DOI: 10.1002/cbic.200800298

Addressing the Numbers Problem in Directed Evolution

Manfred T. Reetz,* Daniel Kahakeaw, and Renate Lohmer[a]

Our previous contribution to increasing the efficiency of directed evolution is iterative saturation mutagenesis (ISM) as a systematic means of generating focused libraries for the control of substrate acceptance, enantioselectivity, or thermostability of enzymes. We have now introduced a crucial element to knowledge-guided targeted mutagenesis in general that helps to solve the numbers problem in directed evolution. We show that the choice of the amino acid (aa) alphabet, as specified by the utilized codon degeneracy, provides the experimenter with a powerful tool in designing "smarter" randomized libraries that require considerably less screening effort. A systematic comparison of two different codon degeneracies was made by examining the relative quality of the identically sized enzyme libraries in relation to the degree of oversampling required in the screening process. The specific example in our case study concerns the conventional NNK codon degeneracy (32 codons/20 aa) versus NDT (12 codons/12 aa). The model reaction is the hydrolytic kinetic resolution of a chiral trans-*disubstituted epoxide, catalyzed by the epoxide hydrolase from* Aspergillus niger. *The NDT library proves to be of much higher quality, as measured by the dramatically higher frequency of positive variants and by the magnitude of catalyst improvement (enhanced rate and enantioselectivity). We provide a statistical analysis that constitutes a useful guide for the optimal design and generation of "smarter" focused libraries. This type of approach accelerates the process of laboratory evolution considerably and can be expected to be broadly applicable when engineering functional proteins in general.*

Introduction

Directed evolution has emerged as a useful means to engineer the catalytic properties of enzymes, including thermostability, robustness in hostile organic solvents, and enantioselectivity.[1] Although an ever increasing number of successful studies continue to appear, the overall process remains time-consuming and generally requires robotic equipment to screen relatively large libraries of enzyme variants, which typically comprise 10^3–10^6 transformants (clones). Probing the vast protein sequence space, which of course is much larger, is a central challenge in directed evolution. Selection based on growth advantage of bacteria can, in principle, handle much larger numbers,[1,2] but these systems have not been developed to include such catalytic properties as enantioselectivity and/or substrate acceptance in a general way.[3]

Since the "numbers problem" in directed evolution persists to this day,[1,2,4] methodology development in the quest to probe the vast protein sequence space more efficiently than in the past has become a pressing issue. One strategy calls for even larger libraries in the order of 10^7–10^9 or more transformants, thereby covering a greater portion of protein sequence space, but requiring ultra-high-throughput screening systems.[1,2–6] In spite of progress, such as in in vitro compartmentalization,[5] pooling techniques[6] and various display systems,[2,7–10] it is currently unclear how these systems can be extended to cover such important catalytic properties as the aforementioned enantioselectivity and/or substrate acceptance in a general way. The other strategy is to consider the opposite, namely to construct smaller libraries characterized by higher quality.[1g,h] In this endeavor the comparative evaluation of two different libraries is meaningful only if the *same* number of transformants are screened in each case. Higher quality then means a higher frequency of improved enzyme variants (hits) in a given library *and* a higher degree of improvement of a given catalytic property. Consequently, library quality needs to be viewed in the light of the screening effort needed to identify beneficial hits, which is the persisting issue in directed evolution. Ideally, the size of the libraries would be reduced to such an extent that conventional GC or HPLC would suffice even in the face of such analytically challenging problems as enantio- and/or regioselectivity.

Whatever strategy is chosen, the question regarding the best choice of the mutagenesis method remains to be answered. To this day the most often applied technique in directed evolution is error-prone polymerase chain reaction (epPCR),[1,11] in addition to other methods such as saturation mutagenesis[1,12,13] and DNA shuffling[1,14] as well as variations and combinations thereof.[1] Unfortunately, few studies illuminate systematically the relative merits of the various approaches.[1,15,16] Random mutagenesis based on epPCR is often considered to address the whole gene (and thus the entire enzyme), but for several reasons it has severe disadvantages theoretically and in practice.[17,18] Due to the degeneracy of the genetic code, epPCR has significant biases. Taking this into account, it has been shown that the actual diversity of a library can be as low as 20% of the "theoretical" size, based on the usual algorithm that neglects this aspect.[17] This means that large numbers of enzyme variants are not accessible in a prac-

[a] *Prof. Dr. M. T. Reetz, Dr. D. Kahakeaw, R. Lohmer*
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

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

tical way, regardless of whether super-high-throughput screening or selection is used to find the rare positives.[1,11,17,18]

One way to address the numbers problem is to consider saturation mutagenesis, a method that restricts randomization to predetermined sites in the enzyme with the creation of focused libraries.[12,13,16,19–23] Such targeted randomization reduces the extent of protein sequence space drastically, but it requires structural information in order for the correct mutagenesis sites to be chosen. Fortunately, such data are available in most cases of interest. For example, in the quest to enhance the enantioselectivity of a lipase, we demonstrated some time ago that simultaneous randomization at a site composed of four amino acid (aa) positions next to the binding pocket is far more efficient than four consecutive rounds of epPCR, despite the fact that, in the latter case, more transformants were actually screened.[16]

We recently introduced a more systematic approach to focused library generation, namely iterative saturation mutagenesis (ISM), which is a structure-based strategy integrating knowledge-guided design and evolutionary randomization.[20–23] Two forms of ISM were developed, combinatorial active-site saturation test (CAST), for controlling substrate scope and/or enantioselectivity,[20,21] and the B-FIT technique, for increasing thermostability.[22,23] In both, sites in the enzyme, denoted as A, B, C, D, etc., are first chosen as loci for saturation mutagenesis with formation of focused libraries, a given site being composed of one, two or three (or more) aa positions in the enzyme. Following the identification of a hit from a given library, that mutant gene is used as a template for performing saturation mutagenesis at a different site, and the process is repeated as often as needed at other sites. When considering enantioselectivity and/or substrate acceptance, the criterion for choosing the proper sites is straightforward: all sites around the complete binding pocket are identified on the basis of crystal structural data or homology models.[22]

In our studies[16,20–22] and in reports by other groups regarding saturation mutagenesis[12,13,19] the issue of "oversampling", a parameter that refers to the number of enzyme variants needed to be screened in order to ensure a certain percentage coverage of a given library, was neglected.[23] However, it is essential to consider this aspect when estimating the quality of a library correctly. Fortunately, algorithms developed by Patrick and Firth,[24] Boseley and Ostermeier,[25] and Denault and Pelletier[26] are available that can be used as a basis for designing and assessing all kinds of mutagenesis libraries.

In previous studies[20–23] we generally applied NNK codon degeneracy (N: adenine/cytosine/guanine/thymine; K: guanine/thymine), which is the conventional way to perform saturation mutagenesis. It involves 32 codons and relates to all 20 proteinogenic aa's as building blocks. In the hope of profound improvements in library quality as defined above, we addressed the numbers problem from a different angle, specifically by considering codon degeneracies that encode smaller aa alphabets. Reduced aa alphabets in enzymology studies have been utilized for various purposes,[27–30] the question of minimal requirements for proper folding and enzyme activity[27,29] and binary patterning[28] being two prominent examples. For our purpose, a variety of different codon degeneracies can be considered, NDT being one of several possibilities (D: adenine/guanine/thymine; T: thymine). This choice involves 12 codons and reduces the number of aa's to twelve (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly), which is a balanced mix of polar and nonpolar, aliphatic and aromatic, and negatively and positively charged representatives, while excluding most cases of structurally similar aa.

In this model study we focused on the relative merits of NNK versus NDT[20c] codon degeneracy. The diversity of an NNK library in terms of the number of structurally different (distinct) transformants is obviously higher than that of the respective NDT library, provided essentially full library completeness (for example, $>95\%$ coverage) is ensured by the required degree of oversampling in each case. However, the percentage of improved variants relative to the total number of transformants present in the two libraries, that is the frequency of occurrence of hits with distinct sequences, cannot be expected to be identical. Moreover, if oversampling is limited to a specified number of screened clones, which in one library correlates with full or nearly full coverage, while in the other library the same number means a significantly lower percentage coverage, then vast differences in quality can arise. As shown in this study, consideration of these facets is indispensable when designing and generating higher-quality ("smarter") libraries. We thereby provide a practical tool for increasing the efficiency of saturation mutagenesis as a method in directed evolution.

Results and Discussion

Statistical analysis

We first wanted to visualize the relationship between percentage coverage of a library and the degree of oversampling, irrespective of codon usage or type of mutagenesis method. To this end we employed the previously mentioned algorithms, which are based on Poisson statistics and on the assumption that all sequences occur with equal probability.[24,25] The proposed algorithm for estimating completeness as a function of the number of transformants (clones) actually screened,[24] T, can be transformed into Equation (1), where P_i denotes the probability that a particular sequence occurs in the library, and F_i is the frequency.

$$T = -\ln(1-P_i)/F_i \quad (1)$$

Upon substituting for F_i, the relationship then reduces to Equation (2), where V is the number of gene mutants comprising a given library:

$$T = -V\ln(1-P_i) \quad (2)$$

This relationship defines the correlation between the number of mutants V of a given library and the number of transformants T that have to be screened for a specified degree of completeness. On this basis we define O_f as the "oversampling factor" [Eq. (3)]:

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

$$O_f = T/V = -\ln(1-P_i) \quad (3)$$

Upon calculating the oversampling factor O_f as a function of the percentage coverage, one obtains the curve shown in Figure 1, which should be kept in mind when designing and analyzing libraries. It can be seen that for ensuring 95% coverage, for example, the oversampling factor O_f amounts to about three, which means that a threefold excess of transformants needs to be screened. Due to the exponential character of the relationship, degrees of coverage beyond 95% require vastly higher screening efforts. Of course, lower degrees of library coverage might suffice in a given experiment,[16] but decisions regarding this important aspect can be guided by consulting Figure 1.

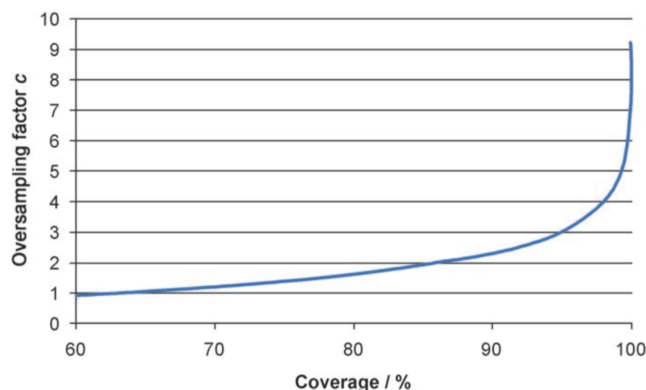


Figure 1. Correlation between library coverage and oversampling of enzyme variants.

In many saturation mutagenesis experiments sites that are composed of more than one aa position are randomized.[13,16,19–23] Using the appropriate algorithm,[24] we calculated for the NNK and NDT libraries the amount of oversampling, in absolute numbers, necessary for 95% coverage of relevant protein sequence space (Table 1). It is immediately clear that the potential screening effort is very different for the NNK versus NDT systems. For example, in the case of a site composed of three aa's, NNK requires almost 100000 clones, whereas NDT needs only about 5000 for 95% coverage.

Table 1. Oversampling necessary for 95% coverage as a function of NNK and NDT codon degeneracy.

No.[a]	NNK Codons	NNK Transformants needed	NDT Codons	NDT Transformants needed
1	32	94	12	34
2	1028	3066	144	430
3	32768	98163	1728	5175
4	1048576	3141251	20736	62118
5	33554432	100520093	248832	745433
6	$>1.0\times10^9$	$>3.2\times10^9$	$>2.9\times10^6$	$>8.9\times10^6$
7	$>3.4\times10^{10}$	$>1.0\times10^{11}$	$>3.5\times10^7$	$>1.1\times10^8$
8	$>1.0\times10^{12}$	$>3.3\times10^{12}$	$>4.2\times10^8$	$>1.3\times10^9$
9	$>3.5\times10^{13}$	$>1.0\times10^{14}$	$>5.1\times10^9$	$>1.5\times10^{10}$
10	$>1.1\times10^{15}$	$>3.4\times10^{15}$	$>6.1\times10^{10}$	$>1.9\times10^{11}$

[a] Number of aa positions at one site.

Figures 2 and 3 show this dependency over the whole range of coverage (0–95%) for sites composed of 1, 2, 3, 4, and 5 aa positions as a function of NNK or NDT codon degeneracy. Here

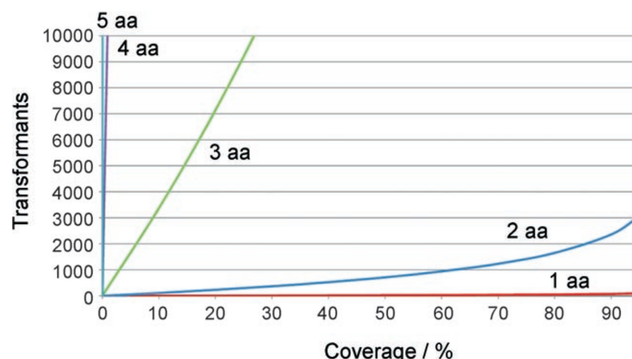


Figure 2. Library coverage calculated for NNK codon degeneracy at sites comprising 1, 2, 3, 4 and 5 amino acid positions.

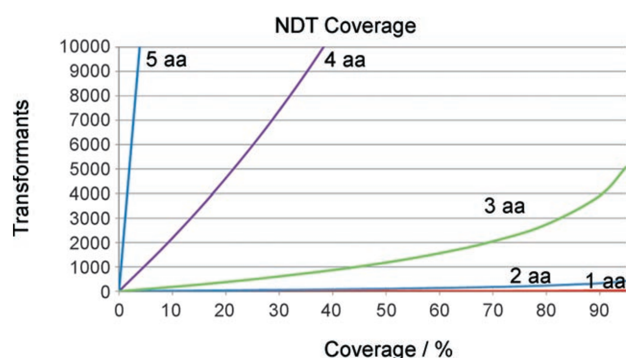


Figure 3. Library coverage calculated for NDT degeneracy at sites comprising 1, 2, 3, 4 and 5 amino acid positions.

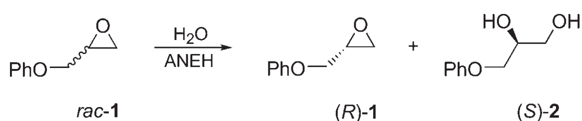
again the pronounced differences in NNK and NDT libraries become apparent. For example, if in the case of a three-aa site, the experimenter restricts the screening to 5000 transformants for practical and economical reasons, the use of an NDT library correlates with about 95% coverage, whereas the same number of screened clones in an NNK library allows for only 15% coverage. One can suspect that in such size-restricted libraries, the NDT library should be characterized by higher quality.

It is straightforward to analyze in the same way other codon degeneracies that might be more suitable for a given task. In general we conclude that it is best to choose systems in which the number of codons is equal to the number of aa's. This reduces the inherent bias and over-representation of certain aa's. Assuming all aa's are equally probable, a maximum number of distinct protein variants in a given library is then ensured. Moreover, codon degeneracies such as NDT exclude the occurrence of WT transformants in the enzyme libraries. Both facets can be expected to contribute to the quality of the focused libraries.

© 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

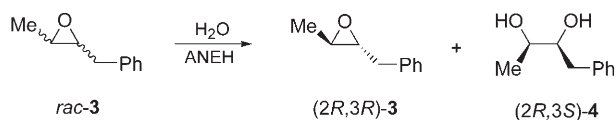
Saturation mutagenesis experiments

In order to test these expectations in the laboratory, we devised an experimental platform upon which NNK and NDT libraries can be compared systematically under identical conditions. The epoxide hydrolase (EH) from *Aspergillus niger* (ANEH)[31] served as the enzyme, which had been employed previously in the investigation into iterative CASTing as a means to enhance the enantioselectivity of the hydrolytic kinetic resolution of glycidyl phenyl ether (*rac*-**1**; Scheme 1).[21] In that study five cycles of ISM based on NNK codon degeneracy at five different sites led to a stepwise increase in the selectivity factor *E* from 4.6 to 115 in favor of (*S*)-**2.**



Scheme 1.

In this study, we first employed a more "difficult" substrate that is not accepted by the WT ANEH, namely the *trans*-disubstituted epoxide *rac*-**3** (Scheme 2). Thus, substrate acceptance



Scheme 2.

(rate) had to be engineered before considering the enantioselectivity of the respective hydrolytic kinetic resolution, making this transformation a particularly challenging model reaction for directed evolution. We return to the reaction of *rac*-**1** in the latter part of this study.

The substrate *rac*-**3** was manually docked into the binding pocket of the WT ANEH; this revealed the geometric relationship between the relevant saturation sites A–F (Figure 4). These are identical to the sites utilized in the earlier study.[21] On the basis of preliminary saturation mutagenesis experiments, sites A (Ile193/Ser195/Phe196), B (Leu215/Ala217/Arg219), and F (Phe244/Met245/Leu249) were chosen for further randomization experiments. Note that these sites are all composed of three aa positions and that positions 217 (part of site B) and 245 (part of site F) harbor Ala and Met in the WT, which are not encoded by NDT.

We first applied saturation mutagenesis[12] at site B, employing NNK and NDT degeneracy to form two separate focused libraries. By using an activity screen based on the known adrenaline test developed by Reymond,[32] which we modified to be suitable for whole-cell catalysis,[33] a total of 5000 transformants were evaluated in each library. The conditions were adjusted so that a positive hit was indicated when ≥10% conversion of

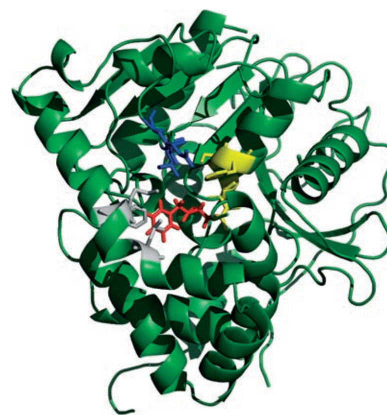


Figure 4. Predicted CAST sites of ANEH harboring the substrate *rac*-**3** (red) based on the X-ray structure of the WT.[31d] A) positions 193/195/196 (yellow), B) positions 215/217/219 (blue), and F) positions 244/245/249 (gray).

rac-**3** has been reached within 10 h of reaction time under the conditions operating in the wells of the microtiter plates. At a later stage the enantioselectivities of the hits were established by HPLC analysis of the reaction products.

The results from the two libraries proved to be drastically different (Figure 5). Whereas the NDT library provided 511 hits, the NNK library was found to contain only 38 positives. We therefore conclude from this initial observation that the utilization of NDT codon degeneracy constitutes the superior strategy. This conclusion was corroborated upon applying a more stringent catalytic threshold, namely ≥20% conversion (10 h). The NDT library was then found to contain 180 active transformants, whereas the NNK library provided only ten positives. Finally, upon tightening the restriction further to ≥40% conversion (10 h), the NDT library was found to contain 78 positives, whereas only a single hit remained in the case of NNK. Rather than sequencing all of the original 549 transformants found in the two libraries as a result of the initial screen, sequence characterization was confined to the 79 positives discovered in the final screening step. This procedure revealed 26

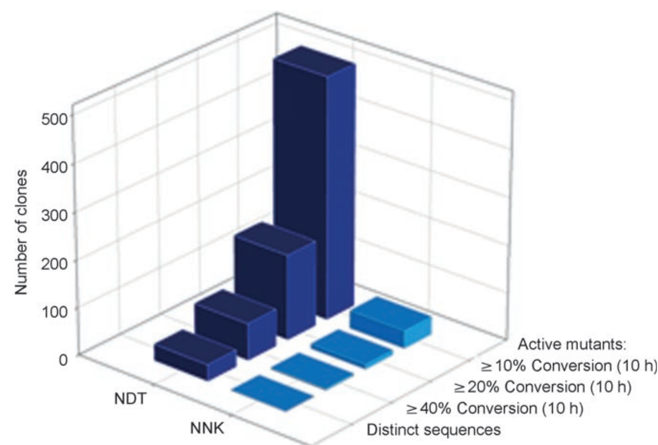


Figure 5. Frequency of active transformants in the NNK and NDT library from site B as catalysts in the hydrolytic kinetic resolution of *rac*-**3**.

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

structurally different (distinct) variants in the NDT library, which all solved the problem of substrate acceptance of *rac*-**3**. In striking contrast, only one such variant was found in the NNK library.

The observation that some of the transformants occur more than once is to be expected when applying saturation mutagenesis.[12,13] However, the number of redundant positives in the libraries described herein is gratifyingly small. Only a fraction of the original 511 (NDT) and 38 (NNK) active transformants in the initial screens were actually sequenced; this means that numerous ones displaying significant (although lower) activity have been excluded from further consideration. Interestingly, of the 26 final hits from the NDT library (site B) displaying distinct sequences, 22 are characterized by three-aa substitutions and four are double mutants, but none is a single mutant. The final hit from the NNK library is a double mutant. These results show that the highest increase in enzyme activity is achieved when all three aa's at site B are exchanged. Since the catalytic activity of the WT ANEH is extremely low (essentially no substrate acceptance), and a background reaction occurs to a small degree, it is difficult to specify accurately the actual rate acceleration. However, a rate factor of at least 10^3 can be estimated as a lower limit.

Some of the most active hits were then tested for enantioselectivity in the hydrolytic kinetic resolution of *rac*-**3**. Table 2 shows that most, but not all, of the mutants of the NDT library identified and sequenced in the final screening step show high degrees of enantioselectivity in favor of (2*R*,3*S*)-**4**, two having *E* values of >200. The kinetic property of the most active and enantioselective mutant (DK-35) was studied more closely. We used turnover frequency (TOF) as the most reliable measure of enzyme activity in this kind of transformation, as recommended in earlier studies regarding other epoxide hydrolases.[34,35] The TOF value of 55 (mol product per mol catalyst per second) proved to be considerably higher than in related cases with a different EH as a catalyst in the hydrolysis of *cis*-1,2-disubstituted epoxides (TOF = 0.07–7.2 mol product per mol catalyst per s).[34] For further comparison (Table 2), the final mutant from the NNK library was found to lead to a lower activity (TOF = 17) and to a selectivity factor of only *E* = 101.

Table 2. Selected ANEH variants (site B) as catalysts in the hydrolytic kinetic resolution of *rac*-**3** favoring (2*R*,3*S*)-**4**.

Variant	Codon degeneracy	Mutations	TOF[a]	*E*
DK-09	NDT	Leu215Ser/Ala217Tyr	14	6
DK-11	NDT	Leu215Asp/Ala217Tyr/Arg219Phe	29	56
DK-17	NDT	Leu215Gly/Ala217Tyr/Arg219Cys	12	>200
DK-21	NDT	Leu215Phe/Ala217Tyr/Arg219Leu	10	85
DK-26	NDT	Leu215Tyr/Ala217Ile/Arg219Asn	12	64
DK-27	NDT	Leu215Ser/Ala217Val/Arg219Leu	18	180
DK-28	NDT	Leu215Gly/Ala217Arg/Arg219Leu	26	62
DK-35	NDT	Leu215His/Ala217Tyr/Arg219Val	55	>200
DK-36	NDT	Leu215Phe/Ala217Cys/Arg219Ser	11	27
DK-10	NNK	Leu215Met/Arg219Ser	17	101

[a] TOF: Mol product per mol catalyst per s.

With respect to methodology development in directed evolution, which is the purpose of this study, we note that an unusually small experimental effort encompassing only 5000 transformants has generated at least two dozen solutions to the catalytic problem originally defined, namely the rapid evolution of substrate acceptance and high enantioselectivity.

Saturation mutagenesis was also performed at sites A and F, each harboring three aa positions. Here the search for improved ANEH mutants was confined to the respective NDT libraries, in each case 5000 transformants being screened. Figure 6 shows that initial saturation mutagenesis at these

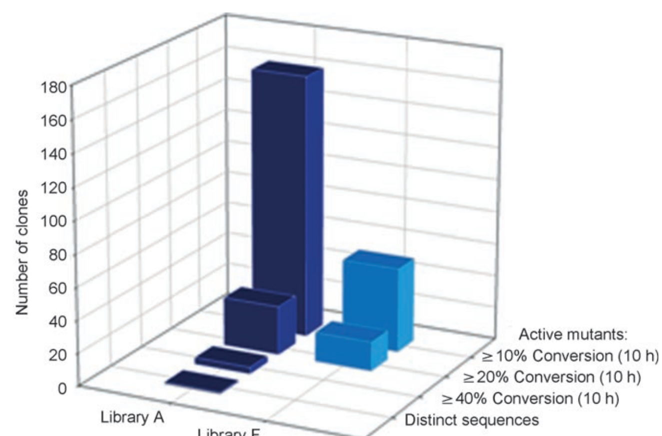


Figure 6. Frequency of active transformants in the NDT libraries from sites A and F as catalysts in the hydrolytic kinetic resolution of *rac*-**3**.

sites is not as productive as the one executed previously at site B; this need not come as a surprise. The initial test (≥ 10% conversion in 10 h) in the case of library A showed the presence of 162 active transformants, this number being reduced to 30 at ≥ 20% conversion (10 h) and to four at ≥ 40% conversion (10 h). The most active mutant arising from the library generated at site A (DK-01) proved to be fairly enantioselective (*E* = 57 in favor of (2*R*,3*S*)-**4**). Library F provides fewer active hits (Figure 6). The lower catalytic performance of these mutants does not mean that their genes are of little use in future iterative saturation mutagenesis experiments, since they can be employed as starting points for CASTing at the other sites. Alternatively, the genes of some of the best mutants obtained from the library at site B could serve as logical templates for saturation mutagenesis at sites A or F.

Finally, we considered the hydrolytic kinetic resolution of substrate *rac*-**1**, which had served as the model reaction in the original study when using NNK codon degeneracy.[21] Rather than screening the already available NDT library at site B in its entirety, we simply restricted this part of the study to some of the best hits identified earlier as active catalysts in the reaction of *rac*-**3**. Upon testing only ten transformants, this search led to the discovery of several highly active and surprisingly enantioselective ANEH mutants (Table 3). For example, mutant DK-27 has an *E* value of 35 in favor of (*S*)-**2**. In our original study such respectable enantioselectivity was not achieved until the

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 3. Selected ANEH variants as catalysts in the hydrolytic kinetic resolution of *rac*-**1** favoring (*S*)-**2**, obtained previously in the case of substrate *rac*-**3** by saturation mutagenesis at site B by using NDT codon degeneracy.

Variant	Mutations	*E*
DK-04	Leu215Ser/Ala217Ile/Arg219Arg	16
DK-27	Leu215Ser/Ala217Val/Arg219Leu	35
DK-35	Leu215His/Ala217Tyr/Arg219Val	25

fourth cycle of iterative CASTing at four different sites (B, C, D, and F) along the upward pathway when using NNK codon degeneracy.[21]

Conclusions

We have addressed once more the numbers problem in directed evolution by considering statistical aspects of library generation and performing appropriate laboratory experiments based on saturation mutagenesis. The model study involves the epoxide hydrolase from *Aspergillus niger* (ANEH) as the enzyme with enhancement of the reaction rate (substrate acceptance) and enantioselectivity in the hydrolytic kinetic resolution of a chiral substrate serving as the catalytic parameters. Whereas previous studies have demonstrated the merits of targeted randomization,[12,13,16,19] culminating in iterative saturation mutagenesis (ISM),[22,23] this contribution provides a tool that is useful in the quest to generate even "smarter" enzyme libraries. Its basis is the optimal choice of the codon degeneracy when tailoring saturation mutagenesis experiments. In the model study, we have compared systematically two different codon degeneracies that relate to two differently sized aa alphabets, specifically the conventional NNK (32 codons) encoding 20 aa versus NDT (12 codons) encoding only 12 aa's. This led to the discovery that vastly different library qualities result. Upon comparing identically sized NNK and NDT libraries each limited to 5000 transformants, the frequency of hits in the NDT case turned out to be much higher. The catalytic profiles of the best hits in terms of activity and enantioselectivity also proved to be different, those resulting from NDT codon usage again being superior. These experimental results are in line with a statistical analysis based on algorithms proposed earlier.[24,25] The basic prediction regarding the main trends is not expected to change by considering possible aa bias. Nevertheless, this aspect is the subject of an ongoing project in our laboratories.

We conclude that it is possible to decrease the size of focused libraries generated by targeted mutagenesis, while increasing their quality in terms of the frequency of beneficial variants *and* the degree of catalyst improvement. This is possible for two reasons. The correct choice of codon degeneracy ensures the proper minimal set of aa's in terms of structural and electronic characteristics *and* allows for nearly complete coverage of a given size-restricted library. Both factors working together maximize the probability of hits and reduce the number of "junk" transformants. These features have been uncovered experimentally for the first time in a model case, but we expect the trends to be general.

NDT is not the only codon degeneracy for constructing reduced aa alphabets, others can likewise be considered depending upon the nature of the protein property to be engineered (catalytic profile of enzymes or binding property of proteins). Excluding those aa's that are similar to those that are needed minimally is one important recommendation. The other general guideline for achieving more efficient and therefore faster directed evolution is derived from our latest data and from the statistical analysis, namely that the number of codons should be equal to the number of aa's being used as building blocks when performing saturation mutagenesis.

This approach is not limited to CASTing, other forms of targeted mutagenesis as in the B-Fit method[22,23] are also relevant. Moreover, reduced aa alphabets as described herein can also be used when combining site randomization with DNA shuffling[16] or synthetic shuffling.[36]

Finally, we point out that the screening/selection problem when applying targeted mutagenesis becomes even more acute when randomizing sites composed of five or more aa positions. Even if assay systems capable of rapid evaluation of 10^8–10^9 transformants for activity/enantioselectivity were to be developed in a general way,[37] this would not suffice to cover the relevant protein sequence space when applying saturation mutagenesis based on the conventional NNK approach (Table 1). We suggest that in this kind of broadly targeted mutagenesis, which might be beneficial in some systems (for example, randomization at aa positions of an enzyme domain), the strategy described herein could provide the experimenter with an effective tool for handling the numbers problem in directed evolution.[38]

Experimental Section

Vector construction: The PCR-amplified ANEH genes were cloned into vector pET22b(+) (Novagen, Madison, USA) between the EcoRI and MscI restriction sites, and *E. coli* BL21Gold(DE3) (Novagen, Madison, USA) was transformed with the resulting plasmids.

Oligonucleotides: The NDT and NNK libraries at site B (215, 217, 219) were prepared by using the complementary primers (Invitrogen) NDT-B-for (5′-CGGTTCATTTGAACNDTTGCNDTATGNDTGCTCCCCCTGAG-3′) and NDT-B-rev (5′-CTCAGGGGGAGCAHNCATAHNGCAAHNGTTCAAATGAACCG-3′) as well as NNK-B-for (5′-CGGTTCATTTGAACNNKTGCNNKATGNNKGCTCCCCCTGAG-3′) and NNK-B-rev (5′-CTCAGGGGGAGCMNNCATMNNGCAMNNGTTCAAATGAACCG-3′). Primers for NDT library at site A (193, 195, 196) were: NDT-A-for (5′-GGGAGGTGATNDTGGTNDTNDTGTTGGACGACTGTTGG-3′) and NDT-A-rev (5′-CCAACAGTCGTCCAACAHNAHNACCAHNATCACCTCCC-3′). For NDT library at site F (244, 245, 247), NDT-F-for (5′-CGAATGGAGAAGNDTNDTACCGATGGCNDTGCTTATGCCATG-3′) and NDT-F-rev (5′-CATGGCATAAGCAHNGCCATCGGTAHNAHNCTTCTCCATTCG-3′) were used.

Library construction: Saturation mutagenesis libraries were generated by the QuikChange™ mutagenesis method[12] with the above degenerate primers. Briefly, each PCR contained (25 μL final volume): 10×KOD buffer (Novagen; 2.5 μL), $MgCl_2$ (1 μL, 25 mM), dNTP (5.0 μL, 2 mM each), appropriate degenerate primers (5.0 μL,

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

2.5 mM), template plasmid (0.5 μL, 10 ng μL^{-1}) and 0.5 units of KOD Hot start polymerase. PCRs were carried out on a Biometra Thermocycler (Whatman Biometra, Göttingen, Germany). Thermal cycling consisted of an initial denaturation of 3 min at 94 °C, followed by 25 cycles of 60 s at 52 °C, 60 s at 72 °C, 480 s at 72 °C, and a final elongation of 14 min at 72 °C. Residual template in each PCR reaction was removed by double digestion with 1 unit of DpnI (New England Biolabs) in 1× manufacturer's buffer for 2 h at 37 °C, followed by further addition of 1.0 unit of DpnI for 1 h. The PCR product was used to transform competent BL21Gold(DE3) (Novagen) cells, the cells were plated on LB$^{CB/TET}$ plates.

Gene sequencing: Mutant genes were sequenced by using the standard T7 and T7-Ter primer (Medigenomix, Martinsried, Germany).

Expression for screening: Colonies were picked with a colony picker QPIX (Genetix, New Milton, UK), and precultures were grown overnight at 3 °C and 800 rpm in deep-well plates with lactose-free 505 medium (800 μL per well; formula per liter: 20 mL of 50×505 solution (250 g L^{-1} glycerol, 25 g L^{-1} glucose, and distilled water), 50 mL of 20×NPS solution (66 g L^{-1} $(NH_4)_2SO_4$, 136 g L^{-1} KH_2PO_4 and 142 g Na_2HPO_4), 2 mL of 1 mM $MgSO_4$, and filled to a volume of 1 L with ZY medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract)). The preculture (50 μL) was used to inoculate an expression culture with 5052 medium (800 μL per well) containing lactose as inducer (formula per liter: 20 mL of 50×5052 solution (100 g L^{-1} α-lactose, 250 g L^{-1} glycerol, 25 g L^{-1} glucose, and distilled water), 50 mL of 20×NPS solution, 2 mL of 1 mM $MgSO_4$, and filled to a volume 1 L with ZY medium)). In both media, carbenicillin was used as antibiotic (100 mg L^{-1}). After being grown at 37 °C and 800 rpm for 6 h, cultures were centrifuged at 4000 rpm for 5 min. Cells were washed by centrifuging for 3 min at 4000 rpm and 4 °C with PBS buffer (800 μL; Na_2HPO_4 1.1 g L^{-1}, NaH_2PO_4 0.3 g L^{-1}, NaCl 9.0 g L^{-1}). Supernatants were discarded, and cells were resuspended in PBS buffer (800 μL).

Screening: Stock solutions of *rac*-1-phenyl-2,3-epoxy butane (*rac*-**3**; 72 mmol substrate in acetonitrile) were stored at −20 °C. The reactions were performed as follows: washed expression culture (20 μL) was added to PBS buffer (150 μL, 57 mmol, pH 7.2), and the substrate was dissolved in acetonitrile (10 μL, 72 mmol). The reaction suspension was incubated for 10 h at 800 rpm and 37 °C. The hydrolytic reaction was monitored by using a cell-based adrenaline assay for high-throughput screening.[33] The change of absorption was obtained with a Spectramax UV/Vis spectrophotometer (Molecular Devices Corp.). Initial activity was assessed with a threshold value of 10%, incorporating procedural variabilities and the expected conversion. Active clones were collected and reproduced in triplicate based on the aforementioned preculture. Mutant genes were sequenced by using the standard T7 and T7-Ter primers (Medigenomix). A similar procedure was used for 20% and 40% conversions.

Variant expression and purification: For all variants, expression was performed in a ZY 5052 expression culture (100 mL) supplemented with carbenicillin (100 mg L^{-1}). The culture was incubated at 30 °C and 250 rpm overnight. The cells were chilled on ice for 20 min, harvested (5 min, 10000 rpm, 4 °C), washed with PBS buffer (57 mmol, pH 7.2), and centrifuged again (3 min, 10000 rpm, 4 °C). The cell pellet was typically resuspended in PBS buffer (8 mL, 57 mmol, pH 7.2) containing DNase I (0.5 mg mL^{-1}; Applichem, Darmstadt, Germany) and lysozyme (1.0 mg mL^{-1}?; GERBU Biotechnik, Gaiberg, Germany). The suspension was chilled on ice for 1 h before sonication (Bandelin, 2×30 s, 40% pulse, on ice). After centrifugation (14000 rpm, 45 min, 4 °C) a clear supernatant was collected. The variants were purified by ion-exchange chromatography (HiTrap Q HP, 5 mL, GE Healthcare) with a step-gradient and a 5 mL min^{-1} flow rate (starting buffer A: 57 mM phosphate buffer pH 7.4; elution buffer B: 57 mM phosphate buffer pH 7.4 with 1 M NaCl).

HPLC analysis: The chiral analyses of the hydrolytic kinetic resolution reaction of *rac*-**3** were performed by using Chiracel AD-RH chiral column (2.5 m, 4.6 mm i.d., Daicel Chemical Industries, Tokyo, Japan). Conditions: methanol/H_2O, 0.5 mL min^{-1}, UV 210 nm.

Acknowledgements

This work was supported by the Fonds der Chemischen Industrie (FCI) and the DFG (SSP 1170 "Directed Evolution to Optimize and Understand Molecular Biocatalysis"; Project RE 359/13-1). D.K. thanks the FCI for a Kekulé Stipend. We thank A. Vogel, J. D. Carballeira, F. Schulz, J. Sanchis, L. Fernández, and H. Hinrichs for helpful discussions.

Keywords: codon degeneracy · enantioselectivity · enzymes · epoxides · hydrolases · 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, *Directed Enzyme Evolution: Screening and Selection Methods, Vol. 230*, 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) J. Kaur, R. Sharma, *Crit. Rev. Biotechnol.* **2006**, *26*, 165–199; f) S. Brakmann, A. Schwienhorst, *Evolutionary Methods in Biotechnology: Clever Tricks for Directed Evolution*, Wiley-VCH, Weinheim, **2004**; g) S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291–297; h) R. J. Fox, G. W. Huisman, *Trends Biotechnol.* **2008**, *26*, 132–138; i) 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; j) S. B. Rubin-Pitel, H. Zhao, *Comb. Chem. High Throughput Screening* **2006**, *9*, 247–257; k) S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. S. Tawfik, *Nature* **2006**, *444*, 929–932; l) J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, F. H. Arnold, *Curr. Opin. Struct. Biol.* **2005**, *15*, 447–452; m) M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5716–5722; n) M. T. Reetz in *Advances in Catalysis, Vol. 49* (Eds.: B. C. Gates, H. Knözinger), Elsevier, San Diego, **2006**, pp. 1–69.

[2] a) J.-L. Reymond, *Enzyme Assays: High-Throughput Screening, Genetic Selection and Fingerprinting*, Wiley-VCH, Weinheim, **2005**; b) H. Leemhuis, V. Stein, A. D. Griffiths, F. Hollfelder, *Curr. Opin. Struct. Biol.* **2005**, *15*, 472–478; c) A. Aharoni, A. D. Griffiths, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2005**, *9*, 210–216; d) Y. L. Boersma, M. J. Dröge, W. J. Quax, *FEBS J.* **2007**, *274*, 2181–2195.

[3] a) M. T. Reetz, L.-W. Wang, *Comb. Chem. High Throughput Screening* **2006**, *9*, 295–299; b) M. T. Reetz, C. J. Rüggeberg, *Chem. Commun.* **2002**, 1428–1429; c) Y. L. Boersma, M. J. Dröge, A. M. van der Sloot, T. Pijning, R. H. Cool, B. W. Dijkstra, W. J. Quax, *ChemBioChem* **2008**, *9*, 1110–1115.

[4] G. A. Weiss, *Curr. Opin. Biotechnol.* **2007**, *11*, 241–243.

[5] A. D. Griffiths, D. S. Tawfik, *Trends Biotechnol.* **2006**, *24*, 395–402.

[6] K. M. Polizzi, M. Parikh, C. U. Spencer, I. Matsumura, J. H. Lee, M. J. Realff, A. S. Bommarius, *Biotechnol. Prog.* **2006**, *22*, 961–967.

[7] a) S. Y. Lee, J. H. Choi, Z. Xu, *Trends Biotechnol.* **2003**, *21*, 45–52; b) N. Varadarajan, J. Gam, M. J. Olsen, G. Georgiou, B. L. Iverson, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6855–6860; c) M. J. Dröge, Y. L. Boersma, G. van Pouderoyen, T. E. Vrenken, C. J. Rüggeberg, M. T. Reetz, B. W. Dijkstra, W. J. Quax, *ChemBioChem* **2006**, *7*, 149–157.

[8] Y. L. Boersma, M. J. Dröge, W. J. Quax, *FEBS J.* **2007**, *274*, 2181–2195; M. Widersten, L. O. Hansson, L. Tronstad, B. Mannervik, *Methods Enzymol.* **2000**, *328*, 389–404.

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[9] D. Lipovsek, E. Antipov, K. A. Armstrong, M. J. Olsen, A. M. Klibanov, B. Tidor, K. D. Wittrup, *Chem. Biol.* **2007**, *14*, 1176–1185.
[10] S. Becker, H. Höbenreich, A. Vogel, J. Knorr, S. Wilhelm, F. Rosenau, K.-E. Jaeger, M. T. Reetz, H. Kolmar, *Angew. Chem.* **2008**, *120*, in press; DOI: 10.1002/ange.200705236; *Angew. Chem. Int. Ed.* **2008**, *47*, in press; DOI: 10.1002/anie.200705236.
[11] a) D. W. Leung, E. Chen, D. V. Goeddel, *Techniques* **1989**, *1*, 11–15; b) R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1992**, *2*, 28–33.
[12] Details of QuikChange^{TM} protocol of Stratagene: H. H. Hogrefe, J. Cline, G. L. Youngblood, R. M. Allen, *BioTechniques* **2002**, *33*, 1158–1165.
[13] Numerous different molecular biological approaches to saturation (or cassette) mutagenesis are known,[12] for example: a) C. N. Dominy, D. W. Andrews in *Methods in Molecular Biology, Vol. 235: E. coli Plasmid Vectors, Methods and Applications* (Eds.: N. Casali, A. Preston), Humana, Totowa, **2003**, pp. 209–223; b) M. A. Vandeyar, M. P. Weiner, C. J. Hutton, C. A. Batt, *Gene* **1988**, *65*, 129–133; c) R. D. Kirsch, E. Joly, *Nucleic Acids Res.* **1998**, *26*, 1848–1850; d) L. Zheng, U. Baumann, J.-L. Reymond, *Nucleic Acids Res.* **2004**, *32*, e115; e) M. Smith, *Ann. Rev. Genet.* **1985**, *19*, 423–462; f) R. Georgescu, G. Bandara, L. Sun in *Directed Evolution Library Creation*, (Eds.: F. H. Arnold, G. Georgiou), Humana, Totowa, **2003**, pp. 75–84; g) J. A. Wells, M. Vasser, D. B. Powers, *Gene* **1985**, *34*, 315–323; h) D. Barettino, M. Feigenbutz, R. Valcarcel, H. G. Stunnenberg, *Nucleic Acids Res.* **1994**, *22*, 541–542; i) M. S. Z. Horwitz, L. A. Loeb, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7405–7409; j) A. R. Oliphant, A. L. Nussbaum, K. Struhl, *Gene* **1986**, *44*, 177–184; k) J. F. Reidhaar-Olson, R. T. Sauer, *Science* **1988**, *241*, 53–57; l) J. D. Hermes, S. M. Parekh, S. C. Blacklow, H. Köster, J. R. Knowles, *Gene* **1989**, *84*, 143–151; m) G. Sarkar, S. S. Sommer, *BioTechniques* **1990**, *8*, 404–407; n) A. Martin, E. Toselli, M.-F. Rosier, C. Auffray, M.-D. Devignes, *Nucleic Acids Res.* **1995**, *23*, 1642–1643; o) B. Borrego, A. Wienecke, A. Schwienhorst, *Nucleic Acids Res.* **1995**, *23*, 1834–1835.
[14] 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.
[15] a) L. A. Rowe, M. L. Geddie, O. B. Alexander, I. Matsumura, *J. Mol. Biol.* **2003**, *332*, 851–860; b) J. M. Joern, P. Meinhold, F. H. Arnold, *J. Mol. Biol.* **2002**, *316*, 643–656; c) R. M. Kelly, H. Leemhuis, L. Dijkhuizen, *Biochemistry* **2007**, *46*, 11216–11222; d) T.-W. Wang, H. Zhu, X.-Y. Ma, T. Zhang, Y.-S. Ma, D.-Z. Wie, *Mol. Biotechnol.* **2006**, *34*, 55–68.
[16] M. T. Reetz, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem.* **2001**, *113*, 3701–3703; *Angew. Chem. Int. Ed.* **2001**, *40*, 3589–3591.
[17] T. Eggert, M. T. Reetz, K.-E. Jaeger in *Enzyme Functionality-Design, Engineering, and Screening* (Ed.: A. Svendsen), Marcel Dekker, New York, **2004**, pp. 375–390.
[18] T. S. Wong, K. L. Tee, B. Hauer, U. Schwaneberg, *Nucleic Acids Res.* **2004**, *32*, e26; b) D. A. Drummond, B. L. Iverson, G. Georgiou, F. H. Arnold, *J. Mol. Biol.* **2005**, *350*, 806–816; c) S. Shafikhani, R. A. Siegel, E. Ferrari, V. Schellenberger, *BioTechniques* **1997**, *23*, 304–312.
[19] Examples of focused libraries:[16] a) Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585–592; b) M. S. Warren, S. J. Benkovic, *Protein Eng.* **1997**, *10*, 63–68; c) G. P. Horsman, A. M. F. Liu, E. Henke, U. T. Bornscheuer, R. J. Kazlauskas, *Chem. Eur. J.* **2003**, *9*, 1933–1939; d) J. Yang, Y. Koga, H. Nakano, T. Yamane, *Protein Eng.* **2002**, *15*, 147–152; e) L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *J. Biol. Chem.* **2004**, *279*, 46810–46817; f) S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45–54; g) 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; h) K. Miyazaki, F. H. Arnold, *J. Mol. Evol.* **1999**, *49*, 716–720; i) G. Valinger, M. Hermann, U. G. Wagner, H. Schwab, *J. Biotechnol.* **2007**, *129*, 98–108; j) S. Bartsch, R. Kourist, U. Bornscheuer, *Angew. Chem.* **2008**, *120*, 1531–1534; *Angew. Chem. Int. Ed.* **2008**, *47*, 1508–1511; k) L. Liang, J. Zhang, Z. Lin, *Microb. Cell Fact.* **2007**, *6*, 36; l) J. Chelliserrykattil, A. D. Ellington, *Nat. Biotechnol.* **2004**, *22*, 1155–1160; m) E. M. Gabor, D. B. Janssen, *Protein Eng. Des. Sel.* **2004**, *17*, 571–579; n) M. Fa, A. Radeghieri, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 1748–1754; o) A. Juillerat, T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick, H. Vogel, K. Johnsson, *Chem. Biol.* **2003**, *10*, 313–317; p) N. U. Nair, H. Zhao, *ChemBioChem* **2008**, *9*, 1213–1215.
[20] a) 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; b) C. M. Clouthier, M. M. Kayser, M. T. Reetz, *J. Org. Chem.* **2006**, *71*, 8431–8437; c) M. T. Reetz, J. J.-P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun.* **2006**, 4318–4320.
[21] 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.
[22] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem.* **2006**, *118*, 7909–7915; *Angew. Chem. Int. Ed.* **2006**, *45*, 7745–7751.
[23] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891–903.
[24] a) W. M. Patrick, A. E. Firth, *Biomol. Eng.* **2005**, *22*, 105–112; b) W. M. Patrick, A. E. Firth, J. M. Blackburn, *Protein Eng.* **2003**, *16*, 451–457; see also c) M. A. Mena, P. S. Daugherty, *Protein Eng. Des. Sel.* **2005**, *18*, 559–561.
[25] A. D. Bosley, M. Ostermeier, *Biomol. Eng.* **2005**, *22*, 57–61.
[26] M. Denault, J. N. Pelletier in *Methods in Molecular Biology, Vol. 352: Protein Engineering Protocols* (Eds.: K. M. Arndt, K. M. Müller), Humana, Totowa, **2007**, pp. 127–154.
[27] A. R. Davidson, K. J. Lumb, R. T. Sauer, *Nat. Struct. Biol.* **1995**, *2*, 856–864.
[28] M. H. Hecht, A. Das, A. Go, L. H. Bradley, Y. Wei, *Protein Sci.* **2004**, *13*, 1711–1723; and references therein.
[29] K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem.* **2005**, *280*, 37742–37746.
[30] a) T. Li, K. Fan, J. Wang, W. Wang, *Protein Eng.* **2003**, *16*, 323–330; b) S. Akanuma, T. Kigawa, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13549–13553; c) S. S. Sidhu, A. A. Kossiakoff, *Curr. Opin. Chem. Biol.* **2007**, *11*, 347–354; d) A. Rajpal, N. Beyaz, L. Haber, G. Cappuccilli, H. Yee, R. R. Bhatt, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8466–8471; e) E. Susko, A. J. Roger, *Mol. Biol. Evol.* **2007**, *24*, 2139–2150; f) A. D. Solis, S. Rackovsky, *Proteins Struct. Funct. Bioinf.* **2007**, *67*, 785–788; g) N. Doi, K. Kakukawa, Y. Oishi, H. Yanagawa, *Proteins Eng. Des. Sel.* **2005**, *18*, 279–284.
[31] a) C. Morisseau, A. Archelas, C. Guitton, D. Faucher, R. Furstoss, J. C. Baratti, *Eur. J. Biochem.* **1999**, *263*, 386–395; b) K. M. Manoj, A. Archelas, J. C. Baratti, R. Furstoss, *Tetrahedron* **2001**, *57*, 695–701; c) M. Arand, H. Hemmer, H. Dürk, J. Baratti, A. Archelas, R. Furstoss, F. Oesch, *Biochem. J.* **1999**, *344*, 273–280; d) J. Y. Zou, B. M. Hallberg, T. Bergfors, F. Oesch, M. Arand, S. L. Mowbray, T. A. Jones, *Structure* **2000**, *8*, 111–122.
[32] D. Wahler, J.-L. Reymond, *Angew. Chem.* **2002**, *114*, 1277–1280; *Angew. Chem. Int. Ed.* **2002**, *41*, 1229–1232.
[33] D. Kahakeaw, M. T. Reetz, *Chem. Asian J.* **2008**, *3*, 233–238.
[34] L. Zhao, B. Han, Z. Huang, M. Miller, H. Huang, D. S. Malashock, Z. Zhu, A. Milan, D. E. Robertson, D. P. Weiner, M. J. Burk, *J. Am. Chem. Soc.* **2004**, *126*, 11156–11157.
[35] Reviews of epoxide hydrolases: a) K. Faber, R. V. A. Orru in *Enzyme Catalysis in Organic Synthesis*, 2nd ed. (Eds.: K. Drauz, H. Waldmann), Wiley-VCH, Weinheim, **2002**, pp. 579–608; b) C. Morisseau, B. D. Hammock, *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 311–333; c) A. Archelas, R. Furstoss, *Curr. Opin. Chem. Biol.* **2001**, *5*, 112–119.
[36] a) D. Zha, A. Eipper, M. T. Reetz, *ChemBioChem* **2003**, *4*, 34–39; b) W. M. Coco, L. P. Encell, W. E. Levinson, M. J. Crist, A. K. Loomis, L. L. Licato, J. J. Arensdorf, N. Sica, P. T. Pienkos, D. J. Monticello, *Nat. Biotechnol.* **2002**, *20*, 1246–1250; c) J. E. Ness, S. Kim, A. Gottman, R. Pak, A. Krebber, T. V. Borchert, S. Govindarajan, E. C. Mundorff, J. Minshull, *Nat. Biotechnol.* **2002**, *20*, 1251–1255; d) R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338–344.
[37] Such high numbers (Table 1) might not be realistic in a practical sense, that is, it could well be difficult to obtain exceedingly large libraries of this kind in real laboratory experiments, a limitation that is different from the classical screening/selection problem.
[38] In certain cases it might be beneficial to limit the aa alphabet to 4–5 members.

Received: April 30, 2008
Published online on June 20, 2008

 © 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim