CHEMBIOCHEM

ChemPubSoc Europe

DOI: 10.1002/cbic.201100784

Many Pathways in Laboratory Evolution Can Lead to Improved Enzymes: How to Escape from Local Minima

Yosephine Gumulya,[a] Joaquin Sanchis,[a] and Manfred T. Reetz*[a, b]

Directed evolution is a method to tune the properties of enzymes for use in organic chemistry and biotechnology, to study enzyme mechanisms, and to shed light on Darwinian evolution in nature. In order to enhance its efficacy, iterative saturation mutagenesis (ISM) was implemented. This involves: 1) randomized mutation of appropriate sites of one or more residues; 2) screening of the initial mutant libraries for properties such as enzymatic rate, stereoselectivity, or thermal robustness; 3) use of the best hit in a given library as a template for saturation mutagenesis at the other sites; and 4) continuation of the process until the desired degree of enzyme improvement has been reached. Despite the success of a number of ISM-based studies, the question of the optimal choice of the many different possible pathways remains unanswered. Here we considered a complete 4-site ISM scheme. All 24 pathways were systematically explored, with the epoxide hydrolase from *Aspergillus niger* as the catalyst in the stereoselective hydrolytic kinetic resolution of a chiral epoxide. All 24 pathways were found to provide improved mutants with notably enhanced stereoselectivity. When a library failed to contain any hits, non-improved or even inferior mutants were used as templates in the continuation of the evolutionary pathway, thereby escaping from the local minimum. These observations have ramifications for directed evolution in general and for evolutionary biological studies in which protein engineering techniques are applied.

Introduction

Directed evolution has emerged as a general way to enhance essentially any property of a protein,[1] including thermal robustness,[1,2] substrate scope, and stereoselectivity of enzymes.[1,3] This allows the traditional limitations of enzymes as catalysts in synthetic organic chemistry and biotechnology to be lifted.[4] The process involves repeated cycles of gene mutagenesis, expression and screening (selection). Many different mutagenesis techniques and strategies have been reported, most of which are based on error-prone PCR (epPCR), saturation mutagenesis, or DNA shuffling.[1] With each approach, enzyme improvement can generally be expected, but the degree of the refinement depends on the amount of laboratory work that the experimenter is willing to invest. In order to maximize efficiency and reliability (necessary in practical applications), recent research has focused on developing advanced ways to probe protein sequence space thereby leading to higher-quality mutant libraries and thus less screening (the bottleneck in laboratory-based evolution).[1,3,5–8] Our contribution in this endeavor is iterative saturation mutagenesis (ISM), which relies on data derived from structural biology in order to identify appropriate randomization sites.[3,9,10] These sites, designated as A, B, C, D, etc., can be composed of one or more amino acid positions. Following saturation mutagenesis and screening for a specified catalytic property, the mutated genes of the improved variants ("hits") are used for saturation mutagenesis at the other sites, and the process is continued until the desired degree of catalyst improvement has been achieved. The possible pathways for four ISM sites are illustrated in Figure 1.

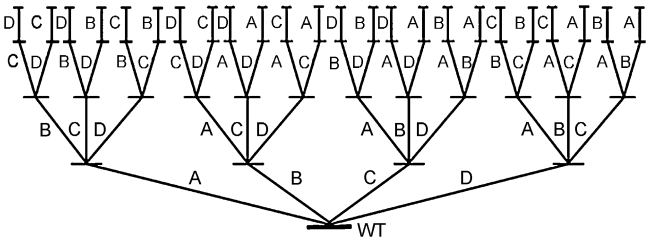


Figure 1. Iterative saturation mutagenesis (ISM) with a 4-site system (A, B, C, and D), each comprising one or more amino acid positions; in all 24 upward pathways each site is considered only once.[3,9] The equivalent 5-site ISM scheme contains 120 different pathways.

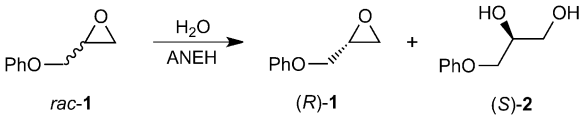
The choice of the randomization sites depends upon the nature of the protein property to be improved. When evolving thermal robustness by ISM, the B-FIT method is suitable:[10] sites composed of residues with the highest average B-factors, as determined from X-ray data, are selected for saturation mutagenesis. In the case of stereoselectivity and substrate scope, the combinatorial active-site saturation test (CAST) is appropriate. For these, the first- or second-sphere sites around the binding pocket are chosen, as revealed by X-ray data or homology models.[11]

[a] *Dr. Y. Gumulya, Dr. J. Sanchis, Prof. Dr. M. T. Reetz*
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr (Germany)

[b] *Prof. Dr. M. T. Reetz*
Fachbereich Chemie, Philipps-Universität Marburg
Hans-Meerwein-Strasse, 35032 Marburg (Germany)
E-mail: reetz@mpi-muelheim.mpg.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201100784.

As an example of ISM for enhancing enzymatic enantioselectivity, the *Aspergillus niger* epoxide hydrolase (ANEH) was selected, and the hydrolytic kinetic activity with glycidyl phenyl ether (*rac*-**1**) was employed as the model reaction (Scheme 1).[9] Wild-type (WT) ANEH is a poor catalyst, with a selectivity factor of only $E=4.6$ slightly in favor of (*S*)-**2**. *E* values reflect the relative reaction rate of the two enantiomers, which is independent of the catalytic efficiency.



Scheme 1. Enzymatic reaction of ANEH.

In our previous work using the X-ray structure of WT ANEH,[12] CASTing identified six potential randomization sites: A (positions 193/195/196), B (215/217/219), C (329/330), D (349/350), E (317/318), and F (244/245/249) (Figure 2).[9] Five of these were then selected for saturation mutagenesis in the order B→C→D→F→E, and the final variant LW202 displayed high *S* enantioselectivity ($E=115$). Libraries were generated for all six sites, A–F. The best mutant originated from library B; site A was excluded because sufficient stereoselectivity had already been achieved.[9] At the time (2006), it appeared logical to start with B because in the initial round of mutagenesis experiments library B provided the best hit, yet the order of visiting the other sites was completely arbitrary, not based on structural information or mechanistic insight. In all cases NNK codon degeneracy was used, thus encoding all 20 natural amino acids, but oversampling was not considered at the time. Typically, only about 2000 transformants were screened in each library. For 95% library coverage, as subsequently computed,[10,13] a two-residue site would require the screening of about 3000 transformants, assuming no amino acid bias (98000 for a three-residue site). Variant LW202 is characterized by five sets of mutations: B (Leu215Phe/Ala217Asn/Arg219Ser), C (Met329Pro/Leu330Tyr), D (Cys350Val), F (Leu249Tyr), and E (Thr317Trp/Thr318Val), a total of nine amino acid mutations.[9] Site A was not included because the achieved 25-fold improvement in stereoselectivity fulfilled practical needs at the time. A biophysical follow-up study revealed the explanation for the increased enantioselectivity at the molecular level.[14] Nevertheless, it was unclear whether "fortune" was involved when choosing the particular pathway B→C→D→F→E, one of 120 possible "trajectories" ($5!=120$). This uncertainty also applies to ISM-based studies of other enzymes by us and others that used arbitrarily chosen pathways.[3,15–28]

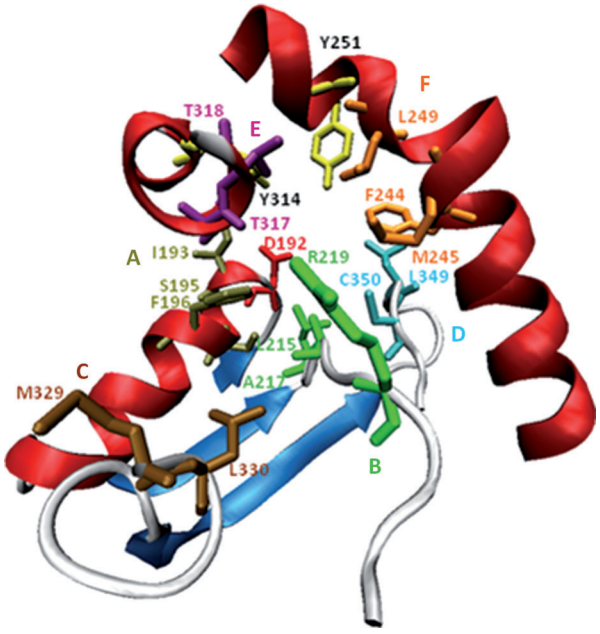


Figure 2. Six potential CAST sites (A, B, C, D, E, and F) for the creation of mutant libraries by saturation mutagenesis.[9] Asp192, the catalytically active residue, undergoes rate- and stereoselectivity-determining S_N2-reaction of the bound substrate *rac*-**1**, the epoxide being activated by two tyrosines by H-bonding. Rapid hydrolysis of the short-lived enzyme–ester intermediate completes the catalytic cycle.

In a previous study we addressed the question of how many pathways lead from WT ANEH to the particular target mutant LW202 by considering all permutations of the five sets of mutations, without performing new saturation mutagenesis experiments.[29] This enabled the construction of a constrained "fitness-pathway landscape" of the 120 pathways from WT to LW202 as determined by experimental free energy values ($\Delta\Delta G^{\neq}$). The analysis of epistatic interactions along each trajectory is also possible. Accordingly, 55 pathways proved to be energetically favorable (absence of local energy dips along a given trajectory). Most importantly in this study, strong cooperative effects (epistasis) between sets of mutations were identified as contributing to the efficacy of ISM. In nature, new amino acid exchange events cannot occur; in contrast, the different permutations of the five sets of accumulated mutations lead to new mutants but not to new point mutations, thereby forming a constrained fitness-pathway landscape.[29]

Such a systematic deconvolution process does not address a very different question that is fundamentally more important: how many pathways in a given ISM scheme lead to notably improved enzyme variants? This would involve an unconstrained fitness-pathway landscape, because new mutations are possible at all evolutionary stages in each trajectory. Some pathways can be expected to involve libraries devoid of improved mutants. Such "local minima" or "dead ends" can have different causes, including 1) actual dips in the landscape as dictated by the corresponding protein sequence space, 2) insufficient oversampling, and 3) amino acid bias due to the degeneracy of the genetic code and/or imperfect PCR conditions, among other factors.

Here we describe the exploration of a complete 24-trajectory fitness-pathway landscape by using an unconstrained ISM scheme based on the enzymatic transformation *rac*-**1**→(*S*)-**2** (Scheme 1)employed previously in the constrained system. The present study led to new perspectives regarding efficient laboratory-based evolution as a means to generate selective biocatalysts for organic chemistry, and also to conclusions concerning the use of protein engineering as a tool in studying Darwinian evolution in nature.

Results and Discussion

Experimental platform

For practical reasons we decided not to consider all five ANEH randomization sites of the original report,[9] because exploration of 120 ISM pathways would entail excessive experimental work. Rather, we designed a four-site system, thereby requiring testing of only 24 pathways (Figure 1); we considered sites B, D, E, and F at the binding pocket of WT ANEH (Figure 2). In order to reduce the screening effort further, two additional measures were taken. Firstly, the three-residue sites B (215/217/219) and F (244/245/249) were truncated to two-residue sites B* (215/219) and F* (244/249), respectively. Secondly, a reduced amino acid alphabet based on NDT codon degeneracy was chosen (only 12 amino acids: Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, and Gly). For 95% library coverage, a two-residue site requires the screening of only 430 transformants (assuming absence of amino acid bias).[13] The amino acid selection includes a balanced set of building blocks, with polar, nonpolar, aromatic, lipophilic, and charged and uncharged side chains. This experimental platform was used throughout the present study.

Saturation mutagenesis experiments

The results of screening the initial four first-generation randomization libraries are summarized in Table 1; in each case, the best variant is shown. Here and in all further NDT libraries, about 430 clones were screened per site, thereby ensuring 95% library coverage. Noteworthy at this stage is the relatively high enantioselectivity of variant GUY-107 ($E=23$) from saturation mutagenesis at site B*. In our original ISM study, which used the same enzyme and the same reaction, we used NNK codon degeneracy at non-truncated site B, and this led to the best mutant Leu215Phe/Ala217Asn/Arg219Ser in that particular library with an E value of only 14.[9] The reason why the superior variant GUY-107 was missed in the earlier study despite theoretically higher structural diversity was probably insufficient sampling. Only about 2200 transformants at site B in the original library were screened at the time, much less than the calculated 98000 colonies required for 95% library coverage of a three-residue site.[10,13]

Table 1. Results of the initial round of saturation mutagenesis using NDT codon degeneracy encoding 12 amino acids. E values were determined from the ANEH-catalyzed hydrolytic kinetic resolution of *rac*-**1**.

Site	Variant	E	Mutations
B*	GUY-107	23	Leu215Phe/Arg219Asp
D	GUY-111	10	Leu349Val/Cys350His
E	GUY-113	21	Thr317Phe/Thr318Ser
F*	GUY-116	12	Leu249Phe

These mutants were then used as templates in the 12 subsequent second-generation saturation mutagenesis experiments, and the ISM process was continued until every site had been considered once in each of the 24 pathways. When encountering libraries lacking any improved variants, which is equivalent to getting caught in a local minimum (dead end), conventional wisdom in directed evolution is to abandon such a pathway or to utilize an alternative mutagenesis method.[1] However, in these cases we decided to systematically test an unorthodox strategy by selecting non-improved and in some cases even inferior mutants (lower enantioselectivity) as templates in the continuation of the respective pathway.

The results of the systematic exploration of the complete ISM Scheme proved to be remarkable in several respects (for the complete data see the Supporting Information). All 24 pathways were found to provide notably improved enantioselectivity in favor of (*S*)-**2** relative to WT ANEH ($E=4.6$), with E values of at least 28 being achieved. Variants displaying enantioselectivity in the lower range ($E=28$–80) can be of practical use, although a higher degree of selectivity is desirable. Of the 24 pathways, 12 lead to variants reaching or surpassing $E=80$ (Figure 3 and the Supporting Information). Of these, GUY-228 (trajectory F*→B*→E→D) was found to be the most selective ($E=158$). This variant and several others surpassed the best mutant (LW202, $E=115$) of our original ISM study, which used NNK codon degeneracy in a five-step evolutionary process.[9] Of the other 12 pathways, eight were found to provide mutants with E values in the range 50 to 80, and only a small minority (four trajectories) led to variants in the lowest category ($E=28$–38; Table S1 and Scheme S1). In summary, all 24 pathways provided notably improved ANEH mutants, and 20 led to significant stereoselectivity ($E=50$–158).

Although insight into the influence of the mutations on stereoselectivity must await a detailed biophysical study, as was performed for the previous mutant (LW202),[14] we note that GUY-228 (B*(Leu215Phe/Arg219Val), F*(Leu249Phe), E(Thr317Phe/Thr318Cys), D(Leu349Asp/Cys350Tyr)) is significantly different to LW202 (B(Leu215Phe/Ala217Asn/Arg219Ser), C(Met329Pro/Leu330Tyr), D(Cys350Val), F(Leu249Tyr), E(Thr317Trp/Thr318Val)). Of the seven point mutations in GUY-228, only Leu215Phe is also found in LW202, which has nine point mutations. Moreover, Leu215Phe occurred in 23 of the 24 final mutants and is therefore important (Supporting Information), while in the exception (GUY-259, $E=127$) Leu215 was not mutated. The results show that many different point mutations induce high stereoselectivity, with several variants surpassing the previously reported mutant. The principal advantage of using reduced amino acid alphabets (for example, NDT codon degeneracy) is the drastically reduced screening effort necessary for 95% library coverage. For higher efficacy, it is better to employ maximal library coverage with lower structural diversity (NDT encoding) than maximal structural diversity but low library coverage (NNK).[13] Other structural trends concerning the 24 final variants became evident upon applying sequence relationships in the construction of a type of phylogenetic tree (Scheme S3).

Careful inspection of Figure 3 shows that in one of the pathways (E→B*→D→F*) a local minimum was encountered (the saturation mutagenesis library at site D). In this case the library did not harbor an improved mutant. The best of many inferior ones was then chosen, and upon continuing to site F*, this led

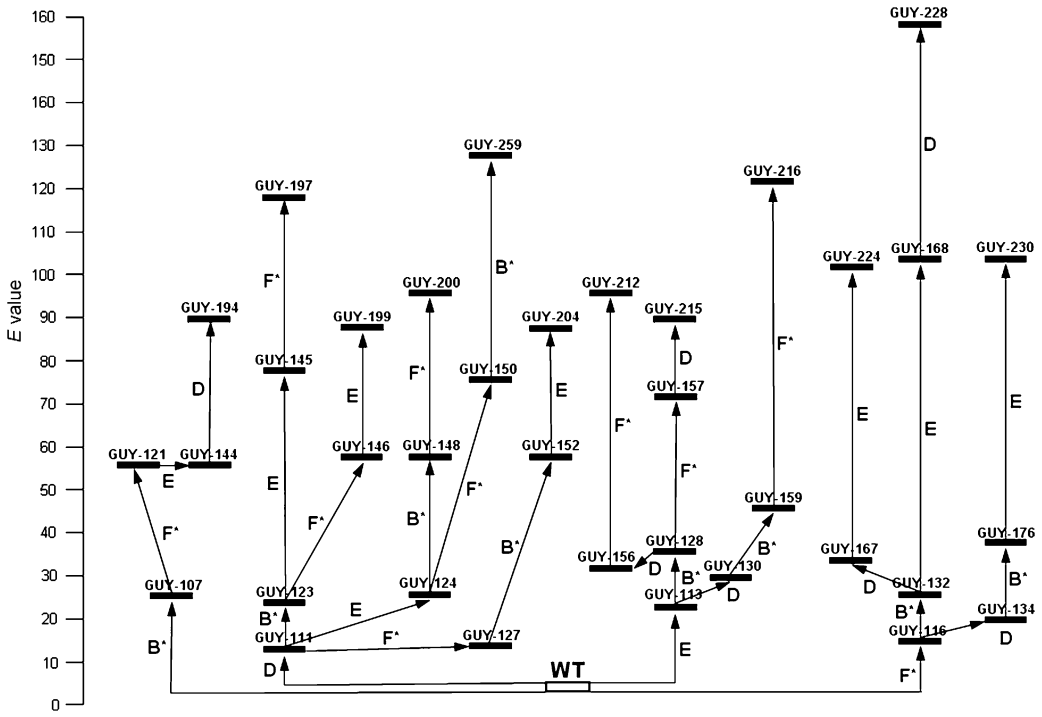


Figure 3. Portion of the 24-pathway ISM scheme showing the 12 pathways that led to ANEH variants displaying $E \geq 80$ in the hydrolytic kinetic resolution of *rac*-**1**.

to a notably improved mutant GUY-212 ($E = 98$; Figure 3). Similar escapes from local minima were achieved in several other cases, such as for the 12 pathways that led to E values in the range 28–78 (Scheme S1). An example is pathway B*→F*→D→E. In three cases the final library did not contain any hits with higher enantioselectivity, but the best mutant in the second to last library characterized by a higher E value was nevertheless identified and sequenced.

Construction of an unconstrained fitness-pathway landscape

The fitness landscape of such a system has a multi-dimensional surface involving all mutational sets at sites B*, C, D, and F* as independent vectors and $\Delta\Delta G^{\neq}$ as the dependent variable, which is difficult to depict graphically. We therefore constructed a "fitness-pathway landscape" by considering the experimental data at all stages of a given pathway in a stacking mode that links WT ANEH (top) with the final mutant in each case (bottom) for all 24 trajectories (Figures 4 and 5). The free energy profiles of the 24 evolutionary pathways have two different types of trajectory. The first is characterized by a smooth continuous decrease in $\Delta\Delta G^{\neq}$ in which each library along a pathway contains an improved (more stereoselective) variant, as for example in trajectory 20 (F*→B*→E→D; solid line in Figure 4). Of the 24 pathways, 16 show such a progressively energetically favored profile. In the second type of pathway at least one saturation mutagenesis library on the evolutionary trajectory lacks a mutant showing enhanced stereoselectivity, as for example in the case of E→F→B*→D* which can be designated as a "less-favored" pathway (dotted arrow in Figure 4).

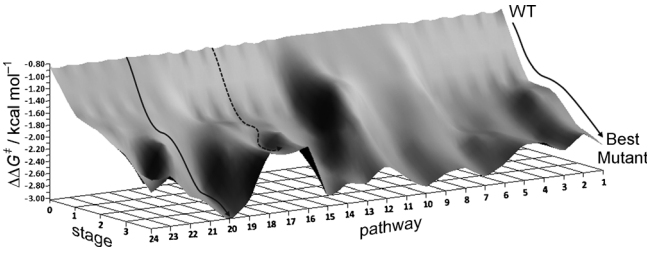


Figure 4. Fitness-pathway landscape featuring the 24 trajectories leading from WT ANEH to the final enantioselectivity at the end of each pathway as specified by the respective $\Delta\Delta G^{\neq}$ values. Solid line: typical pathway in which each mutant library contains a variant displaying improved enantioselectivity; dotted line: typical pathway in which at least one library is devoid of an improved variant.

In general, favored (solid) trajectories end up with ANEH mutants that display somewhat higher stereoselectivity than final mutants from less-favored (dotted) pathways. However, this does not mean that the dotted pathways (33% of the total) are failures. Such pathways do in fact lead to practical mutants, trajectory E→B*→D→F* being one of several outstanding examples. As noted above, this yielded variant GUY-212 ($E = 98$; Figure 3). The first derivative of the fitness-pathway landscape is in line with these conclusions (Scheme S2).

Conclusions

This study has helped to answer the following questions. 1) In an unconstrained system of laboratory-based evolution (e.g.,

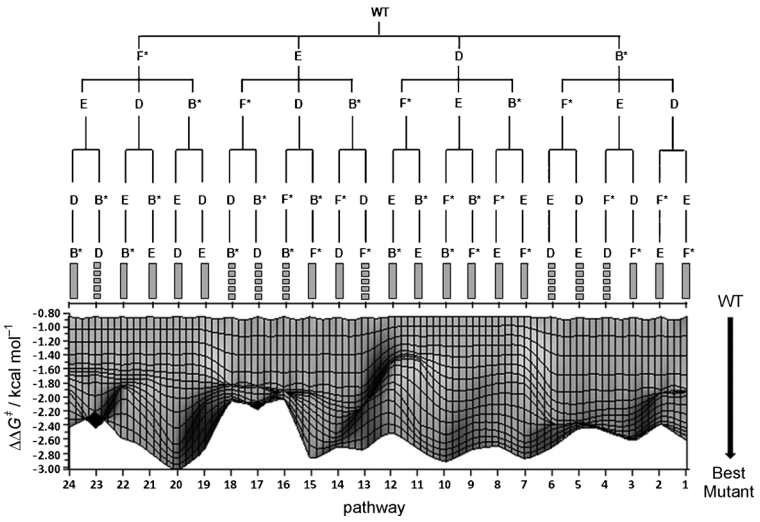


Figure 5. Free energy profiles of the 24 ISM pathways (front view of the fitness-pathway landscape). In the solid pathways all relevant saturation mutagenesis libraries contain improved variants (enhanced enantioselectivity) in the model reaction; the eight red pathways denote those in which at least one library in the four-step evolutionary process is devoid of any improved variants; three occurred in the final library, five in earlier libraries. In the latter, an inferior or non-improved mutant was used in the continuation of the ISM process, leading to a final variant with notably enhanced enantioselectivity.

ISM), how many of the theoretically possible pathways lead to improved mutants, and how does this relate to the efficacy of the method? 2) Is the use of non-improved or even inferior mutants selected in a given library devoid of hits for subsequent ISM rounds a viable strategy for escaping from local minima in the respective fitness-pathway landscape, and how does this relate to directed evolution in general? 3) How relevant are data from constrained and unconstrained laboratory evolution experiments for developing a deeper understanding of natural Darwinian evolution?

We begin with the first question. In the ISM system used here, all 24 pathways led to notably improved mutants as measured by the enhanced stereoselectivity of an epoxide hydrolase in a model reaction. This gives confidence that ISM is a valid method in the toolbox of directed evolution, to complement epPCR and DNA shuffling. This is not surprising, given the rational character of the structure-based approach. Focused libraries that target residues near the binding pocket would be expected to harbor a higher proportion of mutants that influence its shape and electrostatic character and therefore affect stereoselectivity, as was first demonstrated in 2001.[30] Some pathways are likely to be more productive than others. Not all trajectories in any new ISM system will provide highly improved mutants, but the probability of success by arbitrarily choosing a certain pathway without considering others can now be viewed as being reasonably high. If the end result of a given pathway provides an insufficiently improved mutant, then a different route should be explored. Alternatively, the gene of the final variant in a given pathway can be used as a template to "visit" the first site a second time (or any other one).[9] Since the conception of ISM five years ago, more than a dozen ISM-based directed evolution studies have proven to be successful in enhancing stereo- and/or regioselectivity, as well as thermostability; all of these involved arbitrarily chosen pathways, and all required only small libraries and thus minimal screening effort.[3,15–28] Although mutations at remote sites can influence stereoselectivity,[3,30–32] the probability of positive mutational effects is greatest at amino acid positions closer to the active center.[3,32] Indeed, the first case of saturation mutagenesis for increasing enantioselectivity involved a four-residue randomization site at the binding pocket of a lipase;[30] this later led to the acronym CAST. Moreover, when applying ISM, pronounced cooperative epistatic effects occur between sets of mutations as the evolutionary process proceeds (more than additivity) and between point mutations within a set.[3,29] In the present study, for practical reasons we utilized CAST sites with two amino acid positions and a 12-membered reduced amino acid alphabet (NDT codon degeneracy). As a possible alternative, extended sites (e.g., 4–8 residues) around the binding pocket of an enzyme could be used, in conjunction with highly reduced amino acid alphabets for reducing the screening effort.[33] Hilvert has reported that a reduced amino acid alphabet (only nine members) can be used to construct an enzyme with proper folding and activity,[34] so it can be concluded that restricting the number of building blocks in directed evolution to only a few appropriate amino acids yields a viable strategy, especially when combined with bioinformatic techniques.[33]

When applying ISM, or indeed any other directed evolution method, the generation of a mutant library devoid of any improved variants is normally interpreted as the occurrence of a local minimum.[1] Such libraries are traditionally discarded, thereby forcing the experimenter to devise alternative strategies for altering protein sequence, or to abandon the task. In the present study we encountered several libraries of this type, but decided not to abandon them. By choosing non-improved mutants (or even inferior ones) as templates in subsequent saturation mutagenesis experiments, it was possible in all cases to escape from what is normally considered a dead-end. Recently, we report isolated examples of this behavior,[35a] although it was not clear how general this might be. The probabilistic analysis of saturation mutagenesis reported by Nov is also revealing.[35b] The present systematic exploration of this phenomenon lends support to the previous preliminary conclusion regarding the viability of this approach for escaping from a local minimum.[35a] This not only opens further perspectives for ISM, but for the field of directed evolution in general, independent of the gene mutagenesis method. What we observe bears some similarity to the neutral drift concept[36,37] and to the Eigen–Schuster postulate of quasi-species in natural evolution.[38] Mannervik refers to the quasi-species concept in other directed evolution studies, but inferior mutants were not used for subsequent mutagenesis.[39] Our finding that inferior mutants can be used successfully in subsequent mutagenesis experiments is reminiscent of Lenski's digital organisms postu-

late for "computer programs that self-replicate, mutate, compete and evolve", where "deleterious mutations… served as stepping-stones in the evolution of complex features".[40]

The complete exploration of all possible evolutionary pathways as described herein is the first of its kind, but from a practical point of view we do not suggest that such a systematic procedure needs be followed in all ISM-based laboratory evolution studies aimed at improving the catalytic parameters of enzymes. Rather, the exploration of a single pathway generally suffices, the order of the randomization sites being arbitrary.[3,15–28] If in a given ISM study a local minimum is encountered early on, it may well be better to test an alternative pathway rather than choosing an inferior mutant for continuing laboratory evolution. However, if the experimenter has already invested significant effort in reaching a notable but not yet fully satisfactory degree of catalyst improvement by going through several steps in an arbitrarily chosen pathway, then it is more practical to choose an inferior mutant as a stepping stone for continuing evolutionary optimization rather than abandoning the trajectory due to the local minimum and starting over from the WT.

Finally, we also conclude that it is possible to learn from the mechanisms of natural evolution when attempting to devise more effective directed evolution strategies, but that the reverse (using data from laboratory evolution in order to understand natural evolution) can be problematic, although not impossible if the experiments are devised carefully. The type of evolution that we practice in the laboratory cannot be equated with the Darwinian evolution that occurs in nature: our system is based on screening and not on selection in which the host organism has a growth (survival) advantage because it harbors mutants with increased stereoselectivity. To construct such selection systems is extremely difficult.[41]

These points need to be viewed in the light of a study by Weinreich et al., who concluded that "Darwinian evolution can follow only very few mutational paths to fitter proteins".[42] They used a selection system (not screening) involving a β-lactamase, which allowed bacterial resistance to an antibiotic (cefotaxime) to be determined experimentally, as this depends on enzyme activity. In doing so, a *constrained* experimental platform was devised in the following way.[42] Five point mutations of a β-lactamase variant were first considered, based on a DNA shuffling paper by Stemmer.[43] All permutations of these five point mutations enabled 120 pathways (5!=120) to be constructed that *linked WT enzyme to the specific final mutant*. The chance of transversing a pathway was inferred from the probability of fixation. Thus, new mutants were generated without introducing new point mutations. This procedure led to the conjecture that 102 of the 120 pathways "are inaccessible to Darwinian selection and that many of the remaining trajectories have negligible probabilities of realization",[42] thereby leading to this general conclusion regarding Darwinian evolution. However, natural evolution does not occur in a constrained system in which all mutations are predetermined and in which new mutations are not allowed to occur. Therefore, the number of pathways linking WT with a specified mutant of a defined sequence involves a constrained system, which is not an ideal model for studying Darwinian evolution. The number of mutational paths to fitter proteins is likely to depend on the mutagenesis method and on the particular model system investigated. Indeed, by using a related constrained system in an evolving bacterial population, the effects of epistasis on fitness for the first five mutations to fix in a population of *E. coli* were recently studied, leading Cooper et al. to conclude that of the 120 possible trajectories linking the ancestor with the most-fit genotype, 86 increase monotonically for fitness.[44] Likewise, a recent study by Salverda and co-workers showed that several rounds of epPCR in the β-lactamase/cefotaxime system provided active mutants in alternative pathways not previously observed.[45] Clearly, fundamental differences exist between constrained[42] and unconstrained systems.[45] Finally, we note that an unconstrained system based on the application of ISM (as in the present study) but applied to a selection system such as that for example the β-lactamase/cefotaxime setup, would constitute an alternative experimental platform in evolutionary biological studies.

Experimental Section

Library generation: Saturation mutagenesis libraries were created by using QuikChange (Stratagene). Each PCR reaction (50 μL) contained 10× KOD Buffer (5 μL; Novagen), dNTPs (2 mM each, 5 μL) $MgCl_2$ (25 mM, 2 μL), degenerate primers (2.5 μM, 2.5 μL), template plasmid (20–50 ng μL^{-1}, 0.5 μL), KOD Hot start polymerase (0.5 U). Instead of using the Pfu Turbo DNA polymerase normally used in the Stratagene kit, PCR was performed with KOD hotstart polymerase from Novagen due to its higher fidelity DNA copy. The thermal cycler programme was 95 °C 3 min, 20× (95 °C 30 s, 59 °C 1 min, 72 °C 7 min0, 72 °C 16 min. The PCR product was transformed in *E. coli* DH5α after two consecutive treatments with DpnI (1 U; New England Biolabs) at 37 °C for 2 h each. The libraries generated and ANEH mutants were analyzed by sequencing using primers pQE for (GTATC ACGAG GCCCT TTCGT CT) or pQE rev (CATTA CTGGA TCTAT CAACA GGAG; Eurofins MWG, Ebersberg, Germany).

Expression of gene library: Bacterial colonies were picked with a QPIX colony picker (Genetix, Sunnyvale, CA), inoculated into 96-well deep plates containing LB (800 μL per well, supplemented with 100 μg mL^{-1} carbenicilin), and incubated overnight (37 °C with shaking at 800 rpm). Overnight culture (100 μL) was transferred to a glycerol plate (96 wells, 100 μL glycerol (70%) in each well) and kept at −80 °C. Another 100 μL of this overnight culture was used to inoculate fresh medium and allowed to grow (12 h, 30 °C, 800 rpm).

Library screening: Cultures (90 μL) were transferred into the wells of a 96-well micro-plate, and 600 μL of a mixture comprising *rac*-**1** (625 mg in acetonitrile (25 mL; HPLC grade)) and sodium phosphate buffer (280 mL, 20 mM, pH 7.2) was added. After shaking (1 h, 30 °C), the reaction was stopped by centrifugation (4000 rpm, 20 min), and 150 μL of the reaction mixture was then transferred into a new reaction plate and mixed with the same volume of methanol (HPLC grade). This mixture was further centrifuged (4000 rpm, 30 min), and 200 μL of this solution was finally transferred to a 96-well microtiter plate to be measured analytically by HPLC. For reactions that exceeded 50% conversion, the reactions were repeated with a lesser amount of culture (45 μL). For confirmations of potential hits, an internal standard solution ((*R*)-(+)-1-phenyl-1-butanol (6.66 mM in methanol)) was used. The HPLC con-

ditions were: column, Kromasil 3-cellucoat RP 100, 4.6 mm I.D. (Akzo Nobel, Schweden); pre-column, Kromasil 5-Cellucoal 10, 4.6 mm I.D.; eluent, methanol/water (70:30); flow rate 1 mL min^{-1}. The retention times were: (*R*)-diol (**2**), 1.9 min; (*S*)-diol (**2**), 2.1 min; (*R*)-epoxide (**1**), 4.1 min; (*S*)-epoxide (**1**), 4.7 min. The possible influence of the mutations on the expression efficiency was not studied (the amount of ANEH in each well of the microtiter plates was not measured).

Acknowledgements

Support by the Arthur C. Cope Foundation (USA) and the Max Planck Society (Germany) is gratefully acknowledged.

Keywords: directed evolution · enantioselectivity · enzyme catalysis · fitness landscapes · iterative saturation mutagenesis

[1] Recent reviews of directed evolution: a) T. W. Johannes, H. Zhao, *Curr. Opin. Microbiol.* **2006**, *9*, 261 – 267; b) *Protein Engineering Handbook, Vols. 1 – 2* (Eds.: S. Lutz, U. T. Bornscheuer),Wiley-VCH, Weinheim, **2009**; c) N. J. Turner, *Nat. Chem. Biol.* **2009**, *5*, 567 – 573; d) C. Jäckel, P. Kast, D. Hilvert, *Annu. Rev. Biophys.* **2008**, *37*, 153 – 173; e) S. Bershtein, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2008**, *12*, 151 – 158; f) P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 866 – 876; g) M. T. Reetz in *Asymmetric Organic Synthesis with Enzymes* (Eds.: V. Gotor, I. Alfonso, E. García-Urdiales), Wiley-VCH, Weinheim, **2008**, pp. 21 – 63; h) L. G. Otten, F. Hollmann, I. W. C. E. Arends, *Trends Biotechnol.* **2010**, *28*, 46 – 54; i) A. V. Shivange, J. Marienhagen, H. Mundhada, A. Schenk, U. Schwaneberg, *Curr. Opin. Chem. Biol.* **2009**, *13*, 19 – 25; j) A. S. Bommarius, J. K. Blum, M. J. Abrahamson, *Curr. Opin. Chem. Biol.* **2011**, *15*, 194 – 200.

[2] Reviews covering protein engineering of thermostability: a) T. Oshima, *Curr. Opin. Struct. Biol.* **1994**, *4*, 623 – 628; b) C. Ó'Fágáin, *Enzyme Microb. Technol.* **2003**, *33*, 137 – 149; c) V. G. H. Eijsink, S. Gåseidnes, T. V. Borchert, B. van den Burg, *Biomol. Eng.* **2005**, *22*, 21 – 30; d) A. S. Bommarius, J. M. Broering, *Biocatal. Biotransform.* **2005**, *23*, 125 – 139; e) N. Amin, A. D. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti, V. Schellenberger, *Protein Eng. Des. Sel.* **2004**, *17*, 787 – 793; f) P. L. Wintrode, F. H. Arnold, *Adv. Protein Chem.* **2001**, *55*, 161 – 225.

[3] M. T. Reetz, *Angew. Chem.* **2011**, *123*, 144 – 182; *Angew. Chem. Int. Ed.* **2011**, *50*, 138 – 174.

[4] a) K. Faber, *Biotransformations in Organic Chemistry*, 5th ed., Springer, Berlin, **2004**; b) *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook, Vols. I – III* (Eds.: K. Drauz, H. Waldmann), 2nd ed., Wiley-VCH, Weinheim, **2002**; c) *Industrial Biotransformations* (Eds.: A. Liese, K. Seelbach, C. Wandrey), 2nd ed., Wiley-VCH, Weinheim, **2006**; d) J. Tao, G.-Q. Lin, A. Liese, *Biocatalysis for the Pharmaceutical Industry*, Wiley-VCH, Weinheim, **2009**; e) V. Gotor, I. Alfonso, E. García-Urdiales, *Asymmetric Organic Synthesis with Enzymes*, Wiley-VCH Weinheim, **2008**.

[5] S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291 – 297.

[6] M. R. Parikh, I. Matsumura, *J. Mol. Biol.* **2005**, *352*, 621 – 628.

[7] R. J. Kazlauskas, U. T. Bornscheuer, *Nat. Chem. Biol.* **2009**, *5*, 526 – 5297.

[8] R. J. Fox, G. W. Huisman, *Trends Biotechnol.* **2008**, *26*, 132 – 138.

[9] M. T. Reetz, L.-W. Wang, M. Bocola, *Angew. Chem.* **2006**, *118*, 1258 – 1263, *Erratum* 2556; *Angew. Chem. Int. Ed.* **2006**, *45*, 1236 – 1241, *Erratum* 2494.

[10] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891 – 903.

[11] 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.

[12] J. Zou, B. M. Hallberg, T. Bergfors, F. Oesch, M. Arand, S. L. Mowbray, T. A. Jones, *Structure* **2000**, *8*, 111 – 122.

[13] a) M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem* **2008**, *9*, 1797 – 1804; b) M. T. Reetz, D. Kahakeaw, J. Sanchis, *Mol. BioSyst.* **2009**, *5*, 115 – 122.

[14] M. T. Reetz, M. Bocola, L.-W. Wang, J. Sanchis, A. Cronin, M. Arand, J. Zou, A. Archelas, A.-L. Bottalla, A. Naworyta, S. L. Mowbray, *J. Am. Chem. Soc.* **2009**, *131*, 7334 – 7343.

[15] K. Engström, J. Nyhlén, A. G. Sandström, J.-E. Bäckvall, *J. Am. Chem. Soc.* **2010**, *132*, 7038 – 7042.

[16] K. Okrasa, C. Levy, M. Wilding, M. Goodall, N. Baudendistel, B. Hauer, D. Leys, J. Micklefield, *Angew. Chem.* **2009**, *121*, 7827 – 7830; *Angew. Chem. Int. Ed.* **2009**, *48*, 7691 – 7694.

[17] R. Hawwa, S. D. Larsen, K. Ratia, A. D. Mesecar, *J. Mol. Biol.* **2009**, *393*, 36 – 57.

[18] L. Liang, J. Zhang, Z. Lin, *Microb. Cell Fact.* **2007**, *6*, 36.

[19] M. R. M. De Groeve, M. De Baere, L. Hoflack, T. Desmet, E. J. Vandamme, W. Soetaert, *Protein Eng. Des. Sel.* **2009**, *22*, 393 – 399.

[20] M.-W. Bhuiya, C.-J. Liu, *J. Biol. Chem.* **2010**, *285*, 277 – 285.

[21] R. M. Kelly, H. Leemhuis, L. Dijkhuizen, *Biochemistry* **2007**, *46*, 11216 – 11222.

[22] M. Ivancic, G. Valinger, K. Gruber, H. Schwab, *J. Biotechnol.* **2007**, *129*, 109 – 122.

[23] D. F. Visser, F. Hennessy, J. Rashamuse, B. Pletschke, D. Brady, *J. Mol. Catal. B* **2011**, *68*, 279 – 285.

[24] D. J. Bougioukou, S. Kille, A. Taglieber, M. T. Reetz, *Adv. Synth. Catal.* **2009**, *351*, 3287 – 3305.

[25] M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.* **2010**, *132*, 9144 – 9152.

[26] H. Zheng, M. T. Reetz, *J. Am. Chem. Soc.* **2010**, *132*, 15744 – 15751.

[27] S. Prasad, M. Bocola, M. T. Reetz, *ChemPhysChem* **2011**, *12*, 1550 – 1557.

[28] a) Y. Miyauchi, R. Kourist, D. Uemura, K. Miyamoto, *Chem. Commun.* **2011**, *47*, 7503 – 7505; b) M. A. Norrgård, B. Mannervik, *J. Mol. Biol.* **2011**, *412*, 111 – 120.

[29] M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260 – 2267.

[30] M. T. Reetz, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem.* **2001**, *113*, 3701 – 3703; *Angew. Chem. Int. Ed.* **2001**, *40*, 3589 – 3591.

[31] 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.

[32] S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45 – 54.

[33] a) M. T. Reetz, S. Wu, *Chem. Commun.* **2008**, 5499 – 5501; b) A. G. Sandström, Y. Wikmark, K. Engstöm, J. Nyhlén, J.-E. Bäckvall, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 78 – 83.

[34] K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem.* **2005**, *280*, 37742 – 37746.

[35] a) Y. Gumulya, M. T. Reetz, *ChemBioChem* **2011**, *12*, 2502 – 2510; b) for a probabilistic look at saturation mutagenesis, see: Y. Nov, *Appl. Environ. Microbiol.* **2012**, *78*, 258 – 262.

[36] S. G. Peisajovich, D. S. Tawfik, *Nat. Methods* **2007**, *4*, 991 – 994.

[37] M. A. DePristo, *HFSP J.* **2007**, *1*, 94 – 98.

[38] M. Eigen, J. McCaskill, P. Schuster, *J. Phys. Chem.* **1988**, *92*, 6881 – 6891.

[39] a) L. O. Emrén, S. Kurtovic, A. Runarsdottir, A.-K. Larsson, B. Mannervik, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10866 – 10870; b) S. Kurtovic, B. Mannervik, *Biochemistry* **2009**, *48*, 9330 – 9339.

[40] R. E. Lenski, C. Ofria, R. T. Pennock, C. Adami, *Nature* **2003**, *423*, 139 – 144.

[41] a) M. T. Reetz, H. Höbenreich, P. Soni, L. Fernández, *Chem. Commun.* **2008**, 5502 – 5504; b) 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.

[42] 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.

[43] W. P. C. Stemmer, *Nature* **1994**, *370*, 389 – 391.

[44] A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science* **2011**, *332*, 1193 – 1196.

[45] M. L. M. Salverda, E. Dellus, F. A. Gorter, A. J. M. Debets, J. van der Oost, R. F. Hoekstra, D. S. Tawfik, J. A. G. M. de Visser, *PLoS Genet.* **2011**, *7*, e1001321.

Received: December 20, 2011
Published online on April 20, 2012

