance: a) C.-C. Hsu, Z. Hong, M. Wada, D. Franke, C.-H. Wong, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9122–9126; b) I. Lavandera, S. Fernández, J. Magdalena, M. Ferrero, H. Grewal, C. K. Savile, R. J. Kazlauskas, V. Gotor, *ChemBioChem* **2006**, *7*, 693–698; c) S. Oue, A. Okamoto, T. Yano, H. Kagamiyama, *J. Biol. Chem.* **1999**, *274*, 2344–2349; d) S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45–54.
[18] 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.
[19] K. Faber, *Biotransformations in Organic Chemistry*, 5th ed., Springer, Berlin, **2004**.

Received: June 3, 2008
Published online on August 19, 2008

