M. T. Reetz

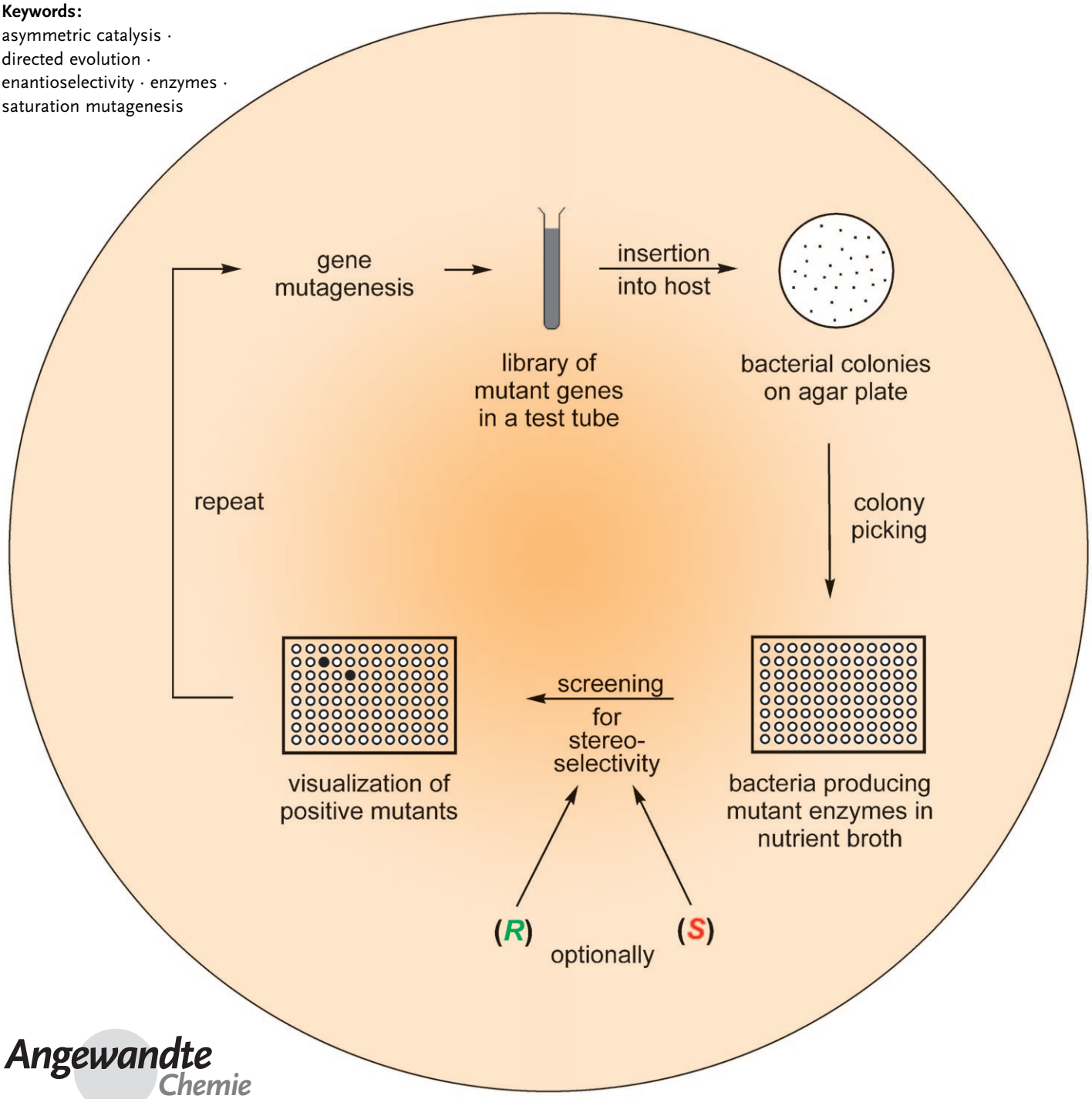
DOI: 10.1002/anie.201000826

Protein Engineering

Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions

*Manfred T. Reetz**

Keywords:
asymmetric catalysis · directed evolution · enantioselectivity · enzymes · saturation mutagenesis



Asymmetric catalysis plays a key role in modern synthetic organic chemistry, with synthetic catalysts and enzymes being the two available options. During the latter part of the last century the use of enzymes in organic chemistry and biotechnology experienced a period of rapid growth. However, these biocatalysts have traditionally suffered from several limitations, including in many cases limited substrate scope, poor enantioselectivity, insufficient stability, and sometimes product inhibition. During the last 15 years, the genetic technique of directed evolution has been developed to such an extent that all of these long-standing problems can be addressed and solved. It is based on repeated cycles of gene mutagenesis, expression, and screening (or selection). This Review focuses on the directed evolution of enantioselective enzymes, which constitutes a fundamentally new approach to asymmetric catalysis. Emphasis is placed on the development of methods to make laboratory evolution faster and more efficient, thus providing chemists and biotechnologists with a rich and non-ending source of robust and selective catalysts for a variety of useful applications.

From the Contents

1. Introduction

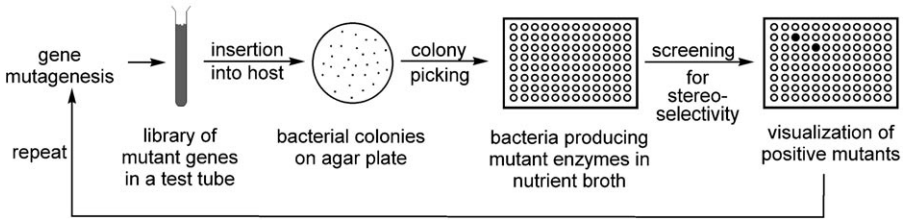
Methodology development in synthetic organic chemistry provides chemists with an ever-expanding toolbox of reagents, catalysts, and techniques, and thus contributes to the economic and ecological viability of the field. Asymmetric catalysis plays a key role in this "green" endeavor, with chiral synthetic catalysts[1] and enzymes[2] being the two most important options from which to choose. It is currently impossible to provide general recommendations, because the choice depends upon the particular problem at hand. For example, it makes a difference whether a catalyst is employed for small-scale applications, or whether it is used in an industrial process. Enzymes cannot catalyze the myriad transformations that transition-metal catalysts make possible, such as, for example, in the synthesis of complex natural products![3] Instead, they are restricted to the major enzyme types, namely oxidoreductases (reduction or oxidation), transferases (transfer of aldehydic, ketonic, acyl, sugar, phosphoryl, or methyl groups), hydrolases (hydrolysis or formation of esters, amides, lactones, lactams, epoxides, nitriles, anhydrides, or glycosides), lyases (addition/elimination of small molecules), isomerases (isomerizations, for example, racemization or epimerization), and ligases (formation of C–C, C–O, C–S, or C–N bonds). This is all that nature needs for the production of simple and highly complex molecules! Can this ever be achieved by chemists using only enzymes in vitro?

The application of enzymes in synthetic organic chemistry and biotechnology has traditionally suffered from the following limitations:[2] 1) A given compound of interest is not accepted by the enzyme because of limited substrate scope. 2) The enzyme's activity is sufficient, but stereoselectivity is poor. 3) The enzyme is too unstable to perform properly under operating conditions. 4) Product inhibition prevents high turnover.

During the last 10–15 years the genetic technique of directed evolution has been developed and refined to such an extent that it is now possible to address all of these problems.[4] Directed evolution, or laboratory evolution as it is sometimes called ("evolution in the test tube"), strives to imitate natural evolution by undergoing repeated cycles of gene mutagenesis, expression, and screening (or selection) until the desired degree of biocatalyst improvement has been achieved. The logic behind this concept is quite different from traditional forms of protein engineering that utilize site-directed mutagenesis based on "rational design", which is sometimes successful but far from general.[5] In the early days of directed evolution the goal was usually to increase the stability of enzymes.[6,7] Shortly thereafter, my research group of organic chemists became interested in the directed evolution of a very different catalytic property, namely enantioselectivity.[8] In our approach, the gene of a stereorandom or poorly enantioelective enzyme is first subjected to gene mutagenesis with formation of a gene library, which is subsequently inserted into a bacterial host such as *Escherichia coli*. The latter is then plated out on agar plates, and following expression, the transformants are harvested. This is followed

[*] Prof. Dr. M. T. Reetz
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr (Germany)
Fax: (+49) 208-306-2985
E-mail: reetz@mpi-muelheim.mpg.de
Homepage: http://www.mpi-muelheim.mpg.de/mpikofo_home.html

by a screening procedure which evaluates the enantioselectivity of the enzyme mutants as catalysts in a given transformation. The gene of an improved mutant is then used as the starting point (template) for another cycle, and the evolutionary process is continued until the desired degree of stereoselectivity has been achieved (Scheme 1).[8,9] It is the reliance on evolutionary pressure which makes the overall protocol "rational", and is quite different from the development of synthetic chiral catalysts, which require consideration of steric and electronic effects and/or utilize combinatorial methods.[1]



Scheme 1. Individual steps in the laboratory evolution of stereoselective enzymes.[8,9]

The challenges in putting the Darwinian approach to asymmetric catalysis into practice are twofold. Firstly, we had to develop high-throughput enantioselectivity assays which did not exist at the time.[8,9] Secondly, strategies had to be designed to probe protein sequence space efficiently; this in turn relates to the screening effort. To this day, the screening step, be it for stereoelectivity, activity, or thermostability, remains the bottleneck of directed evolution.[4,10] Consider an enzyme composed of X amino acids (residues). The number of enzyme variants (mutants) N at the theoretically maximum degree of diversity is described by the algorithm $N = 19^M X!/X-M)!M!$, where M denotes the total number of amino acid substitutions per enzyme molecule.[4] Application of this algorithm to an enzyme composed of 300 residues, for example, shows that 5700 mutants are possible if one amino acid is substituted randomly, but 16 million in the case of two simultaneous substitutions, and about 30 billion if three amino acids are exchanged simultaneously.

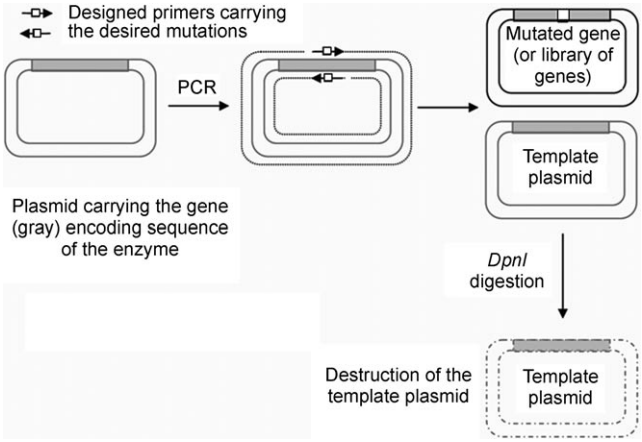
Originally, random mutagenesis methods were based on the use of chemicals, light, or mutator strains.[4] Today, the most often used gene mutagenesis method is the error-prone polymerase chain reaction (epPCR), in which the fidelity of the PCR process is perturbed by varying the concentrations of $MgCl_2$ (or $MnCl_2$) and/or of nucleotides.[4,11] The mutation rate can be controlled empirically. In early examples of protein eingineering[6] some sort of random gene mutagenesis method was used to enhance the thermostability of enzymes. However, with rare exceptions,[6f] only initial libraries were considered, which by itself does not (yet) constitute evolution. In a seminal study in 1993, Chen and Arnold applied several cycles of epPCR at low mutation rate to increase the robustness of the protease Subtilisin E toward the enzyme-damaging solvent dimethylformamide.[7] Since epPCR involves the whole gene, and thus targets the entire enzyme, it is a kind of "shotgun" method.[4] However, the degeneracy of the genetic code and other factors cause amino acid bias.[12]

A very different approach is DNA shuffling, which was devised by Stemmer;[13] this is a recombinant method for which different variations have been developed.[4] In general, one or more genes are first digested with a DNase to yield double-stranded oligonucleotide fragments of 10–50 base pairs, which are then amplified in a PCR process. Repeated cycles of strand separation and reannealing in the presence of a DNA polymerase, followed by final PCR amplification, result in the reassembly of full-length mutant genes. Family shuffling is a particularly efficient version, in which homologous genes from different species are chosen with formation of mutant libraries that are characterized by high catalyst diversity.[14]

Yet another commonly used mutagenesis method is saturation mutagenesis, sometimes also called cassette mutagenesis, combinatorial saturation mutagenesis, or site saturation mutagenesis.[4,15] It involves amino acid randomization at a predetermined position in the enzyme. This entails the introduction of all of the other 19 canonical amino acids with formation of focused libraries of mutants. Extension to the simultaneous randomization at two or more amino acid positions is also possible. Saturation mutagenesis is knowledge-based, since it requires structural, mechanistic, and/or bio-informatic data to make an appropriate choice regarding the putative randomization sites. In effect, it is a fusion of rational design and "blind" directed evolution. Since the mid-1980s many molecular biological variations of saturation mutagenesis have been reported, generally based on the use of appropriate oligonucleotides.[4,15] Degenerate primers need to be designed and prepared (commercially) which allow the genetic information encoding the desired mutational changes to be inserted into the plasmid. The currently most widely used procedure is the QuikChange protocol of Stratagene,[16]

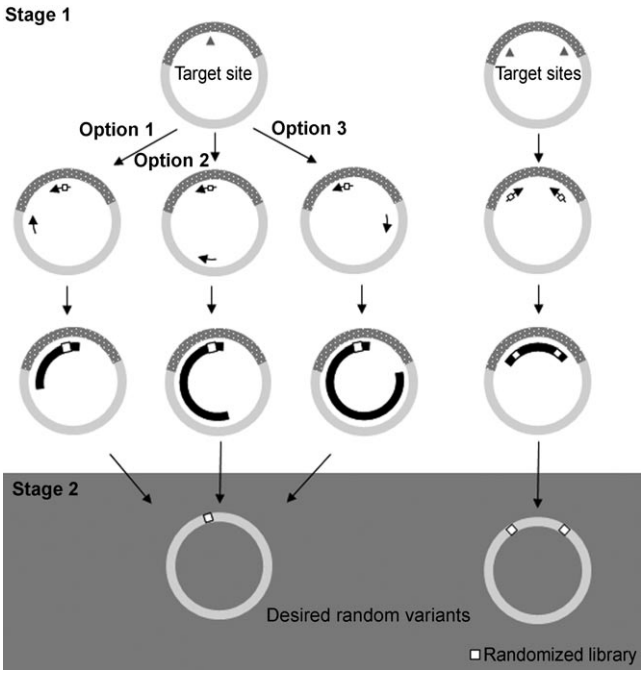
Manfred T. Reetz was born in 1943 in Germany and studied chemistry in the USA (BS from Washington University (St. Louis) in 1965 and a MS from the University of Michigan in 1967). He received his PhD in 1969 from the University of Göttingen (Germany) with U. Schöllkopf. Following postdoctoral research with R. W. Hoffmann in Marburg, he obtained his Habilitation there in 1974. After two years as Associate Professor at the University of Bonn, he became Full Professor in Marburg in 1980. In 1991 he joined the Max-Planck-Institut für Kohlenforschung in Mülheim an der Ruhr and was Managing Director 1993–2002, during which he created five Departments. As Director of the Department of Synthetic Organic Chemistry his interests focus on directed evolution of stereoselective enzymes and combinatorial transition-metal catalysis.

which is based on earlier studies.[4,15] It was originally developed for traditional site-specific substitution of single amino acids, but can also be employed for simultaneous randomization at up to five amino acid positions. The protocol is summarized in Scheme 2. Most often, NNK codon degeneracy is chosen (N: adenine/cytosine/guanine/thymine; K: guanine/thymine), which encodes all 20 canonical amino acids. Other codon degeneracies can also be chosen, specifically when utilizing reduced amino acid alphabets (Section 3.3).



Scheme 2. Steps required in saturation mutagenesis according to the QuikChange protocol.[16]

Recently, we noted problems with this protocol in certain cases.[17] Simultaneous randomization at two distal sites proved to be difficult, if not impossible, and large plasmids also caused problems. To handle such difficult-to-amplify templates we modified and extended previous improvements[15d–f] based on non-overlapping oligonucleotides. This resulted in the emergence of the two-stage process shown in Scheme 3.[17] In the first stage, both the mutagenic primer and the antiprimer, which are not complementary, anneal to the template. The amplified sequence is then used in the second stage as a megaprimer. In this straightforward process, sites composed of one or more residues can be randomized in a single PCR reaction, irrespective of their location in the gene sequence. In a comparative study,[17] the virtues of the new method relative to QuikChange and related protocols were demonstrated by using four different enzymes (lipases from *Candida antarctica* and *Pseudomonas aeruginos*a, epoxide hydrolase from *Aspergillus niger*, and the P450 enzyme BM3).[17] We have also developed a second approach and applied it in the evolution of stereoselective mutants of the enoate reductase YqjM[18] (Section 3.4). This approach is based on the previously reported sequence- and ligation-independent cloning (SLIC) method devised by Li and Elledge.[19] A set of primers encompassing the desired degenerated codons is used to amplify a part of the YqjM gene by using pET21a-YqjM as a template. The rest of the gene and the vector are then amplified with a second set of primers that have about 30 nucleotides identity with the first pair. After elimination of the template, the two PCR products



Scheme 3. Saturation mutagenesis method for difficult-to-amplify templates.[17] The gene is represented by the dotted section, the vector backbone is shown in light gray, and the formed megaprimer in black. In the first stage of the PCR, both the mutagenic primer (white squares represent positions randomized) and the antiprimer (or another mutagenic primer, shown to the right) anneal to the template, and the amplified sequence is used as a megaprimer in the second stage. Finally, the template plasmids are digested with *Dpn*I and the resulting library is transformed in a bacterial host. The scheme on the left illustrates the three possible options in the choice of the megaprimer size for a single-site randomization experiment. The scheme on the right represents an experiment with two sites simultaneously randomized.

are treated separately with T4 DNA polymerase. The vector and the pool of degenerated inserts are subsequently remixed and incubated in *E. coli* RecA, which facilitates the assembly of the homologous single strands in vitro. This is followed by transformation in the *E. coli* DH5α strain.[18]

The size of a library may vary between a few hundred and a few million, depending on the laboratory effort invested.[4] Clearly the high end of the scale poses severe screening problems,[10] especially if enantioselectivity needs to be evaluated. In principle, selection systems can handle very large numbers (10^6–10^7),[4] but they have to be properly designed so that the host organism experiences a growth advantage, since it harbors an enzyme with an improved catalytic profile. Unfortunately, it is not a trivial task to develop selections systems suitable for the directed evolution of enantioselective enzymes, and general solutions to this challenge have not yet been presented.[21]

The currently known *ee*-screening systems developed by us and others have been reviewed elsewhere,[9c,10] and for illustrative purposes only the Mülheim mass spectrometry (MS) based assay is mentioned here, which in favorable cases allows about 5000 *ee* determinations per day.[22] Since enantiomers have identical mass, the relative amounts of

the *R* and *S* forms present in a given sample cannot be measured by conventional MS techniques. However, if the substrates are isotopically labeled, then differences in the mass spectra of the pseudoenantiomers are easily detected. The method is limited to desymmetrization reactions of *meso* compounds with reactive enantiotopic groups and to kinetic resolutions. However, the technique has to be optimized for each new substrate and product, and it requires an expensive multiplexed MS instrument. Thus far, truly universal *ee* screening methods have not been developed, which means that research needs to be continued in this challenging field. Medium-throughput, meaning the *ee* screening of 300–800 samples per day, is possible by using automated GC[23a] or HPLC.[23b] This may suffice in some cases, specifically when high-quality focused libraries of mutants are generated (Section 3). A prescreen for activity, preferably on-plate, is of great advantage.[10b]

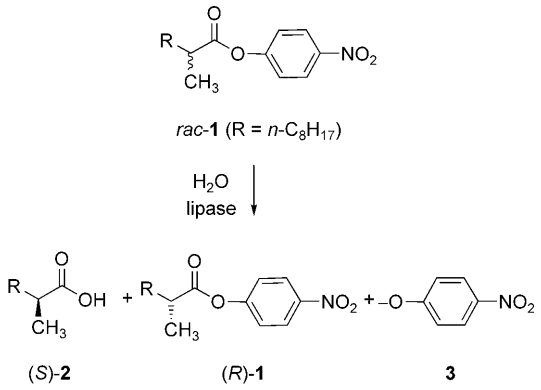
A very different way to handle the crucial problem of high-throughput screening relates to pooling, which entails the simultaneous analysis of multiple samples followed by deconvoluting those pools that signal the presence of improved mutants. One intriguing approach is the Bommarius system based on a Monte Carlo simulation model which allows the testing of various scenarios prior to experimentation.[24] In general, it requires a sensitive test for activity. This interesting technique is most likely to be successful when the background level of activity is minimized, so that it is easier to detect small increases in rate. An alternative pooling strategy was developed recently[18] (Section 3.4.1).

This Review emphasizes the importance of methodology development in the quest to make directed evolution more efficient than in the past. The rapid development of stereoselective (bio)catalysts is not only important for basic research, industry also requires catalytic processes within a defined timespan.

2. Early Examples of Directed Evolution of Enantioselective Enzymes

In an initial proof-of-principle study we focused on enhancing the enantioselectivity of the lipase from *Pseudomonas aeruginosa* (PAL) as a catalyst for the hydrolytic kinetic resolution of ester *rac*-**1**.[8] Wild-type (WT) PAL shows a selectivity factor of only $E = 1.1$ in slight favor of (*S*)-**2** (Scheme 4). After four cycles of epPCR at a low mutation rate, a mutant with four point mutations (Val47Gly/Ser149Gly/Ser155Leu/Phe259Leu) and a selectivity factor of $E = 11.3$ was identified.

Since a fifth round of epPCR improved enantioselectivity only marginally ($E = 13$), it became clear that better strategies had to be developed. We therefore spent several years testing various approaches, including saturation mutagenesis experiments at the four hot spots[25] identified earlier by the epPCR process.[8] In some cases this procedure provided improved mutants, for example, when randomizing position 155 remote from the active site ($E = 20$), but in others no hits could be detected. The idea of focusing on hot spots identified by epPCR was independently suggested by Miyazaki and Arnold



Scheme 4. Hydrolytic kinetic resolution of *rac*-**1** catalyzed by PAL mutants.[8]

in a study on thermostabilization,[26] and has since been used many times in directed evolution.[4] The problem with this strategy is the fact that epPCR generally leads to the accumulation of some superfluous point mutations, which means that saturation mutagenesis may not result in a positive response.

In a seminal experiment, a focused library was generated by saturation mutagenesis at a site consisting of positions 160–163 (Figure 1). This led to the identification of the mutant

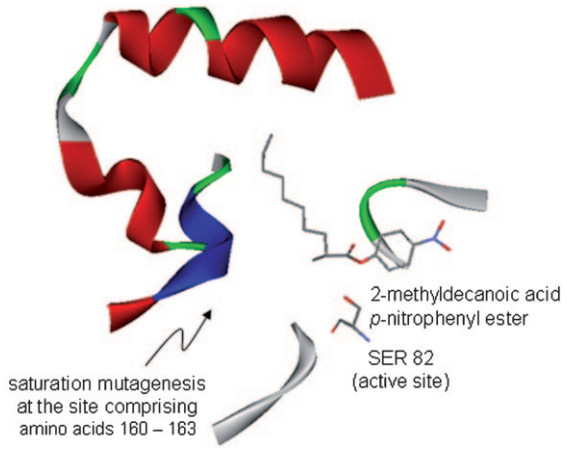


Figure 1. Binding pocket of PAL[27] for the acid part of *rac*-**1** showing the geometric position of amino acids 160–163, which were randomized simultaneously by saturation mutagenesis to enhance the enantioselectivity.[28] Ser82, as part of the usual catalytic triad Asp/His/Ser, attacks the carbonyl function nucleophilically with rate- and stereochemistry-determining formation of a short-lived oxyanion.

Glu160Ala/Ser161Asp/Leu162Gly/Asn163Phe with an *E* value of 30 after the screening of 5000 transformants.[28] This is the first case of saturation mutagenesis at a site aligning the binding pocket of an enzyme with the purpose of enhancing the enantioselectivity. The positive result indicated to us at the time that mutations near the active center may have a greater influence on the stereoselectivity than mutations at remote sites.[28] Our proposal was later corroborated by statistical analysis by Kazlauskas and co-workers,[29]

and has now become a generally accepted hypothesis as also suggested by Dalby and co-workers.[30]

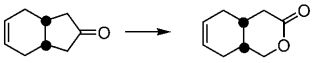
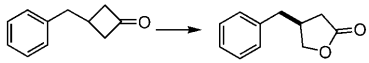
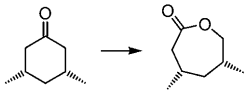
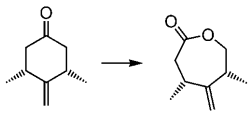
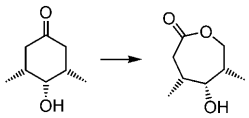
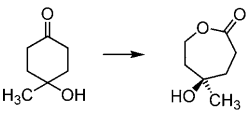
In subsequent studies we also returned to positions 155/162, whose randomization provided two mutants with *E* values of 30 and 34, respectively.[28] These and other early exploratory experiments in our laboratory suggested that randomization at one residue followed by saturation mutagenesis (or epPCR) at another position could constitute a useful way to probe protein sequence space.[25,28] However, we did not systemize this approach until a few years later with the development of iterative saturation mutagenesis (ISM; Section 3),[31] which has proven to be far more successful than originally anticipated.

The best PAL mutant ($E = 51$) was obtained by applying a strategy comprising epPCR with a high error-rate and DNA shuffling with simultaneous randomization again at positions 155/162.[28] This *S*-selective mutant is characterized by six point mutations, only one (Leu162Gly) of which is near the binding pocket. This came as a surprise if enantioselectivity is associated with Emil Fischer's lock and key principle. A QM/MM study not only unveiled that the source of enhanced enantioselectivity is caused by a relay mechanism, it also predicted that only two of the six point mutations are actually necessary, namely Ser53Pro and Leu162Gly.[32] The double mutant Ser53Pro/Leu162Gly was subsequently prepared by site-directed mutagenesis and found to be even more enantioselective ($E = 63$ in favor of (*S*)-**2**)! This was a triumph of theory, but the fact that we had accumulated four superfluous mutations clearly demonstrated that the chosen strategy was far from efficient.[32b] As already pointed out, superfluous mutations are not so unusual in directed evolution, although this facet is rarely illuminated.[4] The accumulation of such mutations means unnecessary laboratory work, especially with regard to screening. The total effort in obtaining the best mutant involved the screening of more than 50000 transformants.[4f,28] In a similar endeavor, we reported the inversion of enzyme enantioselectivity induced by directed evolution. *R*-selective PAL mutants with *E* values of 4–5 were obtained,[9b] which was later improved to $E = 30$.[33]

By using the early strategies outlined above, we then turned to other enzymes, with monoxygenases for enantioselective Baeyer–Villiger (BV) reactions[34] and sulfoxidation of prochiral thioethers being the primary focus of interest.[23b] Biochemists have studied the mechanism of Baeyer–Villiger monooxygenases (BVMOs),[35] and organic chemists have applied them in numerous desymmetrization reactions of prochiral ketones and/or oxidative kinetic resolution of racemic substrates with formation of enantiomerically enriched esters or lactones.[36] The most popular BVMO is the cyclohexanone monooxygenase (CHMO) from *Acinetobacter* sp. NCIMB 9871. Similar to other BVMOs, it is flavin (FAD) dependent and contains an NADPH binding domain.[35,36] The reduced enzyme-bound FAD reacts with oxygen to form an alkylhydroperoxide. This then adds nucleophilically to the carbonyl function with formation of the Criegee intermediate, followed by the usual fragmentation/rearrangement with formation of the ester or lactone product. Reduction of the oxidized FAD by NADPH then regenerates the active flavin. At the time of our first directed evolution study of BVMOs,[34] not a single X-ray study of this type of enzyme had been published. Thus, we applied epPCR. We chose CHMO as the enzyme and studied the oxidative desymmetrization of a substrate that was known to display poor enantioselectivity, namely 4-hydroxycyclohexanone (9% *ee* in slight favor of the (*R*)-lactone). It was possible to evolve *R*- and *S*-selective mutants for this transformation (80–90% *ee*).[34]

The generation of *R*- and *S*-selective CHMO mutants at will is a further remarkable example of directed evolution. It would be at least as impressive if a given mutant would have a broad substrate scope. One of the mutants, Phe432Ser, was tested as a catalyst in the desymmetrization of a set of 4-substituted cyclohexanone derivatives (methyl, ethyl, methoxy, chloro, bromo, iodo), and in all cases highly enantioselective transformations were observed (95–99% *ee*).[34] The same mutant was later applied successfully in the oxidative desymmetrization of a number of other structurally diverse prochiral ketones, again without performing any additional mutagenesis/screening experiments (Table 1).[37] This result shows that the general credo of directed evolution, "you get what you screen for",[4] needs a corollary, namely "you may get more than what you screen for". Synthetic organic chemists, in particular, need enantio-

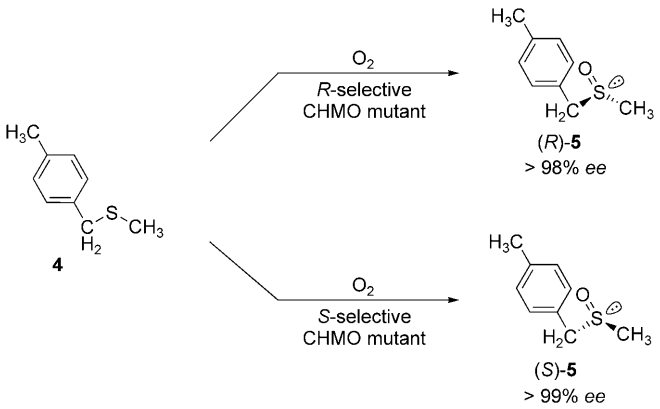
Table 1: Oxidative desymmetrization catalyzed by CHMO mutant 1K2-F5 (Phe432Ser) using air as the oxidant in a whole-cell process.[37]

Substrate	*ee* [%]
	94
	99
	91
	97
	78
	96
	>99
	>99
	>99
	99[a]

[a] Unpublished results of C. Clouthier, M. M. Kayser, and M. T. Reetz.

selective catalysts that are not just effective in a single transformation.

In a separate project, the asymmetric CHMO-catalyzed sulfoxidation of prochiral thioethers such as **4** with formation of chiral sulfoxides **5** was studied by applying the usual epPCR strategy (Scheme 5).[23b] WT CHMO shows an enantioselectivity of only 14% *ee* in favor of (*R*)-**5.** Several



Scheme 5. Enantioselective sulfoxidation catalyzed by CHMO mutant Phe432Ser in a whole cell process using air as the oxidant.[23b]

highly *R*- and *S*-selective mutants were identified (98–99% *ee*), thus demonstrating once more that directed evolution is capable of delivering mutants that reverse the stereoselectivity of an enzyme-catalyzed reaction. Interestingly, one of the highly improved *R*-selective variants (99% *ee*) proved to be the same mutant (Phe432Ser) that we had evolved earlier as a catalyst for asymmetric BV reactions! In some cases the asymmetric sulfoxidation was accompanied by an undesired overoxidation with formation of the corresponding sulfone, in the worst case in up to 20% yield. After performing an additional round of epPCR and screening for low overoxidation, less than 5% sulfone formation was observed while maintaining essentially complete enantioselectivity. This is the first example of directed evolution as a means to suppress an undesired side reaction.[23b]

Whole cells are generally used in these and in other BVMO-catalyzed transformations, thus bypassing the need to apply an in vitro NADPH regeneration system.[34–37] These enzymes are too sensitive to be handled in isolated form, and most organic chemists are not trained to handle whole cells. Consequently, we were attracted by the seminal study of Fraaije, Janssen et al., in which the discovery of the first thermostable BVMO, phenylacetone monooxygenase (PAMO), was announced.[38] This enzyme was characterized by X-ray crystallography by Malito, Mattevi et al.[39] We speculated that this discovery would open the door to in vitro application using isolated PAMA in conjunction with an NADPH regeneration system based on a robust alcohol dehydrogenase and isopropanol as the reductant. PAMO is indeed a robust monooxygenase, but unfortunately it accepts only phenylacetone and a few structurally related aryl acetone derivatives. For example, 2-phenylcyclohexanone reacts extremely slowly, delivering the respective lactone with poor enantioselectivity ($E = 1.5$) to the extent of less than 10% *ee* even after a prolonged reaction time of two days (activity only 7 U mg^{-1}), and substrates such as 2-alkylcyclohexanone derivatives are not at all accepted.[40,41] On the basis of the X-ray structure of PAMO, it was suggested that Arg337 stabilizes the Criegee intermediate in the binding pocket of the enzyme.[39] This provided us with a means to begin the interpretation of the enhanced and inverted enantioselectivity of the CHMO mutants, especially of variant Phe432Ser. After constructing a homology model of CHMO based on the X-ray data of PAMO[39] (Figure 2), we suspected that serine in mutant Phe432Ser undergoes a hydrogen-bonding interaction with Arg337, thereby leading to a reshaping of the binding pocket.[40] Moreover, an important structural difference between the two WT-BVMOs lies in the presence of a bulge, as part of a longer loop, next to the binding pocket of PAMO (positions 440–444), which is absent in CHMO. Since the bulge might be the cause of the low substrate scope of PAMO, we decided to apply rational design. We consequently performed several deletion experiments, thereby shortening the critical part of the loop. This approach was only partially successful, since the best rationally designed mutant accepted only 2-phenylcyclohexanone.[40] Later, we returned to this problem by utilizing more efficient strategies[41,42] (Section 4.2).

During the last decade, other research groups joined efforts to generalize the concept of the directed evolution of stereoselective enzymes for application in synthetic organic chemistry (Scheme 1) by using the strategies outlined above. These advances include the use of enzymes of the type esterase,[29,43] lipase,[44a–b] hydantoinase,[45] epoxide hydrolase,[46] nitrilase,[47] aldolase,[48] monoamine oxidase,[49] Baeyer–Villiger monooxygenase,[50] transaminase,[51] benzoylformate decarboxylase,[52] phosphotriesterase,[53] P450,[54] reductase,[55] oxynitrilase,[56] and horse radish peroxidase.[57] Circular permutation, that is, the intramolecular relocation of the C and N termini of a protein, has been applied to the evolution of enantioselectivity with the lipase *Candida antarctica* B (CALB).[44c] These contributions have been reviewed elsewhere[58] and only selected cases are highlighted here before turning to the main focus of this Review.

One important study carried out by scientists at Diversa (Verenium) concerns the directed evolution of a nitrilase as a catalyst for the desymmetrization of the prochiral nitrile **6** with formation of (*R*)-**7**,[47] which is an intermediate in the synthesis of the cholesterol-lowering therapeutic drug Lipitor (**8**; Scheme 6). The original patents, including the protection of synthetic organic pathways, are to expire soon, which means that a number of companies are poised to enter the generic market of this multiblockbuster (> $12 billion per year), probably by using more efficient synthetic schemes. Screening genomic libraries from environmental samples resulted in the discovery of about 200 new nitrilases showing high reaction rates and medium to very good enantioselectivity. One of them was found to catalyze the formation of the desired (*R*)-**7** with an enantioselectivity of 94.5% *ee*.[47] Unfortunately, when experiments were performed at a more industrially practical concentration of 2.25 M, the activity and

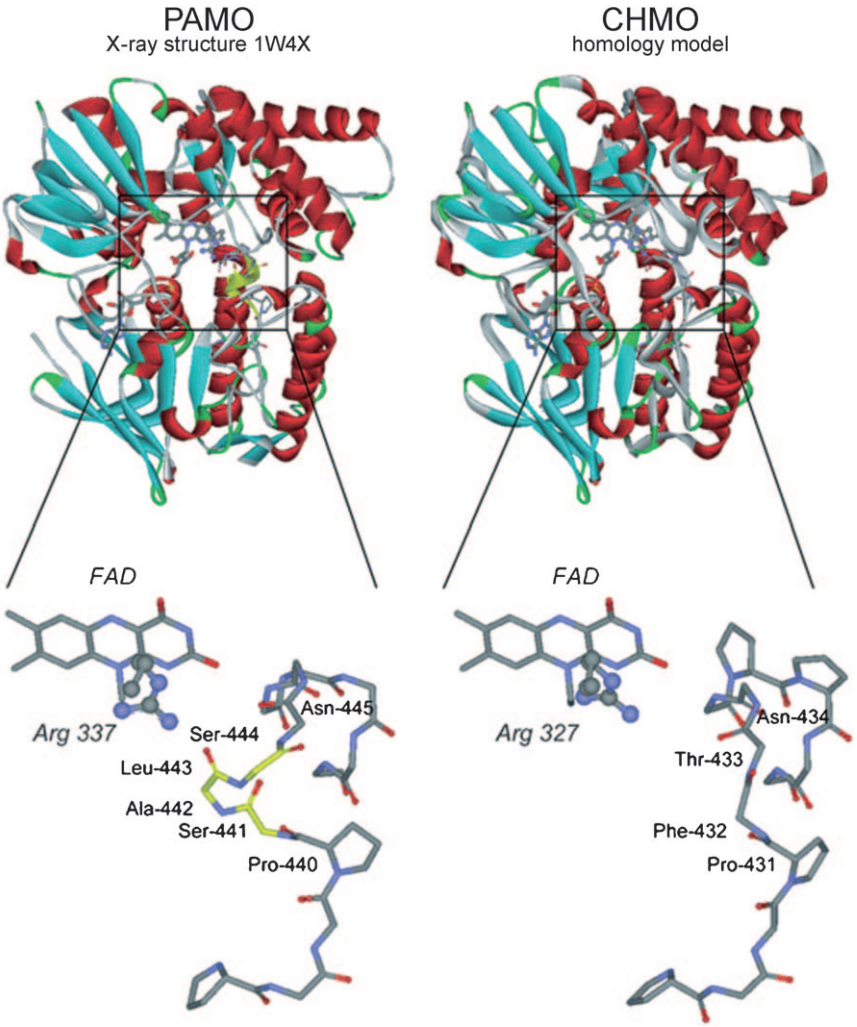
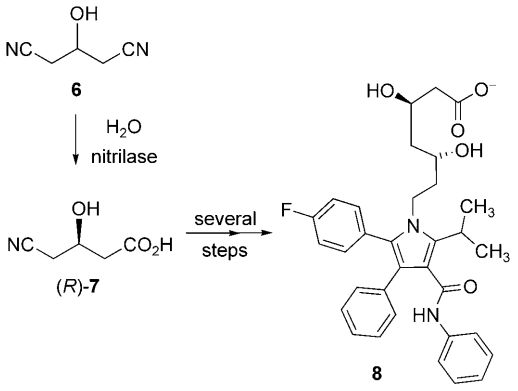


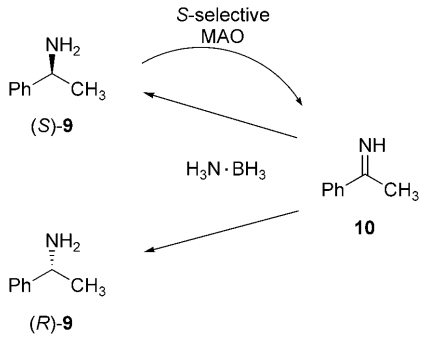
Figure 2. Comparison of the crystal structure of PAMO[39] (left) and the homology model of CHMO[40] (right) with 40.3% sequence identity. The upper part shows the overall fold, and the lower part is an enlargement of the active site showing the FAD cofactor as solid sticks and the catalytic arginine as a ball and stick model. The yellow color highlights the presence of two additional amino acids in the arginine-stabilizing loop of PAMO compared with CHMO.

enantioselectivity suffered (only 87.8% *ee*). This finding was a sign that product inhibition was occurring. Therefore, directed evolution was considered, in this case by applying saturation mutagenesis systematically one site at a time at all 330 amino acid positions in the enzyme. The total number of clones amounted to 31585, which were screened by applying the Mülheim MS-based screening system,[22] in this case based on ^{15}N-labeling at one of the nitrile functions.[47] About 17 hits were found to display enhanced enantioselectivity in favor of (*R*)-**7**. Some of them also showed improved performance at the higher concentration, the best one (Ala190His) being the most active and enantioselective catalyst under the operating conditions. Complete conversion was observed within 15 h at 98% *ee*. Even at substrate concentration of 3 M, conversion is 96% (98.5% *ee*). This example shows that product inhibition, which is sometimes a problem in biotechnology, can be suppressed completely by the technique of directed evolution, while increasing the enantioselectivity.[47]

Another impressive example of the directed evolution of enantioselective enzymes concerns the genetic manipulation of monoamine oxidases, which in nature catalyze the racemization of amino acids. Both *R*- and *S*-selective monoamine oxidases are known which selectively catalyze the reaction of either the *R* or *S* enantiomer. This property was exploited by Turner and co-workers in an ingeneous deracemization scheme with the formation of amines with extremely high enantioselectivity (Scheme 7).[49a] Accordingly, achiral reducing agents such as $NaBH_4$, $NaB(CN)H_3$, or H_3NBH_3 were employed in the presence of either an *R*- or *S*-selective monoamine oxidase. Directed evolution was applied since the



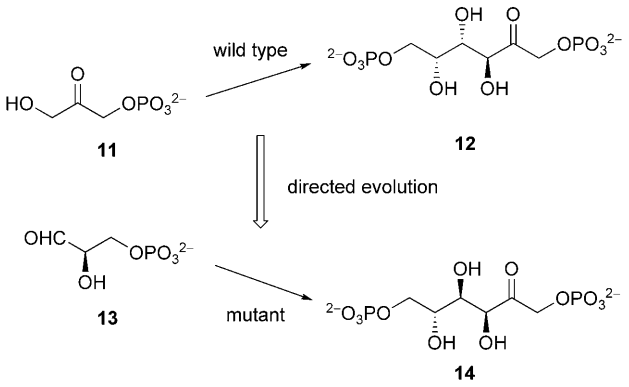
Scheme 6. Formation of the chiral compound (*R*)-**7** as an intermediate in the synthesis of the cholesterol-lowering therapeutic drug **8** (Lipitor), catalyzed by a mutant nitrilase.[47]



Scheme 7. Turner system for deracemizing chiral amines.[49]

activity of the WT enzymes is very low with substrates of the type phenylethylamine (*rac*-**9**), and enantioselectivity is only moderate to good. After several cycles of mutagenesis using the *E. coli* XL1-Red mutator strain and transformation of the plasmid library in *E. coli*, a total of 150000 bacterial colonies were assayed for activity by using a colorimetric prescreen followed by conventional characterization of the enantioselectivity of the best hits. This procedure led to the identification of mutants displaying excellent activity and enantioselectivity (>98% *ee*). Later it was shown that some of the mutants display a remarkably broad substrate range that encompasses primary and secondary amines, and even the difficult class of tertiary amines.[49b] The simple two-step, one-pot process has been commercialized by Ingenza. The experimental platform has been extended to include other types of mutagenesis protocols.[4b,49]

A final enzyme type highlighted here, first reported by Wong and co-workers,[48a,b] concerns the directed evolution of aldolases, which has emerged as a fascinating field in itself.[48] An unusual case concerns the reversal of diasteroselectivity of an aldolase while maintaining essentially complete enantiopurity of the product, as reported by Berry and co-workers.[48c,d] The WT tagatose-1,6-biphosphate aldolase was known to catalyze the aldol addition of **11** to the chiral aldehyde **13** with stereoselective formation of aldol **12**. Three rounds of DNA shuffling were performed to obtain the diastereomeric aldol product **14** (Scheme 8). After a screening procedure, the "tagatose aldolase" was indeed found to have turned into a "fructose aldolase". Laboratory evolution of stereoselective aldolases is clearly of value in the synthesis of complex stereoisomeric products, and one can expect further positive results in this exciting new area of laboratory evolution.[48g]



Scheme 8. Diastereomeric switch in aldol additions induced by directed evolution.[48c–f]

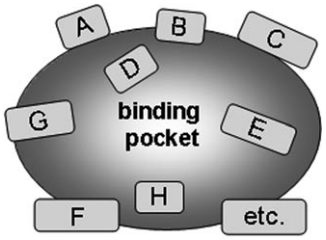
All of the above, as well as numerous further, examples of directed evolution of stereoselective enzymes proved to be successful.[4f,58] In these studies, efficiency in probing protein sequence space was not the primary goal. Indeed, very few studies have compared the use of various mutagenesis methods and strategies in terms of library quality and screening effort.[4,58] Recently, we and other research groups active in methodology development in directed evolution have addressed this pressing issue,[4,58,59] which can be compared to method development in synthetic organic chemistry.[1] We have defined the quality of a mutant library in terms of the frequency of hits and their degree of catalyst improvement,[58,60] which in turn touches on the screening problem.[10] Some research groups have utilized various computational aids in this endeavor, with prominent examples being SCHEMA,[61] ProSAR,[62] FamClash,[63a] GLUE-IT and PEDEL-AA,[63b] and others[63c–h] such as CASTER and B-FITTER developed in my research group.[20,60]

In this Review, the focus is primarily on the development of particularly effective methods in directed evolution for controlling stereoselectivity, substrate scope, and thermostability. In this regard it is shown that the iterative saturation mutagenesis (ISM) developed by us shows outstandingly high efficiency.[18,20,31,60] Applications from my group and more recently from other research groups will be presented. Where possible, the new method is compared in terms of efficacy with the older more traditional approaches to laboratory evolution of stereoselectivity.

3. Iterative Saturation Mutagenesis (ISM) as an Effective Strategy for Laboratory Evolution

3.1. General Concept

Our initial work regarding the lipase PAL had shown that saturation mutagenesis can be an effective tool in the directed evolution of enantioselective enzymes according to Scheme 1.[25,28] A lively discussion regarding the virtues of close versus distal mutations followed.[29,58a] However, at that time we failed to recognize the truly exciting perspectives that saturation mutagenesis at sites aligning the binding pocket can offer, because we did not generalize this approach. In 2005 we reported the first step in this direction with the development of the combinatorial active-site saturation test (CAST).[64] This is simply a systematization of saturation mutagenesis at all the relevant sites around the complete binding pocket of an enzyme, not just at one or two selected sites, as performed earlier by us (Figure 1)[28] and other research groups pursuing other goals.[4,65] The illustration in Scheme 9 shows putative randomization sites aligning the binding pocket of an enzyme, each being composed of one or more amino acid positions. Thus, the term CASTing is a



Scheme 9. General scheme for CASTing.[18,31,60,64] The sites A, B, C, etc. align the binding pocket and can be composed of one or more amino acid positions. The side chains of the amino acids are in general inside the binding pocket.

convenient acronym that defines the knowledge-driven systematization of focused libraries around the binding pocket,[18,31,60,64] which distinguishes it from saturation mutagenesis at remote regions of a protein (Sections 4.3 and 5).

In the light of earlier work on focused libraries for influencing enantioselectivity[9,25,28] or activity,[28,65] we hoped that systematization brought about by CASTing would constitute the fastest and most efficient way to tune these catalytic properties. It was anticipated that in this way the structure and dynamics of a binding pocket could be manipulated at will by reshaping the "lock" in Emil Fischer's lock-and-key hypothesis (or Koshland's induced fit model).[64] In a model study, we returned to the lipase from *Pseudomonas aeruginosa* (PAL) (Figure 1), and on the basis of its X-ray structure[27] defined five potential randomization sites, A, B, C, D, and E (Figure 3). The goal was to expand the substrate scope by generating mutants which are active toward bulky

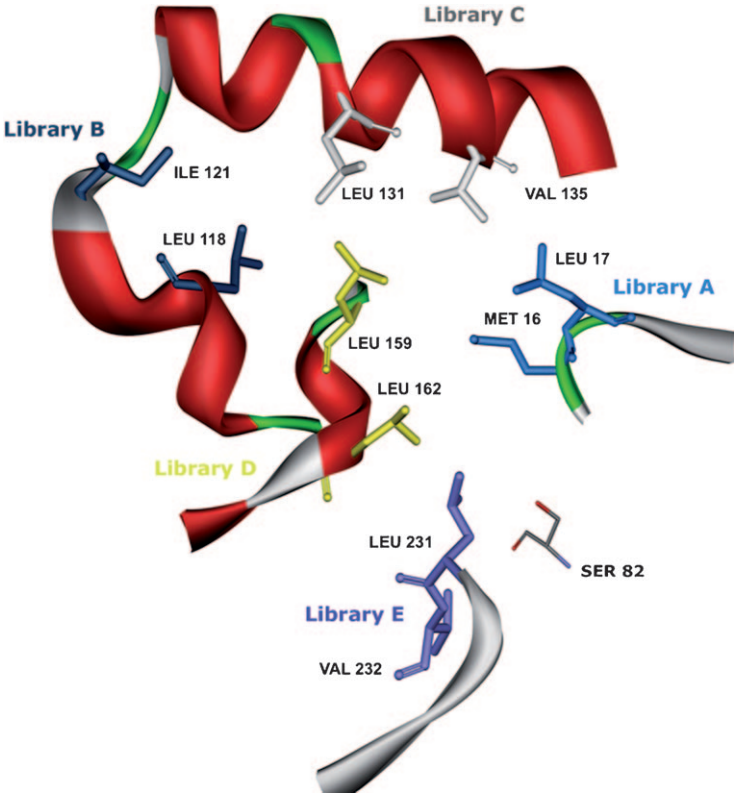
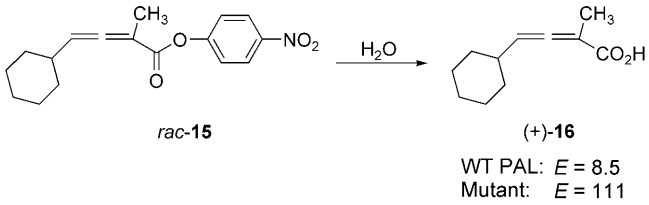


Figure 3. CAST sites A, B, C, D, and E in the lipase PAL.[64]

esters that were not accepted by WT PAL. Upon screening the five libraries, hits were indeed discovered that originated mainly from libraries generated by saturation mutagenesis at sites A and D, respectively. Thus, this is an example of solving the traditional problem of limited substrate scope of enzymes.[64] In some cases, notable enantioselectivity was observed, although this was not the primary goal of the study.

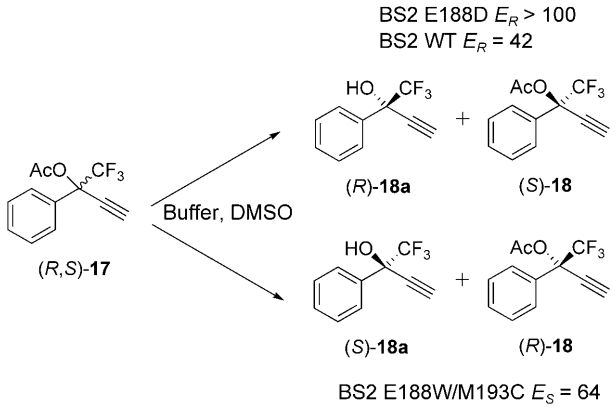
CASTing was also applied to the enantioselective Baeyer–Villiger reaction of 4-methyl- and 4-acetoxycyclohexanone, where the oxidative desymmetrization was catalyzed by cyclopentanone monooxygenase (CPMO).[66] Whereas WT CPMO delivers enantioselectivities of only 46% and 5%, respectively, this approach led to the generation of mutants with >90% *ee*.

In yet another study, Bäckvall and co-workers showed that the same approach could be applied successfully to the lipase-catalyzed hydrolytic kinetic resolution of axially chiral *rac*-**15** (Scheme 10).[67] WT PAL shows only moderate enantioselectivity in slight favor of (+)-**16** ($E = 8.5$). Screening only 600 transformants present in the saturation mutagenesis library generated at site D led to the discovery of a highly enantioselective mutant with an E value of 111, which meant that exploring the other randomization sites was not necessary from a practical point of view (although theoretically interesting).[67]



Scheme 10. Hydrolytic kinetic resolution of an axially chiral ester catalyzed by PAL.[67]

A further example of systematization on the basis of CASTing concerns the complete inversion of the enantioselectivity of the esterase from *Bacillus subtilis* (BS2) as a catalyst in the hydrolytic kinetic resolution of the tertiary acetate *rac*-**17** (Scheme 11).[68] WT BS2 shows reasonable enantioselectivity in favor of (*R*)-**18** ($E = 42$). Guided by the X-ray structure of BS2, Bornscheuer and co-workers first identified several putative randomization sites.[68] A site comprising three amino acid positions (Glu188, Ala190, and Met193) was chosen, because these residues are closest to the catalytic triad and have their side chains pointing toward the binding pocket. Simultaneous randomization using NNK codons theoretically provides $20^3 = 8000$ structurally different mutants, but only 1100 clones were screened for enantioselectivity. Remarkably, even in such a small library, mutants were discovered that displayed enhanced *R* selectivity, for example, single mutant Glu188Asp ($E = > 100$), and reversed enantioselectivity in favor of (*S*)-**18**, for example, double mutant Glu188Trp/Met193Cys ($E = 64$). A fascinating epistatic effect was revealed upon deconvoluting the double mutant. Whereas single mutant Glu188Trp shows modest reversed *S* selectivity, the other single mutant Met193Cys favors the *R* enantiomer, although to a lower degree ($E = 16$) than WT BS2. Therefore, a dramatic cooperative effect is operating through the interaction of two point mutations! Moreover, the authors observed respectable substrate scope when testing the double mutant as a catalyst in the reaction of other tertiary substrates, a finding that did not require any additional mutagenesis/screening experiments.[68]

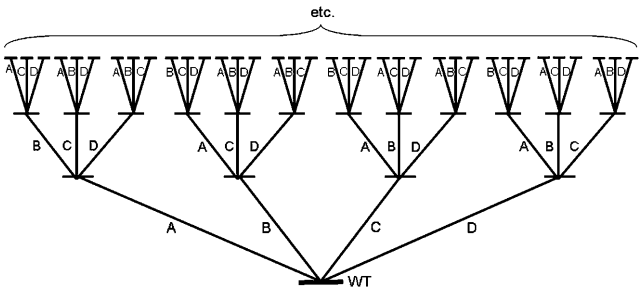


Scheme 11. Hydrolytic kinetic resolution of esters *rac*-**17** using wild-type and mutants of the esterase BS2.[68]

Numerous other studies relying on saturation mutagenesis for practical or mechanistic purposes have appeared recently,[69] although systematization along the lines of CASTing was not strived for. It is important to note once more that identifying hits in an initial library does not constitute directed evolution, although in fortuitous cases this is all that is needed.

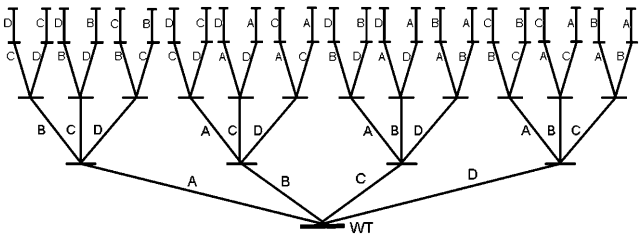
Two strategies appear logical to evolve further improvements in activity and/or enantioselectivity. One is to enter a second round of mutagenesis by combining the point mutations observed in two improved mutants originating from different libraries. This way of exerting evolutionary pressure is fast and straightforward because new libraries are not required. We used this approach to increase the activity of PAL mutants that had previously been obtained by saturation mutagenesis at sites A and D as catalysts in the hydrolytic reaction of bulky esters (Figure 3).[70] Combining the previously identified point mutations observed at sites A and D led to increased rate in most cases. Apparently, this is due to the enlargement of the binding pocket, although the exact interpretation remains difficult. It is not always the introduction of an amino acid with a geometrically smaller side chain that decreases the size of the binding pocket as traditionally thought, because subtle effects may operate, for example, the "turning away" of certain side chains from the wall of the binding pocket as a result of the formation of new hydrogen bonds or π-π interactions elsewhere.[70] Although combining mutations is to some extent successful, it has limitations because of the restricted structural diversity of the mutants.

The second strategy appeared more attractive, namely the idea of iterative saturation mutagenesis (ISM). Rather than moving from one putative randomization site to another in an unmethodical manner as done originally in the PAL project[25,28] and in some other studies with different goals,[4,65,69] we decided to systematize saturation mutagenesis. Scheme 12 illustrates the case of four sites A, B, C, and D.[20,31,60,71] The gene of an improved hit in one library is subsequently used as a starting point (template) for randomization experiments at the other sites, and the process is continued as long as necessary.



Scheme 12. Schematic illustration of iterative saturation mutagenesis (ISM) involving (as an example) four randomization sites A, B, C, and D. Each site is comprised of one or more amino acid positions.[18,20,31,60,71]

In the third and following generations, the number of libraries increases rapidly in a nonconverging manner. Simplification occurs if each site A, B, C, and D is visited only once in an upward pathway. Scheme 13 shows that the system then converges, with only 64 saturation mutagenesis libraries being involved.[20,31,60,71] As we shall see, it is not necessary to test all the pathways in such a scheme.



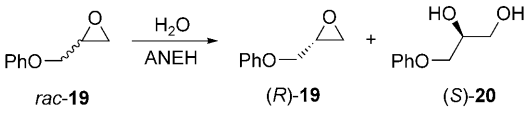
Scheme 13. Simplified ISM with four sites A, B, C and D, where each site in a given upward pathway is visited only once.[18,20,31,60,71]

On a molecular level, if mutational changes occur at one site which enhance the stereoselectivity (or substrate acceptance), and the mutant gene is used as a template to randomize at another site, the system responds to the originally introduced structural perturbation in an epistatic manner. Additive or cooperative interactions are possible, in addition to antagonistic effects which lead to inferior mutants which are eliminated by the screening process.[60,71–73] It is particularly beneficial when cooperative effects dominate. Earlier results show that this is what ocurrs in the ISM path. Synergy of this type can arise from two factors. Possible interactions are those between the mutated amino acids within a set of point mutations, but also and importantly between sets of point mutations. The first effect is logically only possible when the randomization site is composed of at least two amino acid positions. In contrast to other mutagenesis strategies, such as epPCR, the likelihood of observing cooperativity is maximized in ISM methods. Crucial to success is the correct choice of the randomization sites. In the case of enantioselectivity and/or substrate acceptance, the choice is made on the basis of the CAST concept (Scheme 9). It is possible to extend CASTing by considering second-sphere residues (Section 4.2).[42] Re-shaping the structure of the binding pocket to have broader substrate scope and/or enhanced enantiose-

lectivity can also be achieved by choosing randomization at distal sites which induce allostery in the absence of an effector[73] (Section 4.3). Randomization at sites directly aligning the binding pocket in an iterative manner has so far been used most often. ISM can also be applied to tune other catalytic parameters such as thermostability and robustness in the presence of harmful organic solvents by applying the B-FIT method[20] (Section 5).

3.2. First Example of ISM

The first case of ISM in the form of iterative CASTing concerns the directed evolution of the epoxide hydrolase from *Aspergillus niger* (ANEH) as a catalyst for the hydrolytic kinetic resolution of *rac*-**19** (Scheme 14).[31] WT ANEH



Scheme 14. Hydrolytic kinetic resolution of *rac*-**19** catalyzed by ANEH.[31]

favors the formation of (*S*)-**20**, but enantioselectivity is poor ($E = 4.6$). Although Jacobsen's chiral salen complexes are well suited for ring-opening reactions,[74] we employed this transformation as a model reaction to test and compare various strategies in directed evolution. In a preceding study, the traditional use of epPCR had proven to be only marginally successful, with the E value increasing to only 11 after the screening of 20000 transformants.[46a] The X-ray structure of WT ANEH reveals the presence of the binding pocket as a narrow tunnel,[75] which may be the reason why this enzyme appears to be a "difficult" case.

Six CAST sites A, B, C, D, E, and F were identified with the help of the X-ray structure (Figure 4), each one being composed of two or three amino acid positions.[31]

The best hit originated from the library at site B ($E =$ 14), which was then used as a template to randomize site C. Successive "visits" at sites D, F, and E provided the best mutant LW202, which showed unusually high enantioselectivity ($E = 115$), with the order being arbitrarily chosen.[31] As summarized in Scheme 15, five sets of mutations accumulated stepwise, thereby leading to LW202 with a total of nine point mutations. Since the enantioselectivity was already so high, site A was not considered, nor were other pathways explored. A total of about 20000 transformants were screened, which happens to be the same number needed in our earlier study based on epPCR.[46a] Since the results are dramatically superior, $E = 115$ versus $E = 11$, we suspected that ISM constitutes an exceptionally efficient way to generate "smart" mutant libraries.[31] Thermostability was not impaired by the mutational changes.

Three fundamental questions arose from these observations: 1) What is the source of enhanced enantioselectivity on a molecular level? This point was considered in a detailed mechanistic and structural study[76] (Section 6). 2) Is B→C→D→F→E the only pathway to a highly improved mutant, or are there other pathways according to the general concept (Scheme 13) which likewise provide (other) hits of equal or

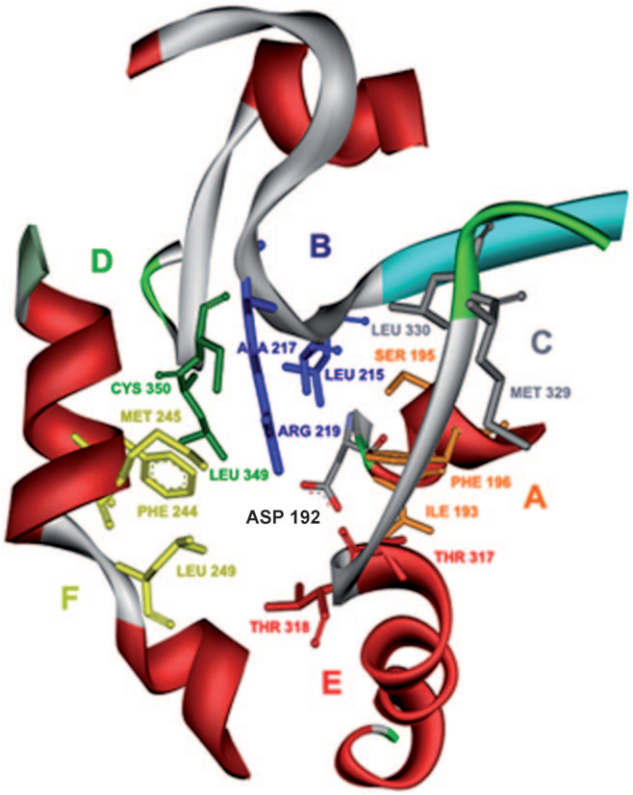
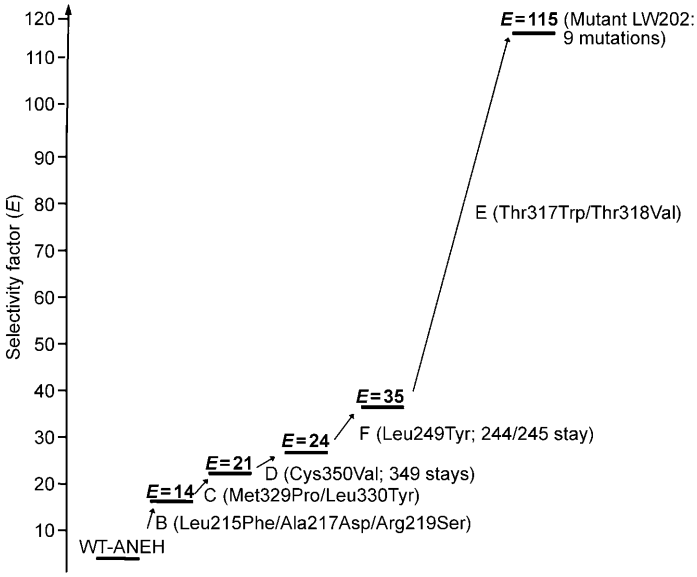


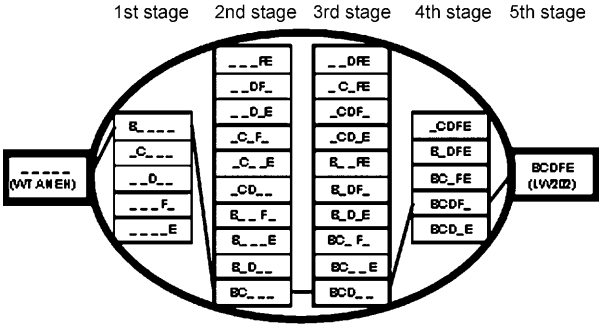
Figure 4. CAST sites A–F[31] of the epoxide hydrolase from *Aspergillus niger* (ANEH) based on the X-ray structure of the wild type.[75] Asp192 (black) initiates rate- and stereoselectivity-determining nucleophilic attack at the less-substituted C atom of epoxide *rac*-**19**.



Scheme 15. Iterative CASTing in the evolution of enantioselective ANEH mutants as catalysts in the hydrolytic kinetic resolution of *rac*-**19**.[31]

even higher enantioselectivity? This question is currently being studied in our laboratory, and preliminary results clearly show that many pathways lead to success.[77] 3) Using the five sets of mutations in all permutationally possible combinations, is the original pathway B→C→D→F→E the only one that leads from WT ANEH to the best mutant LW202? An exploration of all the pathways and analyzing epistatic interactions along each one on the basis of an exhaustive deconvolution strategy was expected to throw light on the nature of ISM. In fact, such a deconvolution strategy turns out to be a type of quality control regarding the efficacy of this type of laboratory evolution or of other mutagenesis strategies.[72]

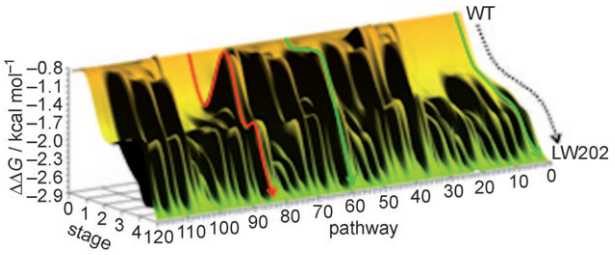
When considering the five sets of mutations that accumulated in the five iterative steps (Scheme 15), it becomes clear that $5! = 120$ pathways lead from WT ANEH to the best mutant LW202. Therefore, 30 mutants are relevant for all permutational combinations (Scheme 16).[72] By using site-



Scheme 16. The 30 possible mutants as intermediate stages between WT-ANEH and enzyme variant LW202 based on the use of five sets of mutations (Scheme 15).[72] The connecting lines indicate the original pathway B→C→D→F→E, which is one of 120 possible trajectories.

specific mutagenesis, 26 new mutants were prepared and tested in the model reaction of *rac*-**19**. Mutants corresponding to B, BC, BCD, and BCDF were already available from the original study. This procedure provided the corresponding selectivity factors *E* and, therefore, the respective free energy values ($\Delta\Delta G^{+}$).

The fitness landscape of this type of system constitutes a six-dimensional surface (mutation sets B, C, D, E, and F are independent vectors, and $\Delta\Delta G^{+}$ is the independent variable), which is difficult to depict graphically. We chose to present the data in the form of a "fitness-pathway landscape", as depicted in Scheme 17. All 120 possible pathways linking WT ANEH (top) with mutant LW202 (bottom), as specified by the experimental results, are stacked. After mapping the free energy profiles of these 120 mutational pathways, we discovered that two different types of trajectories exist. The energetically favored type shows a continuous decrease in free energy, as in the original pathway B→C→D→F→E or typically also in the case of D→C→F→E→B (Scheme 17, green pathways), both devoid of undesired local minima. In contrast, the second type of pathway is energetically disfavored, because it is characterized by the presence of a turning



Scheme 17. Energy profile of the two types of pathways leading from the WT-ANEH to the mutant LW202:[72] Energetically favored (green) as in the original B→C→D→F→E (pathway 2) or D→C→F→E→B (pathway 60) and disfavored (red) as in E→C→F→D→B (pathway 84).

point followed by a peak along the trajectory that is higher in energy than at the previous evolutionary stage, thereby defining a local minimum. A typical example is trajectory E→C→F→D→B (Scheme 17, red pathway). The result of calculating the first derivatives of the free energy differences at every stage of each pathway is in full accord with this analysis.[72]

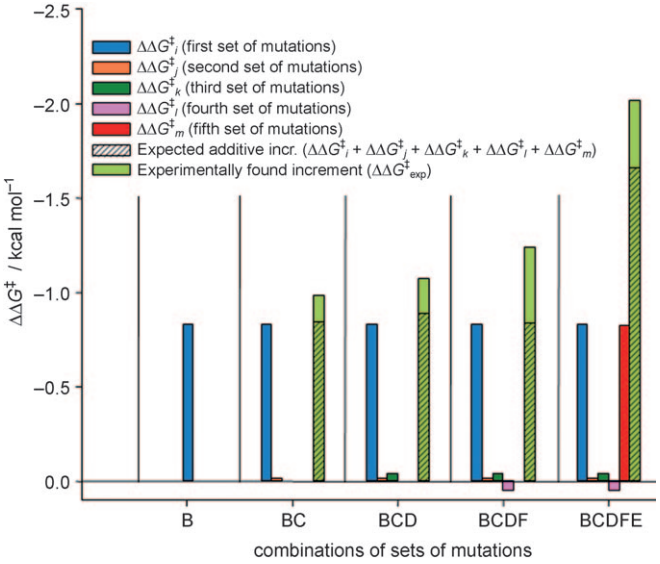
The complete fitness-pathway landscape, included in the original study, reveals some remarkable features.[72] Most importantly, a total of 55 different pathways are energetically favored, which corresponds to 45% of all the possibilities in going from WT ANEH to the specific mutant LW202. This is a high score considering the fact that new mutagenesis experiments with the introduction of additional point mutations were not performed. If this restriction were to be relaxed, one can imagine many favored pathways leading to new mutants. The present data shows that laboratory evolution based on ISM, even in this restricted sense, can follow many energetically favored pathways to generate improved enzyme variants. Comparison with other systems that lead to different conclusions are problematic because of differences in the experimental setups.[78]

All epistatic interactions[79] occurring along each of the 120 evolutionary pathways can be analyzed quantitatively, because the free energy of interaction ($\Delta G_{ij}^{\pm}$) between any two sets of mutations *i* and *j* [Eq. (1)] is experimentally accessible:[72]

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

where $\Delta\Delta G^{+}_{\mathrm{exp}}$ is the difference in the activation energy between both enantiomers obtained experimentally for the binary combination, and $\Delta\Delta G^{+}_{i}$ and $\Delta\Delta G^{+}_{j}$ are the experimental energies obtained for each set of mutants separately. The values of the ΔG^{+}_{ij} interactions are either a measure of cooperative (synergistic) effects ($\Delta G^{+}_{ij} < 0$), of additive effects ($\Delta G^{+}_{ij} = 0$, no interaction), or of partially additive effects ($\Delta G^{+}_{ij} > 0$ and ($|\Delta\Delta G^{+}_{i}|$ and $|\Delta\Delta G^{+}_{j}| < \Delta\Delta G^{+}_{\mathrm{exp}}|$), or they denote antagonistic effects ($\Delta G^{+}_{ij} > 0$ and ($|\Delta\Delta G^{+}_{i}|$ or $|\Delta\Delta G^{+}_{j}|$) $> |\Delta\Delta G^{+}_{\mathrm{exp}}|$).

Scheme 18 features the case of the original trajectory B→C→D→F→E.[72] It can be seen that the combined effect of the mutational sets at each evolutionary stage is more than additive. Cooperativity is, therefore, the underlying factor



Scheme 18. Thermodynamic cycle [Eq. (1)] revealing the interaction of the sets of mutations involved at every stage along the energetically favored pathway B→C→D→F→E.[72]

which explains the notable increase in enantioselectivity, and indeed it is characteristic of ISM in general. Strong cooperative effects operating between the sets of mutations are also observed in the other trajectories. None of the mutational sets are superfluous, that is, all five are required to attain the high enantioselectivity displayed by mutant LW202, which is quite different from other directed evolution studies. The analysis of the energetically disfavored pathways is also illuminating. It shows that local minima are due to antagonistic interactions at specific stages of the evolutionary process. It is possible to escape from a local minimum by backtracking one step, which puts the evolutionary process back on a positive pathway.[72]

3.3. Oversampling and the Numbers Problem in Directed Evolution

The concept of "oversampling" in directed evolution refers to the number of enzyme variants that need to be screened to ensure a certain percentage coverage of a given library, irrespective of the applied mutagenesis method.[4,60,80] This aspect was generally not considered in our initial studies and in reports by other research groups on saturation mutagenesis—after all, the final results proved to be positive. Nevertheless, to maximize the library quality, the issue of oversampling, which relates directly to the screening effort, needs to be considered. Fortunately, algorithms based on Poisson distributions have been developed which allow the design and assessment of all kinds of mutant libraries in protein engineering.[80b–d] Since the absence of amino acid bias is assumed, the results are only approximations, but they are highly useful because clear trends emerge. Oversampling, for example, when utilizing saturation mutagenesis, depends upon the particular codon degeneracy chosen. NNK codon degeneracy (Section 1) is used most often; this involves 32 codons and one stop codon, and encodes all 20 canonical amino acids as building blocks in the randomization process. However, it is also possible to utilize reduced amino acid alphabets,[81] which as we shall see has a profound influence on the degree of oversampling that corresponds to a certain percent coverage of a library.[60,71] As an example, NDT codon degeneracy (D: adenine/guanine/thymine; T: thymine) with only 12 codons and no stop codon defines the 12 amino acids Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, and Gly. This is a structurally balanced mixture of building blocks with polar, nonpolar, aromatic, and lipophilic side chains.

The previously mentioned statistical methods were employed to visualize the relationship between the percentage coverage of a library and the degree of oversampling, irrespective of the codon used or the type of mutagenesis method. The algorithm[80b] for estimating completeness as a function of the number of transformants (clones) actually screened (T) can be transformed into Equation (2), where P_i denotes the probability that a particular sequence occurs in the library and F_i is the frequency.[60b]

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

By substituting for F_i, the relationship reduces to Equation (3), where V is the number of gene mutants comprising a given library.

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

This relationship defines the correlation between the number of mutants V of a given library and the number of transformants T which have to be screened for a specified degree of completeness. We can then define the oversampling factor O_f according to Equation (4).[60b]

$$O_f = T/V = -\ln(1-P_i) \quad (4)$$

The result of calculating O_f as a function of the percentage coverage is shown in Figure 5.[60b] It can be seen that for 95% coverage, for example, the oversampling factor O_f amounts to about 3, which means that a threefold excess of transformants needs to be screened. The exponential character of the relationship means that degrees of coverage beyond 95% require vastly higher screening efforts. Lower degrees of library coverage may suffice in a given experiment, but

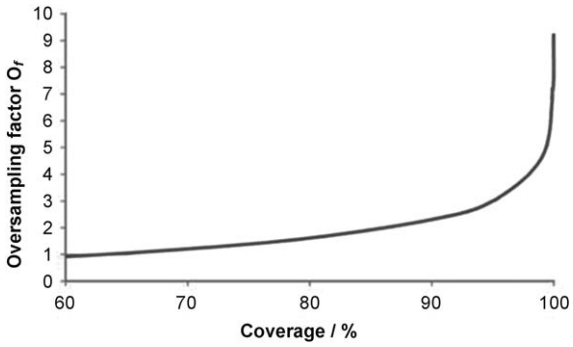


Figure 5. Correlation between library coverage and the oversampling factor O_f of enzyme variants.[60b]

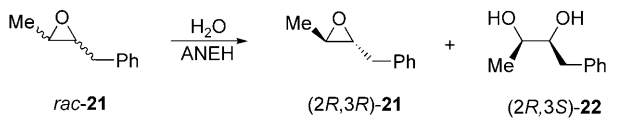
decisions regarding this important aspect should be viewed in light of the curve shown in Figure 5.

In the case of simultaneous randomization at sites composed of more than one amino acid position, we calculated for the respective NNK and NDT libraries the amount of oversampling, in absolute numbers, necessary for 95% coverage of relevant protein sequence space (Table 2).[60] The potential screening effort is very different for the NNK and NDT systems. For example, for 95% coverage in the case of a site composed of three amino acid positions, NNK requires almost 100000 clones, whereas NDT needs only about 5000. We have derived curves for NNK versus NDT codon degeneracy that show the required number of screened clones as a function of library coverage ranging between 0% and 95%. These curves should help when saturation mutagenesis is to be applied.[60b] In a similar way, the analysis of other codon degeneracies is possible by employing the CASTER computer program.[20,60b]

Table 2: Oversampling necessary for 95% coverage as a function of NNK and NDT codon degeneracy (assuming the absence of amino acid bias).[60b]

number of amino acid positions at one site	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	$>1.1\times10^{15}$	$>3.4\times10^{15}$	$>6.1\times10^{10}$	$>1.9\times10^{11}$

The question of the relative quality of NNK versus NDT libraries is exciting, because it has practical ramifications. We therefore constructed an NNK and an NDT library of ANEH mutants, and in each case screened only 5000 clones, which corresponds to 15% versus 95% library coverage, respectively.[60b] The hydrolytic kinetic resolution of the *trans*-disubstituted epoxide *rac*-**21** with formation of (2*R*,3*S*)-**22** was chosen as the model reaction (Scheme 19). Since WT ANEH displays extremely low activity towards this substrate, the goal was to increase the activity, while achieving or maintaining high enantioselectivity. The two same-sized libraries were generated by saturation mutagenesis at site B of WT ANEH (Section 3.2, Figure 4). The NDT library proved to have a considerably higher frequency of hits than the NNK library, with the best mutants showing selectivity factors *E* ranging between 56 and 200.[60b] This transformation is also of synthetic significance, because Jacobsen's chiral salen catalyst is not well suited for *trans*-disubstituted epoxides of type *rac*-**21**.[74] The virtue of using reduced amino acid alphabets to beat the numbers problem was also demonstrated in other studies[18,41,66] (Sections 3.4.1 and 4.2).



rac-**21** (2*R*,3*R*)-**21** (2*R*,3*S*)-**22**

Scheme 19. Hydrolytic kinetic resolution of *rac*-**21** catalyzed by the epoxide hydrolase from *Aspergillus niger* (ANEH).[60b]

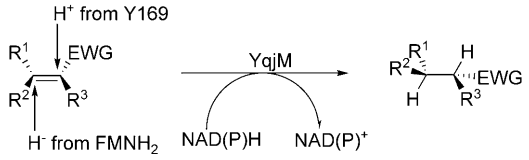
3.4. Further Examples of ISM

The positive experience with ISM raised some questions concerning further research: How general is this strategy? Once appropriate residues around the binding pocket are identified, how should they be grouped into putative sites for randomization? Does it work with enzymes in which domain movements are essential for catalysis? How should ISM be handled when attempting to improve two different catalytic properties simultaneously? Can ISM be applied to protein properties other than catalysis, for example, selective binding? We have addressed some of these issues through a range of projects (Sections 3.4.1 and 3.4.2), and other research groups have likewise contributed to the generalization of ISM (Sections 3.4.3 to 3.4.10).

3.4.1. Increasing Substrate Scope and Enantioselectivity of Enoate Reductase

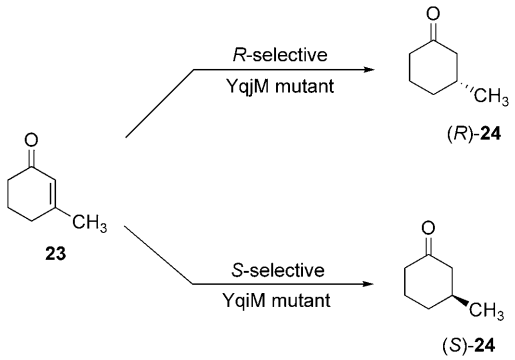
The biocatalytic asymmetric reduction of conjugated olefins mediated by enoate reductases from various sources[2,82] is emerging as an attractive alternative to transition-metal catalysis[83] and organocatalysis.[84] An interesting case regarding laboratory evolution[18] concerns the 37.4 kDa enoate reductase YqjM from *Bacillus subtilis*,[82b] a homologue of the famous Old Yellow Enzyme. Similar to other enoate reductases, YqjM is flavin-dependent. $FMNH_2$ delivers a hydride from one π face of the activated olefin and a proton is provided on the other face in an overall *trans* addition. The X-ray structure of YqjM led to the identification of Tyr169 as the proton source (Scheme 20).[82c] For the transformation to occur, the carbonyl function of the substrate needs to be activated by two hydrogen bonds that originate from His164 and His167.

Unfortunately, WT YqjM shows low or no activity and only moderate to poor enantioselectivity for many substrates of practical interest. For example, the reduction of 3-methylcyclohexenone (**23**) occurs in a slow reaction with an



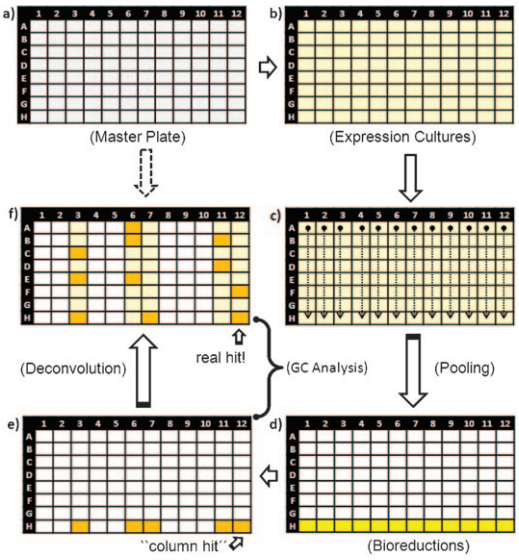
Scheme 20. Mechanism of reduction of activated olefins (EWG: electron-withdrawing group) catalyzed by the enoate-reductase YqjM, derived on the basis of its X-ray structure.[82]

unsatisfactory enantioselectivity of 76% *ee* in favor of (*R*)-**24** (Scheme 21).[82d] Thus, the goal was to test the efficacy of ISM in the quest to enhance the *R* selectivity and to reverse the stereoselectivity in favor of (*S*)-**24**, while increasing the enzyme activity.[18] We hoped that the best *R*- and *S*-selective hits would also be acceptable catalysts for the asymmetric reduction of other substrates.



Scheme 21. Model reaction for the directed evolution of enoate-reductase YqjM.[18]

Although, by nature, ISM entails relatively small mutant libraries, we nevertheless sought to reduce the screening effort even more by pooling[24] (Section 1). Inspired by the idea of Phizicky and co-workers,[85] in which predefined pools of proteins from genome sequences were screened to link biological activities with uncharacterized proteins, we devised a stepwise GC-based protocol (Scheme 22).[18] The application of this protocol to the YqjM project, by using 96 microtiter wells, demonstrated that less screening was necessary. The prescreening of eight pooled wells would require that less than 50% of the columns would have needed to be analyzed. Thus, the amount of screening was reduced by more than half. The benefits of this pooling approach are especially pronounced when a randomization site is incorrectly chosen, because in these cases such a site is rapidly identified and eliminated. For example, instead of 96 GC runs per plate, 12 are sufficient because there is no need for deconvolution.[18]



Scheme 22. Schematic protocol for screening by pooling defined cell cultures overexpressing YqjM variants.[18] a) Pick and inoculate individual colonies. b) Induce expression of YqjM variants. c) Recover by centrifuging individual cell pellets and combine all eight cell pellets belonging to the same column. d) Lyse cells and incubate 12 biotransformations per plate by adding the appropriate reagents. e) Extract product with organic solvent and analyze organic layer by GC. Column hits will be identified in this step by setting an appropriate threshold.

Guided by the X-ray structure of YqjM with the inhibitor *p*-hydroxybenzaldehyde (pHBA),[82c] we considered a total of 20 residues for potential saturation mutagenesis (Figure 6).

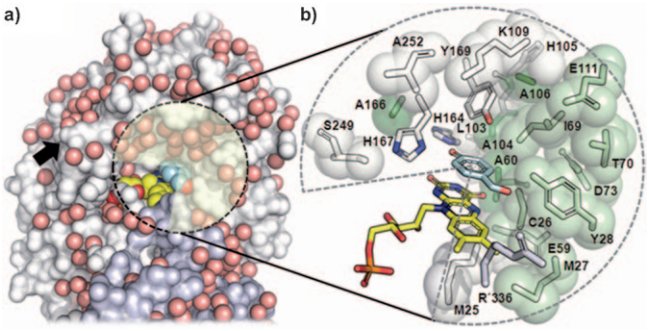


Figure 6. The 20 amino acid positions near the binding pocket of the YqjM flavoprotein considered for saturation mutagenesis (potential sites for CASTing).[18] a) Space-filling model of YqjM (chain A: light gray; chain B: light blue; red spheres: water molecules; yellow spheres: FMN; blue spheres: *p*-hydroxybenzaldehyde as inhibitor). Certain water molecules at the entrance of the catalytic cavity were removed for clarity. The black arrow indicates the distant loop. b) Close-up view of the catalytic active site. Amino acids in green responded positively, gray ones negatively to saturation test.

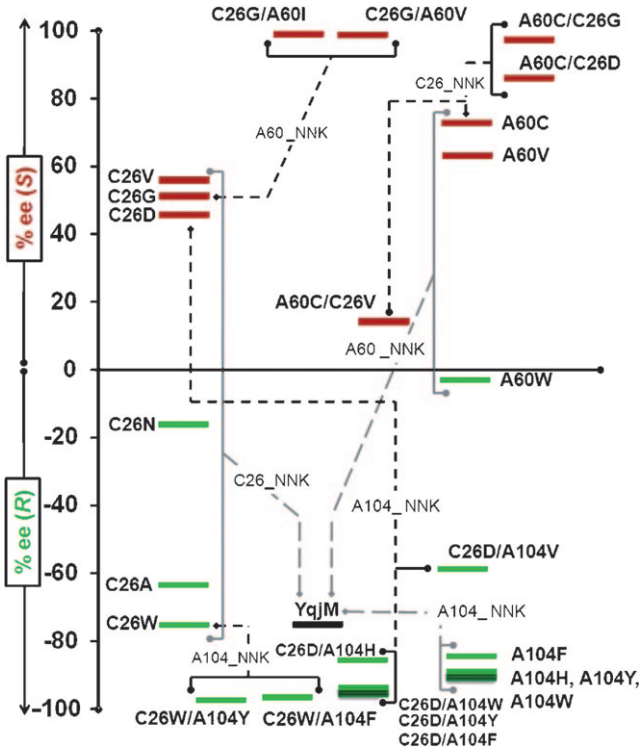
This decision was based on the distance of the residues from the inhibitor. The residues of the catalytic triad His164/His167/Tyr169 were excluded because they participate directly in the catalytic process. The selected residues were roughly assigned to three groups:[18] 1) A "hot zone" composed of positions Cys26, Tyr28, Ile69, Asp73, and Thr70 and to a lesser extent Met25, Met27, and Glu111, where the respective side chains are closest to the β-C atom of the substrate; 2) a presumably "soft zone" comprising Ala60, Ala104, and Ala106 and to a smaller extent Lys109, where the three alanine residues presumably point to the α-C atom of the substrate; 3) a presumably "indifferent zone" composed of Leu103, His105, Ala166, Ser249, Gly251, Ala258, and Arg336, with the latter belonging to chain B.

Following the identification of potential randomization positions, a second decision had to be made, as in any ISM study, of how to group the residues into sites. In the present case of 20 residues, it would be possible to choose, for example, ten sites each comprising two amino acid positions or some different grouping. A site composed of two or more residues makes cooperative effects between the individual residues possible (Section 3.2), in addition to those potentially operating between sets of mutations in the follow-up

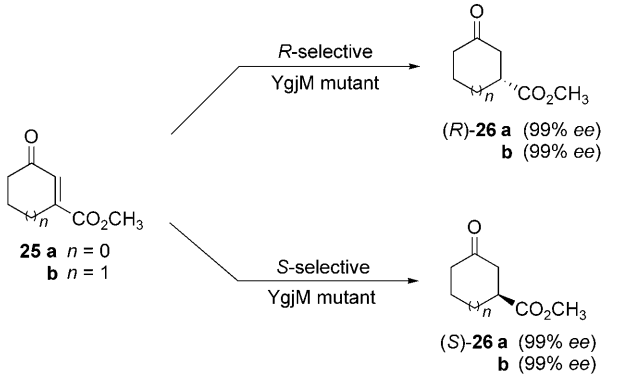
steps of ISM.[72] However, as seen in Section 3.3, such sites require a greater degree of oversampling than in the case of single residue sites (which to some extent can be compensated by the use of reduced amino acid alphabets).[60] Here we opted for 20 single residue randomization libraries using NNK codon degeneracy that encodes all 20 amino acids and requires only about 100 transformants for 95% library coverage.[18] In an additional experiment involving simultaneous randomization at the two distal amino acid positions Cys26 and Ala60, NDT codon degeneracy that encodes 12 amino acids (Section 3.3) as well as the improved SLIC-based method for saturation mutagenesis was applied. Since the model compound **23** is hardly reduced by WT YqjM (only 0.6–1.2% conversion was found to occur under the conditions of 96 deep-well plate conditions), the screening system was designed to identify YqjM mutants with enhanced activity.[18]

Some remarkable observations were made by using our pooling protocol coupled with an automated GC screen for activity and subsequently for enantioselectivity with each of the 100 mutants in the first 20 mutant libraries: 1) out of the 20 libraries, 9 contained improved mutants. 2) Both *R*- and *S*-selective mutants were discovered, often in one and the same library, which means that the nature of the newly introduced amino acid determines the direction of the enantioselectivity. 3) Extrememly high *ee* values of up to 91% (*R*) and 84% (*S*) were achieved in this first round of saturation mutagenesis, although cooperative effects are not (yet) possible. A limited number of ISM experiments were then performed in a second round by choosing the genes of some of the better, but not best, hits in terms of activity as templates for randomizing at another amino acid position which had responded positively in the initial round. It is not always optimal to choose the very best hit for further mutagenesis, especially when optimizing two different catalytic properties simultaneously (see below). Complete systematization was not strived for, because this limited exploration with less than 10 libraries already provided a number of *R*- and *S*-selective double mutants with significantly improved activity and enantioselectivity exceeding 95% *ee* (Scheme 23). A total of only 5000 transformants had to be screened, thus indicating the high efficacy of the ISM approach. It can also be seen that several pathways lead in each case to highly improved catalysts, with the nature of the exchanged amino acids being quite different. This finding shows that more than one mutant solves the problem of increasing activity and enantioselectivity in both the *R*- and *S*-selective regimes. Finally, only the double mutants "do the job", and strong cooperative effects operate between the two point mutations. This finding suggests that grouping the 20 residues into 10 sites each comprising two amino acid positions, for example, can be expected to be an even more efficient ISM strategy, which needs to be checked in future studies.

Some of the mutants were subsequently tested as catalysts in the asymmetric reduction of a number of other substrates such as 3-alkylcyclohexenone derivatives and other compounds. In a remarkably large number of cases both *R* and *S* selectivity was observed (93–99% *ee*). Typical substrates are **25a**,**b** (Scheme 24).[18] Here again, it is a rewarding exercise to evaluate a given mutant, originally evolved for a specific substrate, as a catalyst in the reaction of a set of other substrates for the purpose of defining the substrate scope, without performing additional mutagenesis/screening experiments.



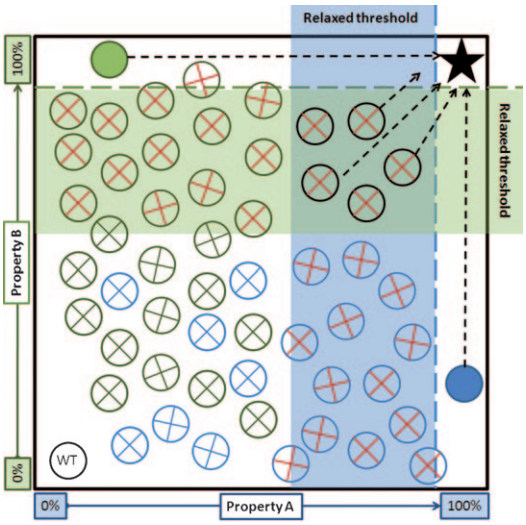
Scheme 23. ISM applied to the YqjM-catalyzed bioreduction of substrate **23** as a model reaction.[18] Light gray dashed lines indicate the first round of mutagenesis, black dashed lines the second round.



Scheme 24. Asymmetric reduction of ketoesters **25a**,**b** catalyzed by YqjM mutants originally evolved for the transformation of the model compound **23**.[18]

Whenever two parameters need to be improved in directed evolution, such as activity and enantioselectivity in the present case, two different strategies can be considered: sequential and simultaneous optimization. In this particular study, the simultaneous evolution of higher activity and enantioselectivity was achieved. In doing so, it was crucial that the constraints applied were not too stringent, such that YqjM variants were selected in the screening process with a range of

activity, rather than utilizing only the most active mutant. The best mutant is not necessarily the optimal one.[18] Especially when choosing single residue saturation mutagenesis, it is wise to accumulate a panel of hits with medium and high activity, which are then screened for the second catalytic parameter (enantioselectivity) before going into the next mutagenesis/screening round. These kinds of nondiscarded variants ("lateral hits")[18] can be related to neutral drifts, proposed in other systems by the research groups of Tawfik,[86] Arnold,[87] and DePristo.[88] We also note a relationship with the Eigen/Schuster notion of quasispecies,[89] which has been invoked by Kurtovic and Mannervik in their directed evolution studies.[90] The approach utilized in the YqjM study is generalized in Scheme 25. The upper right section of the scheme contains the desired variants with improved properties of both catalytic parameters A and B.



Scheme 25. Preferred approach for the simultaneous optimization of two catalyst properties A and B.[18] The black star indicates the desired variant; blue and green dashed lines: stringent thresholds; blue and green rectangles: relaxed thresholds; blue and green filled circles: best mutant for property A and B, respectively, which are not used in further mutagenesis; blue and green circles with red crosses: variants with improved property A or B; black circles with a red cross: mutants with improved A and B properties. Black dashed arrows: second round of mutagenesis.

Finally, in this ISM study the question of library quality with regard to amino acid bias was addressed.[18] Such quality control is rarely performed because of the relatively high experimental effort. However, the increasing emphasis on efficacy in directed evolution[4,18,20,59,60] means that such investigations are in fact rewarding. Indeed, without such checks, screening efforts may be wasted, that is, "You should not search for something that does not exist"![18] Libraries in which the circular template has not been efficiently eliminated will require considerably more oversampling. Excessive occurrence of WT clones reduces library quality. Even worse, in those cases in which certain codons are under-represented or even missing from the library, even higher degrees of oversampling will not solve the problem. Therefore, at least some form of quality control helps these aspects to be assessed and the oversampling adjusted.

Extracting and analyzing all the plasmids of a given library entails a huge if not insurmountable amount of work. In a "short-cut" control, we checked the quality of libraries by performing sequence analyses of pooled plasmids for each library prior to transformation into the expression strain.[18] As anticipated, the four bases are not equally represented as they ideally should be for adenine (A), thymine (T), cytosine (C), and guanine (G; Table 3). A small degree of bias (5–10%) can be expected to occur because of the possible imbalance at the synthesis level of the primers. Yet another problem is the presence of a small but effectively transformed circular template which survives digestion. Finally, some bias as a result of the different annealing efficiency of the pooled primers also needs to be taken into account. In the case of libraries A and B generated by using NNK codon degeneracy, two ANK degeneracy libraries were also created to compensate for the small percent of adenine in the first. After performing simultaneous randomization at the two distal sites

Table 3: Degeneracy of YqjM saturation mutagenesis libraries.[a,18]

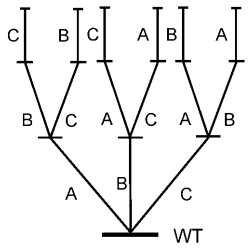
Library	First base	Second base	Third base	Library	First base	Second base	Third base
A	T	G	C	L137	G	A	G
Cys26_NNK				Glu59_NNK			
B	T	A	T	L138	A	A	A
Tyr28_NNK				Lys109_NNK			
C	A	T	C	L139	G	A	G
Ile69_NNK				Glu111_NNK			
L93	G	C	G	L152	A	T	G
Ala60_NNK				Met25_NNK			
L94	A	C	T	L153	G	T	T
Thr70_NNK				Leu103_NNK			
L95	G	C	C	L154	C	A	T
Ala104_NNK				His105_NNK			
L119	G	A	C	L155	G	C	C
Asp73_NNK				Ala106_NNK			
L120	G	G	C	L156	G	C	C
Gly251_NNK				Ala166_NNK			
L121	G	C	C	L157	G	C	G
Ala252_NNK				Ser249_NNK			
L136	A	T	G	LG	A	G	A
Met27_NNK				Arg336_NNK			
L263 C26_NDT	T	G	C	A60_NDT	G	C	G

[a] The percentage of each base was calculated from the sequencing data of the pooled plasmids for each library. Color code: green: adenine; red: thymine: black: guanine; blue: cytosine. The results from the simultaneous randomization at two positions were derived from sequence analysis of 11 individual clones.

in YqjM (positions 26 and 60) by using our SLIC-based method, individual colonies were picked for sequencing. The results showed, inter alia, that almost complete elimination of the pET21a-YqjM WT template had been accomplished with only a small degree of bias remaining for certain bases/codons.[18]

3.4.2. Revisiting Ps. Aeruginosa Lipase (PAL)

As reviewed in Section 2, the lipase from *Ps. aeruginosa* (PAL) served for several years as the model enzyme for testing a variety of mutagenesis methods and strategies. The model reaction was the hydrolytic kinetic hydrolysis of *rac*-**1**. Therefore, it seemed logical to revisit this catalyst system, this time by applying ISM. An additional goal was to explore all theoretically possible pathways in a defined evolutionary route along the lines of Scheme 13.[91] Since the analysis of a complete four-site ISM scheme would entail a great deal of experimental work, we decided to reduce the relevant protein sequence space to a three-site system, thus requiring a maximum of only 15 randomization libraries (Scheme 26).[91] In the particular case at hand, each of the three sites A, B, and C are composed of two amino acid positions.

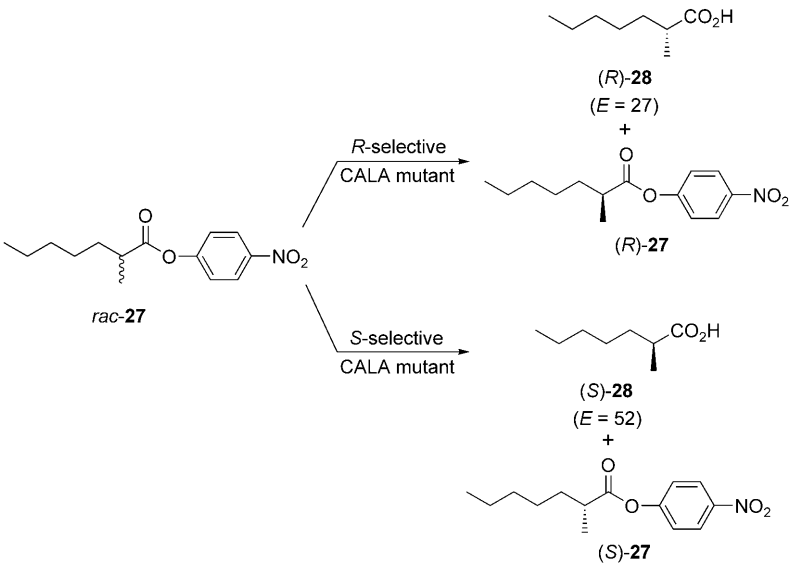


Scheme 26. ISM system comprising three sites A, B, and C.[91]

After partial exploration of Scheme 26 and screening less than 10000 transformants, a very active and highly enantioselective PAL mutant was discovered that showed a selectivity factor of $E = 594$ in the kinetic resolution of *rac*-**1**, with preferential formation of (*S*)-**2**.[91] This value was determined from the kinetics of the *R* and *S* enantiomers separately (not by the Sih formula). This is dramatically better than the best previous mutant with $E = 51$, which had been evolved by a combination of epPCR, DNA shuffling, and saturation mutagenesis and which had required the screening of 50000 transformants.[28,58a] Furthermore, the underlying reason for enhanced *S* selectivity on a molecular level must be different in this case, because remote effects are excluded. We conclude once more that ISM is the superior strategy.[91]

3.4.3. Increasing and Reversing the Enantioselectivity of Lipase A from Candida antarctica

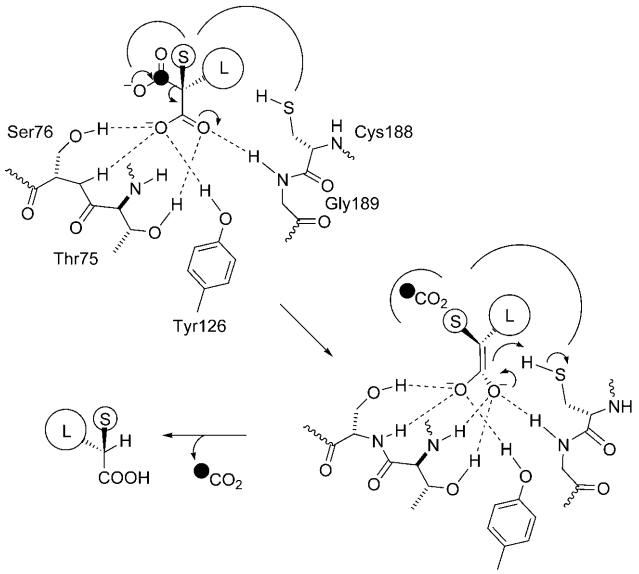
In the first directed evolution study of *Candida antarctic* A (CALA), Bäckvall and co-workers tested iterative CASTing as a means to enhance *S* selectivity and to invert stereoselectivity in the hydrolytic kinetic resolution of *rac*-**27** (Scheme 27).[92a] The use of WT CALA results in a selectivity factor of only $E = 5.1$ in favor of (*S*)-**28**. Guided by the X-ray structure of CALA,[93] the authors defined six potential "hot sites" for saturation mutagenesis, mostly at the binding pocket, but also in one case at its entrance and the lidlike C-terminal flap: double residue sites Phe233/Gly237, Ile336/Val337, Phe149/Ile150, and Leu225/Thr221 (using NDT codon degeneracy; screening requirement for 95% coverage, about 400 transformants), and single residue sites Phe431 and Asp95 (using NNK codon degeneracy; screening requirement for 95% coverage, about 100 transformants). In the initial round of randomization, the two best mutants showed enhanced *S* selectivity (Thr64Met/Phe233Asn/Gly237Leu: $E = 19$) and reversed *R* selectivity (Phe233 Leu/Gly237Tyr: $E = 27$). It should be noted that point mutation Thr64Met is due to a misincorporation during the PCR step, a phenomenon which may occur in any directed evolution process.[4] It was assumed that this point mutation plays no significant role in enantioselectivity. By using the gene of mutant Thr64Met/Phe233Asn/Gly237Leu as a template for saturation mutagenesis at site 149/150 an *S*-selective mutant (Thr64Met/Phe233Asn/Gly237Leu//Phe149Ser/Ile150Asp) was identified that showed markedly enhanced enantioselectivity ($E = 52$).[92a] These results are especially notable because only a very small number of transformants had to be screened. More extensive exploration of the defined protein sequence space can be expected to provide further improvements, and indeed remarkable results have since been obtained by using ISM.[92b]



Scheme 27. Hydrolytic kinetic resolution of *rac*-**27** catalyzed by CALA mutants.[92]

3.4.4. *Increasing Substrate Scope and Enantioselectivity of an Aryl Malonate Decarboxylase.*

The asymmetric decarboxylation of prochiral α-disubstituted malonic acids constitutes an elegant route to chiral carboxylic acids. A variety of metal-catalyzed and organocatalyzed enantioselective decarboxylative protonation (EDP) reactions have been reported, but thus far the enantioselectivity has remained moderate.[94] In contrast, enzymatic reactions based on aryl malonate decarboxylase (AMDase) from *Bordetella bronchiseptica* were shown to be highly enantioselective for a number of substrates.[2,95] Although enzymes of this type are robust and require no cofactor, they all suffer from limited substrate scope in that an α-aryl substituent needs to be present. Replacing the aryl group by alkenyl groups would be of clear synthetic value. Leys, Micklefied, Hauer, and co-workers have undertaken a giant step in solving this problem through application of the ISM concept.[96] Examination of the X-ray structure of AMDase led to the first proposed mechanism involving an intermediate enolate dianion (Scheme 28). A small and a large binding cavity to accommodate small (S) and large (L) substrate substituents, for example, methyl and phenyl groups, respectively, were identified.



Scheme 28. Proposed mechanism of AMDase, where S and L denote small and large groups, respectively.[96]

Five key active site residues were chosen for saturation mutagenesis (Figure 7 A), these being assigned to two groups corresponding to the solvent-exposed phenyl-binding pocket (Pro14, Pro15, and Gly190) and to the smaller hydrophobic cavity (Val43 and Met159).[96] By using phenylmalonic acid as a stereoselectivity nonrelevant substrate for initial activity screening, the phenyl/alkenyl binding pocket was targeted by generating three separate NNK-based saturation mutagenesis libraries at positions 14, 15, and 190, in each case requiring only about 100 transformants for 95 % library coverage. Some

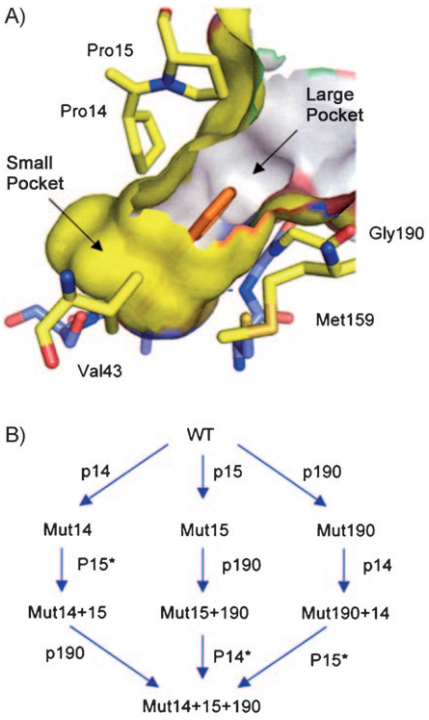
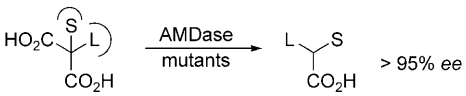


Figure 7. A) Active site of the AMDase of *B. bronchiseptica* showing the large and small binding cavities. B) Strategy for mutagenesis based on restricted ISM.[96]

of the single mutants proved to be more active than WT AMDase, and were subsequently used as starting points for randomization of the other sites, with a third round of ISM carried out in some cases. Thus, this part of the study constitutes a three-site ISM system, as in the PAL case (Scheme 26, Section 3.4.2), but this time with single-residue sites. However, not all the pathways were explored, nor do they need to be. In the case of randomization at sites aligning the smaller hydrophobic cavity, improved single mutants with amino acid exchanges at position 159 were identified, but none at position 43. No improved variants were discovered when the genes of single mutants of Met159 were subjected to mutagenesis at position 49. The possibility of ISM connections between positions 14/15/190 and 49/159 was not studied, because the results were already impressive, with mutants obtained with notably higher enzyme activities (rate increase by a factor of up to 51). The limited ISM exercise is summarized in Figure 7 B.[96]

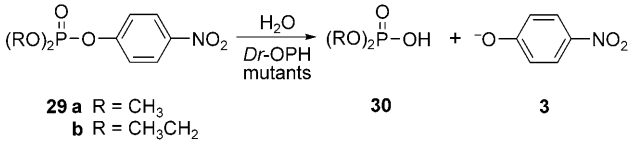
Importantly, the improved mutants showed notably higher activity in the reaction of 2-alkyl-2-alkenyl and 2-amino-2-phenylmalonic acid derivatives relative to WT AMDase, while maintaining high enantioselectivity (> 95 % *ee*; Scheme 29). The double mutant Pro14Val/Pro15Gly turned out to be one of the most versatile catalysts, for which a simple model was proposed. It is likely that further ISM exploration would provide even better biocatalysts for synthetically important transformations of this kind. Such knowledge-guided directed evolution might also be effective in reversing the enantioselectivity, which would be of great synthetic significance.



Scheme 29. Enantioselective decarboxylation of prochiral malonic acid derivatives catalyzed by AMDase mutants.[96]

3.4.5. *Manipulating the Substrate Scope of Amidohydrolase from Deinococcus radiodurans with Generation of a Phosphotriesterase*

In this section a recent study is highlighted in which saturation mutagenesis, performed iteratively at sites aligning the binding pocket of an enzyme, was utilized to influence its substrate scope (rate) in the absence of stereochemical features.[97] Following the pioneering work of Raushel and co-workers[53] on the directed evolution of phosphotriesterases as catalysts for the degradation of organophosphates (pesticides and chemical warfare agents), it was discovered that an enzyme from the amidohydrolase family originating from *Deinococcus radiodurans* (*Dr*-OPH) also shows phosphotriesterase activity.[97] However, as so often in promiscuous enzyme reactions,[98] the activity proved to be very low for most substrates. Since *Dr*-OPH is unusually thermostable and tolerates high phosphoester concentrations, it was an attractive goal to increase the activity by directed evolution, thereby expanding the substrate scope in a practical way. In an extensive and thoroughly performed study involving kinetics, DSC experiments, and X-ray structural analyses of several evolved mutants, Mesecar and co-workers employed a combination of rational site-directed mutagenesis, epPCR, and ISM.[97] The detoxifying hydrolysis of methyl- (**29a**) and ethylparaoxone (**29b**) was employed as the model reaction. The goal was to increase the catalyst activity of *Dr*-OPH in the degradation of these slow-reacting substrates (Scheme 30).



Scheme 30. Detoxifying hydrolysis of methyl- and ethylparaoxone (**29a**,**b**) catalyzed by *Dr*-OPH mutants.[97]

Saturation mutagenesis at selected sites next to the binding pocket proved to be successful, especially when applied iteratively in three rounds. A total of 30000 transformants were screened, which led to several highly improved variants. The catalytic activity of *Dr*-OPH towards the degradation of **29a** and **29b** improved 322- and 126-fold, respectively, and also resulted in specificity. Moreover, a comparison of the X-ray structures of WT *Dr*-OPH and of the improved mutants proved to be helpful in proposing a model that explained the drastic changes in the catalytic profiles. Although a systematizatization of all the possible iterative pathways was not strived for, the practical results are impressive, and led the authors to conclude that the ISM is an ideal way to perform directed evolution.[97]

3.4.6. *Altering Coenzyme Binding Specificity of a Xylose Reductase*

ISM is not only useful for enhancing the enantioselectivity and/or influencing substrate acceptance. As reported recently by Lin and co-workers, it can also be applied when a change in the binding specificity of the coenzyme is desired.[99] The particular goal was to enforce a switch from NADPH to NADH specificity of the NADPH-dependent xylose reductase from *Pichia stipitis* (PsXR). This enzyme plays an important role in xylose fermentation with formation of bioethanol. Being the second most abundant renewable sugar on earth, the fermentation of xylose is considered to have significant economical potential.[100] However, practical problems arise in the fermentation process, since the host organism *Saccharomyces cervevisae* is unable to utilize xylose efficiently. Therefore, strategies have been developed to introduce genes involved in xylose metabolism from other organisms, but this led to the necessity of engineering a NADPH→NADH specificity switch, which would alleviate the inherent redox imbalance in the respective metabolic pathway.[99] Since the X-ray structure of PSXR was not available, a homology model was constructed, guided by crystallographic data of the related xylose reductase from *Candida tenuis*[101] (76% homology). On this basis, six putative randomization sites were identified which were predicted to be important for coenzyme binding: Gln219, Glu223, Lys270, Ser271, Asn272, and Arg276. The six amino acid positions were grouped for structural reasons into three double residue sites, namely 270/272, 271/276, and 219/223, respectively (Figure 8).[99]

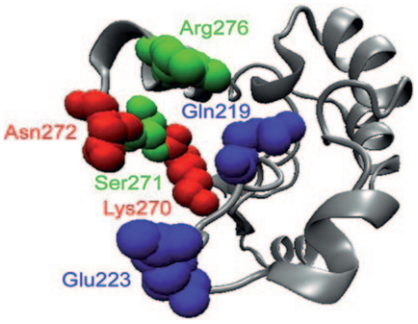


Figure 8. Six residues in the xylose reductase PsXR predicted to participate in NADPH/NADH binding.[99] For the ISM process they were grouped into three double residue sites (red, green, and blue).

As in some of the previous studies, the complete ISM scheme involves six potential upward pathways that require a total of 15 saturation mutagenesis libraries (see Scheme 26, Section 3.4.2), but it is not necessary to explore all the theoretical options. By using NNK codon degeneracy, the authors chose one of the six ISM pathways defined by (270/272)→(271/276)→(219/223). About 1400 transformants were screened in each library, which corresponds to 75% library

coverage assuming the absence of amino acid bias. The starting library contained an improved mutant, Lys270Ser/Asn272Pro, which was then used in the next round to randomize at site 271/276. This led to the quadruple mutant Lys270Ser/Asn272Pro/Ser271Gly/Arg276Phe, which favored NADH over NADPH by a factor of 13. This corresponds to a 42-fold improvement, as measured by the respective k_{cat}/K_M values. The kinetic analysis showed that the switch in coenzyme specificity is caused mainly by a decrease in the k_{cat}^{NADPH} value. The last round of saturation mutagenesis at site 219/223 did not lead to any significant further improvement. However, as the authors point out, this does not mean that 219/223 is not a "hot" site, since a different ISM order could well reveal positive mutations there.[99] Further ISM exploration could well afford even better results, but the practical problem was already solved. The best PsXR mutant can now be cloned into recombinant xylose-fermenting *S. cervevisae* together with PsXDH, thereby allowing for the construction of a new metabolic pathway based on renewable xylose with optimum redox properties.[99] In conclusion, this is the first case of the application of ISM in the quest to alter the binding specificity of a coenzyme, and since the strategy is straightforward and requires relatively small libraries, more cases of this kind can be expected in the future.

3.4.7. *Turning a Cellobiose into a Lactose Phosphorylase*

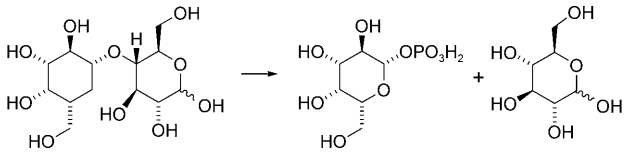
Disaccharide phosphorylases (DSPs) constitute a special class of enzymes (E.C. 2.4.1) that catalyze the reversible phosphorolysis of disaccharides with generation of a glycosyl phosphate and a monosaccharide (Scheme 31). Their natural

Scheme 31. Selective cleavage reactions catalyzed by disaccharide phosphorylase.[102]

role concerns the energy-efficient metabolism of carbohydrates. They also possess glycosyl-transfer and glycosyl hydrolase properties. They have, therefore, been exploited as catalysts in a number of synthetically significant transformations, which are clearly of interest whenever traditional synthetic organic techniques cannot be used.

It is well known that DSPs cannot utilize lactose (**31**) as a preferred substrate,[102] which is unfortunate because then a cheap and simple route to the valuable product α-D-galactose 1-phosphate (αGal1P, **32**) could emerge (Scheme 32). Thus far only one phosphorolytic enzyme (present in human milk) is known to catalyze this particular highly selective reaction, but the protein is difficult to obtain in sufficiently large quantities.[103] De Groeve, Vandamme, Soltaert et al., devised a clever strategy to solve this fundamental problem by turning a cellobiose phosphorylase (CP) into an active lactose phosphorylase (LP) through directed evolution.[104] Initially they applied epPCR, but then turned to saturation mutagenesis in the form of ISM.

The researchers chose the CP from *Cellulomonas uda* as a starting point; this CP shows extremely low activity toward



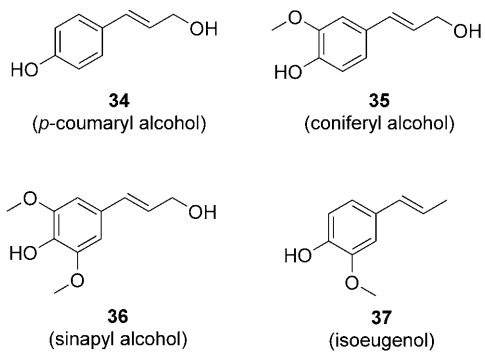
Scheme 32. Selective formation of α-D-galactose 1-phosphate (**32**) catalyzed by a cellobiose phosphorylase evolved on the basis of epPCR and ISM.[104]

lactose (**31**).[104] It should be noted that cellobiose and lactose differ only in the configuration of the C4-hydroxy group. CP from *C. uda* is mechanistically a well-studied enzyme which has been expressed abundantly in *E. coli*. Although its crystal structure has not been solved, a reliable homology model was constructed, thereby making ISM possible. However, a different strategy was initially chosen, namely epPCR in a restricted manner by focusing the amino acid exchange events to the region between Thr216 and Val757, which includes all residues within 15 Å of the catalytic center. In this restricted random mutagenesis, about 10^5 DNA clones were constructed with an average mutation frequency of 6.4 DNA mutations/kb. The mutations were randomly distributed, but they were strongly biased toward GC→N mutations (89%). The most active mutant with respect to lactose activity was variant LP1 with six point mutations: Ala397Val/Thr508Ala/Ala512Thr/Asp557Asn/Asn667Thr/Gly681Ser. It showed a threefold increase in lactose-specific activity compared to WT CP.[104] Partial deconvolution of the six point mutations revealed that some of them are superfluous, which does not come as a surprise when using epPCR (Section 2).

Attention was then turned to the double mutant Thr508Ala/Asn667Thr, which had become accessible by the deconvolution process and which showed reasonable lactose activity. Positions 508 and 667, being next to the binding pocket as shown by the homology model, were singled out for ISM experiments. In this particular case, NNS codon degeneracy (S: cytosine/guanine) which encodes all 20 amino acids and involves 32 codons and one stop codon was chosen rather than NNK, because it is better suited for the applied expression system. In an initial experiment, position 667 was subjected to saturation mutagenesis using the WT CP gene, with lactose and KH_2PO_4 serving as the substrates. The screening of 96 transformants furnished an improved mutant Asn667Ala which proved to be considerably more active toward lactose than WT CP and at least twice as active as the single mutant Asn667Thr. ISM was then initiated by using the gene of variant Asn667Ala as a template for constructing a saturation mutagenesis library at site 508. This led to the discovery of the double mutant Asn667Ala/Thr508Ile, which showed increased LP and CP activities. Importantly, it was shown to have 7.5-times higher specific activity toward lactose than the WT enzyme.[104] Kinetic studies revealed that the enhanced LP activity was due to a higher k_{cat} value. A reasonable model was proposed to explain these intriguing results. A more extensive ISM search can be expected to provide even better mutants for industrial applications.

3.4.8. Modulating Lignin Biosynthesis for Better Utilization of Plants in Paper Making, Biofuel Production, and Agriculture

The plant component lignin is an irregular racemic biopolymer composed of hydroxylated and *O*-methylated phenylpropane units that is formed mainly by the oxidative coupling of three different hydroxycinnamyl alcohols (monolignols), namely *p*-coumaryl (**34**), coniferyl (**35**), and sinapyl



alcohol (**36**).[105] They differ in the degree of *O*-methylation at the *meta* positions as a result of a process catalyzed by *O*-methyltransferases in the biosynthesis of lignin. Free *para*-hydroxy groups are required for the oxidative formation and coupling of phenoxy radicals that leads to the formation of the biopolymer. The relative amount of the chemically more labile *S*-lignin needs to be increased for a better practical utilization of plant materials in paper making, biofuel production, and agriculture. In principle, this can be achieved by creating a 4-*O*-methyltransferase which catalyzes the methylation of *para*-hydroxy groups of the precursor monolignols. The challenge of modulating lignin biosynthesis was the subject of a recent study by Bhuiya and Liu, where the desired *para* regioselectivity of a 4-*O*-methyltransferase as a catalyst in the *O*-methylation of monolignols was engineered by using directed evolution through ISM.[106] For screening purposes isoeugenol (**37**) was used.

Isoeugenol 4-*O*-methyltransfer (IEMT) from *Clarkia breweri* was selected as the enzyme to be engineered, because it was known to catalyze the 4-*O*-methylation of structurally related isoeugenol (**37**). To change the substrate scope so that the monolignols were also accepted, while maintaining the desired regioselectivity, the authors first tried several rational approaches, including swapping of sequence fragments, but to no avail. Then a homology model was constructed from the X-ray structure of a different *O*-methyltransferase, thereby allowing "plastic" sites for saturation mutagenesis to be chosen. Seven residues were saturated separately and screened for coniferyl alcohol (**35**) activity.[106] From a set of 30 active variants, two were identified which were especially prominent: Thr133Leu and Glu165FPhe. Several ISM experiments were performed to increase the activity further while maintaining the *para* regioselectivity. The Glu165Phe variant of IEMT was used as a template for randomization at the other six sites, and Thr133Leu was utilized as a starting point for saturation mutagenesis at position 165, which led to a range of improved double mutants. Some showed the serendipitous combination of single mutations, for example, Thr133Leu/Glu165Phe and Thr133Leu/Glu165Ile. These were then utilized in a final ISM round by visiting sites 139 and 175, which provided, inter alia, the triple mutant Thr133Leu/Glu165Ile/Phe175Ile as the best catalyst. Kinetic studies showed that more than 70-fold increases in activity for **35** and **36** relative to the WT was achieved, while maintaining full *para* selectivity. The reshaped binding pocket is shown in Figure 9, and features the mutations at sites 133, 165, and 175. It would be interesting to test an alternative ISM strategy in which the single residues are grouped into double or triple residue sites. The present results, which required only very small libraries, are already impressive, thereby opening the door for application in plants.[106]

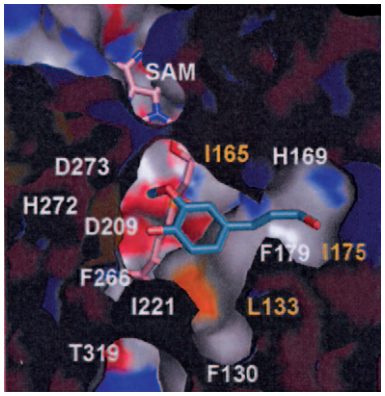


Figure 9. Binding pocket of IEMT reshaped by an ISM process. The mutations of the best triple mutant Thr133Leu/Glu165Ile/Phe175Ile (yellow/orange) and the docked-in substrate **34** (blue/red) are shown.[106]

3.4.9. Converting a Cyclodextrin Glucanotransferase into an α-Amylase

In a revealing study featuring the comparative virtues of epPCR and saturation mutagenesis, Dijkhuizen and co-workers converted a cyclodextrin glucanotransferase into an α-amylase.[107] Mutants displaying switched selectivity at high activity were obtained by using a series of small saturation mutagenesis libraries originating from sites aligning the binding pocket, screening between 320 and 720 trasnsformants, and applying ISM. In contrast, much larger epPCR libraries showed hits having only moderately improved catalytic profiles.

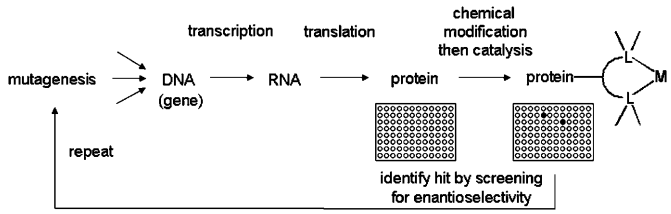
3.4.10. Inverting the Enantioselectivity of the Esterase from Burkholderia gladioli

In a recent study Schwab and co-workers succeeded in inverting the enantioselectivity of the esterase from *Burkholderia gladioli* (EstB) as a catalyst for the hydrolytic kinetic resolution of methyl-β-hydroxyisobutyrate.[108] They demonstrated once again that a semirandom approach guided by the X-ray structure of the enzyme constitutes a powerful approach to fast laboratory evolution, in this case by first

considering epPCR and then turning to the ISM strategy. Whereas WT EstB shows low *S* selectivity in the model reaction, several rounds of saturation mutagenesis at different amino acid positions in the neighborhood of the binding pocket provided a mutant that showed a selectivity factor of $E = 29$ in favor of the *R* product. Studies of this kind are important, because organic chemists are in need of both *R*- and *S*-selective catalysts on an optional basis.

3.4.11. Application of ISM in the Directed Evolution of Hybrid Catalysts

A natural limitation of biocatalysis is the fact that enzymes cannot catalyze the many types of asymmetric transformations that are possible by transition-metal complexes (Section 1).[1,2] To close this gap we proposed some time ago the concept of the directed evolution of enantioselective hybrid catalysts (Scheme 33).[109] It has been known for decades that it is possible to anchor a ligand/metal entity to a



Scheme 33. Concept of directed evolution of hybrid catalysts showing the flow of manipulated genetic information from the gene to transition-metal hybrid catalysts.[109,112,113]

host protein.[110] The first example was the Whitesides system, in which a biotinylated diphosphine/rhodium complex was allowed to interact with avidin.[111] The resulting supramolecular species was used as a catalyst in the asymmetric hydrogenation of a prochiral olefin. Although the enantioselectivity proved to be moderate (33–44% *ee*), this seminal study inspired many chemists to devise related systems.[110,111b] Unfortunately, such procedures deliver only a single catalyst, and enantioselectivity is then a matter of serendipity. In contrast, the concept illustrated in Scheme 33 offers a solution to this fundamental problem, namely to apply directed evolution to hybrid catalysts, which would mean that the enantioselectivity of synthetic transition-metal catalysts can be tuned by applying Darwinian principles on a genetic level.[109,112,113] Parallel to pursuing this concept, Ward and coworkers have cleverly developed the Whitesides system by chemically varying the length of the spacer between the rhodium ligand and the biotin as well as by using site-specific mutagenesis.[111b]

Several requirements need to be considered when developing a practical system for the Darwinian concept (Scheme 33).[109c,112,113] The expression system of the protein host must be prolific, because more catalyst must be present in the wells of the microtiter plates used in directed evolution than in the normal enzyme-catalyzed cases. This arises because of the considerably lower activity of synthetic transition-metal catalysts. Following expression, a simple high-throughput purification step is needed. Finally, bioconjugation must be essentially quantitative. Although not all of these prerequisites were fulfilled, we were able to present proof-of-principle by employing the Whitesides system.[112] The missing prerequisite was a sufficiently prolific expression system of avidin or streptavidin, which has not yet been developed. Consequently, we were forced to use large flasks instead of the usual wells of microtiter plates, and therefore the formation of only minilibraries was possible. Nevertheless, it was possible to apply ISM successfully by using NDT codon degeneracy, with three steps leading to the stepwise improvement of the enantioselectivity from 23% *ee* to 65% *ee* in an asymmetric olefin hydrogenation.[112] Further research by utilizing more appropriate protein hosts is necessary in this exciting new research area, especially by focusing on other types of asymmetric transition-metal-catalyzed reactions. This young research field was reviewed recently.[109c]

4. Saturation Mutagenesis at Sites Distal to the Binding Pocket

4.1. Knowledge-Based Approaches to Saturation Mutagenesis at Distal Sites in Enzymes

Thus far, the discussion has centered around mutagenesis sites which align the binding pocket of the enzyme (CASTing). Nevertheless, the question arises whether saturation mutagenesis can be applied at distal sites to give hits with enhanced stereoselectivity and/or altered substrate scope. It is well known that epPCR and DNA shuffling may lead to point mutations at remote sites[4,25,28,33,58,114] (Section 2), but in the case of saturation mutagenesis a structure-based approach is required. The challenge is to develop reliable criteria for choosing appropriate randomization sites, this time at remote regions of the enzyme. We have proposed two different strategies, where the dynamics of enzymes are involved in both cases. One is to focus on second-sphere residues, that is, on amino acids not in direct contact with the substrate bound at the active site but still within an approximate distance of 10–15 Å, which could lead to a reshaped binding pocket as part of extended CASTing.[42]

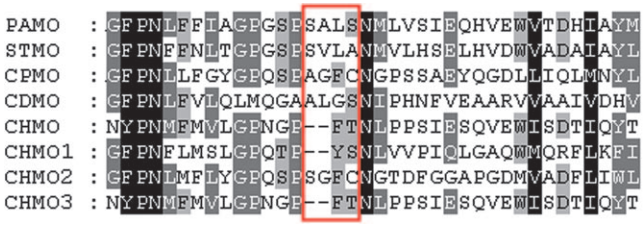
The second approach is defined by the rational choice of remote randomization sites which induce allosteric effects as a result of predictable domain movements in the absence of an effector.[73] The binding pocket of the enzyme would then acquire a new shape, thereby leading to different activity and enantioselectivity. This is a challenging exercise, because the choice of the distal sites requires a detailed analysis of the three-dimensional structure of the enzyme as well knowledge-guided predictive skill.

4.2. Saturation Mutagenesis at Second-Sphere Residues

PAMO is the first and thus far only thermostable Baeyer–Villiger monooxygenase (BVMO; Section 2). This enzyme offers the possibility to circumvent the use of whole cells by

switching to an in vitro procedure by employing the isolated enzyme in combination with an NADPH regenerating enzyme such as a secondary alcohol dehydrogenase.[38–41] Since PAMO accepts essentially only phenylacetone and related linear aryl ketones, the goal was to alter this robust enzyme so that a synthetically attractive substrate scope emerges. Protein engineering of cofactor-dependent enzymes is more challenging than studies of enzymes lacking cofactors, particularly because the former types are known to undergo dynamic structural changes during catalysis. Our first attempt at engineering improved PAMO mutants was directed toward shortening part of a loop from position 441 to 444 (bulge) that aligns the binding pocket (Section 2, Figure 2); however this led to only partial success.[40]

The second attempt was optimization of the loop by CASTing, again in the 441–444 segment comprising first-sphere residues.[41] Randomization of a four-residue site by using NNK codon degeneracy would require the screening of 3.1 million transformants for 95% coverage (Section 3.3, Table 2). The application of NDT codon degeneracy could be an alternative, since it requires only 62000 transformants for 95% library coverage. Grouping the four residues into segments is also an option. Nevertheless, in an attempt to beat the numbers problem in directed evolution even more, we opted for a bioinformatics approach by first aligning eight BVMOs in the loop region (Scheme 34). A limited number of



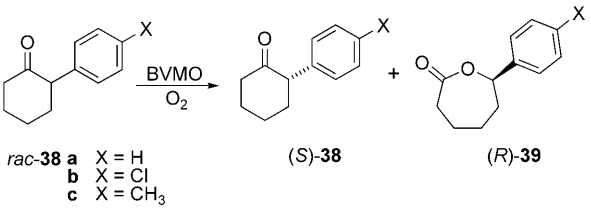
Scheme 34. Sequence alignment of eight different BVMOs (the loop segment 441–444 is shown in a red box).[41]

amino acids are conserved at the four positions 441, 442, 443, and 444, namely Ser/Ala, Ala/Val/Gly/Leu, Leu//Phe/Gly/Tyr, and Ser/Ala/Cys/Thr, respectively. Appropriate codon degeneracies were chosen which match as well as possible the amino acids occurring at the four positions while also introducing a limited number of additional amino acids for slightly higher structural diversity (Table 4).[41]

It was of particular synthetic interest to expand the substrate scope of PAMO so that 2-aryl cyclohexanone derivatives such as *rac*-**38a–c** are transformed to chiral lactones **39a–c** at an acceptable rate and with high enantioselectivity (Scheme 35). WT PAMO is a poor catalyst, with the conversion of ketone *rac*-**38a** being less than 10% even when using large amounts of enzyme for extended reaction times of two days and with the enantioselectivity in the oxidative kinetic resolution amounting to only $E = 1.2$ in slight favor of (*S*)-**39a**. About 2600 transformants would have to be screened for 95% library coverage, but in this case only 1700 were considered. This bioinformatics-guided strategy led to several slightly improved *S*-selective mutants, for example,

Table 4: Choice of codon degeneracies at each position in the 441–444 loop of PAMO.[41] Degenerate codons: A (adenine); B (cytosine/guanine/thymine); C (cytosine); G (guanine); S (cytosine/guanine); for the definition of K and N, see text.

Amino acid positions	Codon degeneracy	Encoded amino acids	Codons	Oversampling for 95% coverage
441	KCA	A, S		
442	KBG	S, A, L, V, W, G		
443	BGC	F, H, L, V, Y, G, D, R, C	864	2587
444	NSC	S, A, P, T, R, G, C		



Scheme 35. Oxidative kinetic resolution catalyzed by PAMO mutants.[41]

Ala441/Leu442/Val43/Ala444 ($E = 2.9$; 36% conversion/24 h), as well as a set of mutants showing complete reversal of enantioselectivity in favor of (*R*)-**39a**, for example, Ala441/Trp442/Tyr443/Thr444 ($E = 70$; 46% conversion/24 h).[41] In both these mutants all of the four original amino acids have been exchanged, but other *R*-selective mutants are characterized by only two or three substitutions. In the case of substrate *rac*-**38b**, some of the *R*-selective mutants discovered in the initial screening process were tested as catalysts in oxidative kinetic resolution without performing any additional high-throughput screening. In this case, the results proved to be even better, with five mutants showing *E* values exceeding 130, for example, Ala441/Trp442/Tyr443/Thr444 ($E > 200$; conversion 59%/24 h). It was also shown that the introduction of the point mutations does not compromise thermostability.[41]

This study demonstrates the utility of a bioinformatics-guided approach to saturation mutagenesis, but these results proved to have limitations. For example, the *p*-tolyl analogue *rac*-**38c** was not accepted by any of the evolved mutants. One possibility for future exploration would be to group the four residues into two sites A and B, each comprising two amino acid positions, and then to apply ISM systematically to form and screen four saturation libraries by using the previous or other codon degeneracies. This kind of systematization would entail the construction of two different ISM pathways, A→B and B→A, which could theoretically lead to the same or different mutants with improved catalytic profiles.

At this point we decided to postpone such an ISM study, which would include all 441–444 residues and possibly other

sites aligning the binding pocket. Instead, curiosity led us to consider second-sphere residues.[42] It is actually not trivial to identify with certainty such sites because of the dynamics of the enzyme. Therefore, in the hope of obtaining a more realistic picture of the occupied binding pocket, as opposed to the published crystal structure in the absence of substrate/inhibitor, we performed induced-fit docking experiments of WT PAMO by using phenylacetone as the substrate.[42] The result indicated that the bound substrate and the catalytically important Arg337 that stabilizes the Criegee intermediate are both located on the *re* side of the flavin (FAD). Inspection of the docking result suggested several putative second-sphere residues spatially located directly beyond the 441–444 loop segment, for example, Pro440 and Pro437 (Figure 10).

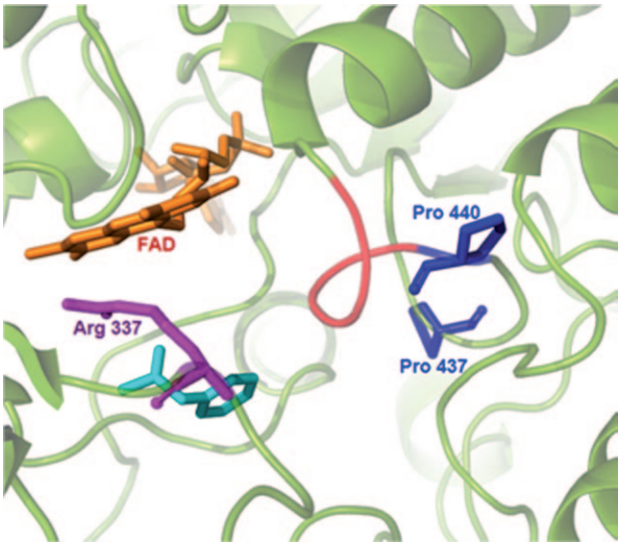
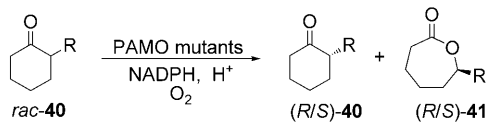


Figure 10. The putative binding pocket of PAMO based on the induced-fit docking model.[42] Phenylacetone and the loop segment 441–444 are shown in cyan and red, respectively, and Pro437 and Pro440 are labeled in blue.

Rather than using 2-phenylcyclohexanone (*rac*-**38a**) as the model substrate, we decided to address the more difficult problem regarding the BV reaction of 2-alkyl-substituted derivatives *rac*-**40a–i**, which are not accepted by WT PAMO (Scheme 36). Ketone *rac*-**40b** was chosen for the screening process, in the hope that an improved mutant would also function as a catalyst in the kinetic resolution of the other substrates. Two separate focused libraries were generated by



a R = Methyl
b R = Ethyl
c R = *n*-Propyl
d R = *n*-Butyl
e R = Allyl
f R = Isopropyl
g R = Cyclohexyl
h R = Benzyl
i R = CH_2CH_2CN

Scheme 36. Oxidative kinetic resolution of *rac*-**40a–i** catalyzed by PAMO mutants.[42]

saturation mutagenesis at positions 437 and 440 by using NNK codon degeneracy, which required in each case the screening of only 200 transformants to ensure > 95% library coverage.[42]

Whereas the library generated by saturation mutagenesis at position 437 failed to contain significantly improved mutants in the oxidative kinetic resolution of *rac*-**40b**, randomization at second-sphere position 440 resulted in the identification of a surprisingly large number of improved hits, ten of which were sequenced. Of these, five showed notably enhanced enzyme activity toward this otherwise inert substrate, with the enantioselectivity also being excellent (Table 5).[42] Since conventional saturation mutagenesis is known to occur with some degree of amino acid bias, it is possible that some potential hits were not formed. Therefore, all the "missing" mutants were prepared by site-specific mutagenesis and tested in the model reaction of **40b**→**41b**. Of the nine relevant cases, namely Pro440Ala, Pro440Val, Pro440Ser, Pro440sp, Pro440Glu, Pro440Gln, Pro440Tyr, Pro440Trp, and Pro440Met, two new mutants proved to be active, namely Pro440Tyr and Pro440Trp (Table 5, entries 6 and 7). It can be seen that in this case that the application of saturation mutagenesis coupled with theoretically sufficient oversampling for ensuring > 95% library coverage results in the identification of 80% hits, which is an acceptable score. This kind of information has not been generated previously, and it should be kept in mind in future saturation mutagenesis studies.

Table 5: Kinetic resolution of *rac*-**40b** by using PAMO mutants derived from a saturation mutagenesis library (entries 1–5) or site-specific mutagenesis (entries 6 and 7) both at position 440.[42]

Entry	Mutant	*E* value	K_m [mM]	k_{cat} [s^{-1}]	k_{cat}/K_m [$M^{-1} s^{-1}$]
1	Pro440Phe	26	0.89	1.2	1300
2	Pro440 Leu	> 200	1.6	0.72	450
3	Pro440Ile	> 200	2.7	0.66	240
4	Pro440Asn	34	2.2	1.5	680
5	Pro440His	> 200	1.0	0.83	830
6	Pro440Trp	> 200	1.3	1.3	1000
7	Pro440Tyr	95	1.9	1.1	580

The best mutants were subsequently tested as catalysts in the oxidative kinetic resolution of the other 2-alkyl cyclohexanone derivatives (Scheme 36). In all cases high activity and excellent enantioselectivities were observed ($E = 30$–200). Even the bulky 2-cyclohexylcyclohexanone derivative *rac*-**40g** reacts rapidly with high enantioselectivity, for example, $E = 69$ (mutant Pro440Ile), $E = 119$ (mutant Pro440Tyr), and $E = > 200$ (mutant Pro440Leu). Certain bicylic ketones are also oxidized with pronounced enantioselectivity. Finally, the point mutations leading to relatively wide substrate scope (activity) and high enantioselectivity do not impair the robustness of the PAMO mutants, namely, there is no trade-off with respect to thermostability.[42]

The above data were obtained in an in vitro system in which isolated PAMO mutants were used in combination with the equally robust secondary alcohol dehydrogenase from

Thermoanaerobacter ethanolicus for NADPH regeneration, with isopropanol serving as the reductant. This study constitutes a major step to more economically and ecologically viable BVMO-mediated stereoselective BV reactions in organic chemistry and biotechnology without the need to use whole cells.[42]

It should be noted that industry may well prefer whole cell systems, but even in such cases BVMOs with high thermostability are a significant advantage, because the lifetime of the system is increased under the operating conditions. All of the laboratory evolution work was performed with a single substrate (*rac*-**38b**) by applying saturation mutagenesis at two sites, which required a screening effort of only 400 transformants.[42] It remains to be seen whether saturation mutagenesis at second-sphere sites as part of CASTing can be generalized as a useful strategy in directed evolution by extension to other types of enzymes, and whether ISM in combination with this approach can be applied successfully in reaching this goal.

4.3. Allostery-Inducing Distal Mutations Generated by Saturation Mutagenesis

Protein allostery is defined as a positive or negative cooperative event, which leads to a structural change at the binding pocket as a result of remote docking of a molecule that acts as an effector.[115] Many different experimental and computational techniques, including directed evolution, have been utilized to unravel the intricacies of this phenomenon.[116] A very different concept is to induce allosteric communication by appropriate mutational changes at a remote site, thereby altering the structure and dynamics at the actual active site in the absence of an effector.[73] This kind of allosteric effect would be expected to cause notable structural rearrangements, such as domain movements, in the enzyme with concomitant changes in the catalytic profile. Remote effects in the evolution of enantioselectivity and/or activity have been noted before,[4,25,28,33,58,114] but these have not been linked to allosteric effects. The concept is not as unlikely as one may think, since related effects have in fact been discovered in nature, specifically in the case of certain diseases caused by mutations which in turn induce allosteric communication. These can shut down the functionality of allosteric enzymes, or they can cause conformational changes at the active site which lead to constitutive activation regardless of whether an effector is bound, as in the case of the G protein.[115a] Nevertheless, it is not a trivial task to identify remote sites in an enzyme at which saturation mutagenesis can be expected to generate mutations that trigger allosterically induced rearrangements with creation of a structurally altered binding pocket in the absence of an effector.

To test this possibility we chose once more phenylacetone monooxygenase (PAMO) and analyzed carefully the published crystallographic features of this enzyme[39] in the hope of locating appropriate randomization sites.[73] Suggestions about movements in the NADP-binding domain believed to be coupled to the catalytic mechanism of PAMO[39] led us to consider mutations strategically located in regions distal from the active site which may induce the required structural rearrangements. Close inspection of the PAMO crystal structure taught us that the number of likely possibilities is in fact limited. One option would be to focus on hinge regions, but in the case of PAMO and other BVMOs, these are characterized by highly conserved sequences. We therefore turned to a different strategy by searching for potential randomization sites at or near two domain interfaces. For example, residues at the interface between the NADP-binding domain and the PAMO helical domain could be chosen in the expectation that the correct mutations would lead to strong attractive interactions with concommitant structural reorganization. However, a different possibility appeared more promising, namely to identify hot spots in PAMO which would bring the NADP- and FAD-binding domains spatially closer together (Figure 11).[73]

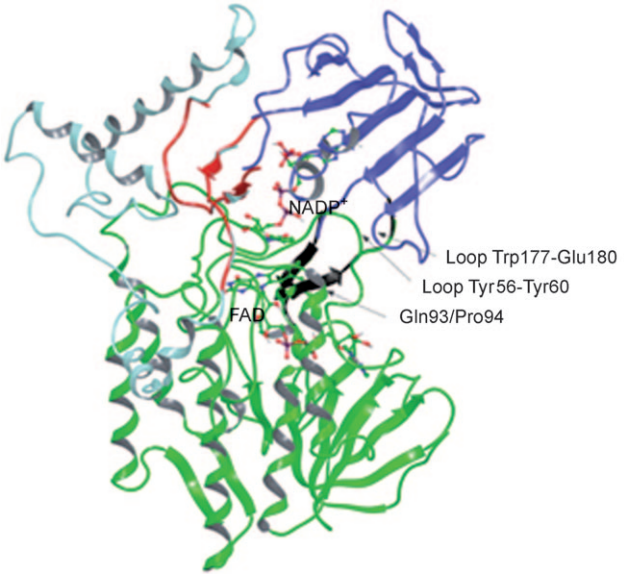


Figure 11. Representations of the different domains in the PAMO structure[73] (FAD domain: green, NADP-binding domain: dark blue, helical domain: light blue). The NADP- and FAD-binding domains are linked by two antiparallel β strands in a hinge-like arrangement (black). The NADP-binding and helical domains can be considered as separate entities connected by two segments (red). The chosen randomization Gln/Pro94 in the N-terminal region of the α helix is also shown.

The WT PAMO X-ray structure[39] reveals several already attractive contacts between the two domains that are structurally not far from the hinge region.[73] They occur between the loop segment Trp177–Glu180 (NADP-binding domain) and the loop segment Tyr56–Tyr60 (FAD-binding domain), with the π-cation interaction between His179 and Arg59 being an example. Located directly behind and in loose contact with the loop segment, Tyr56–Tyr60 is the N-terminal region (Ala91-Glu95) of an α helix, which by nature constitutes a rigid secondary element. We anticipated that appropriate mutations in this part of the helix could lead to strong attractive interactions with the loop segment Trp177–

Glu180 (Figure 11).[73]Consequently, Gln93/Pro94 was chosen as the randomization site (Figure 12). This decision was also guided by the realization that the neighboring amino acid at position 92 is the last member of a long loop that extends parallel to loop segment Asp–Tyr, which is in direct contact with the flavin ring. Docking simulations with phenylacetone bound in the binding pocket of WT PAMO indicated that the substrate is about 18 Å away from position 93.[73]

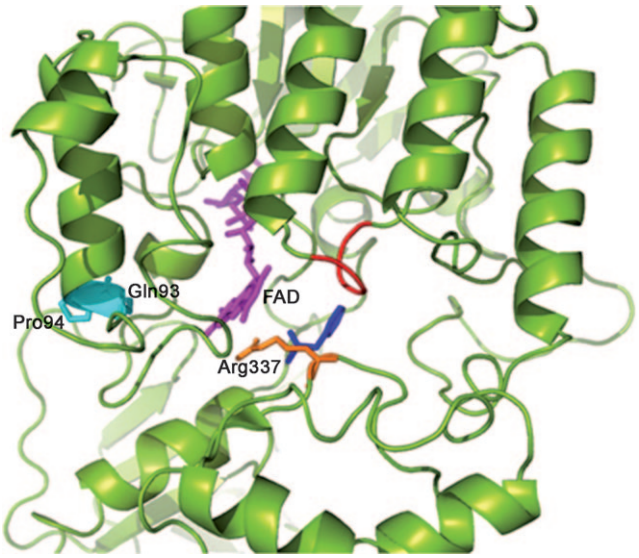


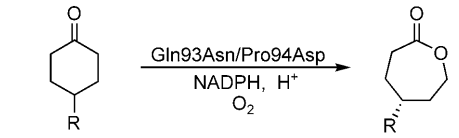
Figure 12. Excerpt of the X-ray structure of PAMO showing the chosen randomization site Gln93/Pro94 (cyan), FAD (magenta), Arg337 (orange), phenylacetone (gray), and loop (red) targeted in previous mutagenesis studies.[73]

Following this analysis, amino acid positions 93/94 were randomized simultaneously by using NDT codon degeneracy encoding a reduced amino acid alphabet of 12 members (Section 3.3), which required an oversampling of 400 transformants for 95% library coverage. About 400 clones were screened for activity and enantioselectivity by using automated gas chromatography. The model reaction was the oxidative kinetic resolution of 2-ethylcyclohexanone (*rac*-**40b**; Scheme 36) in conjunction with an NADPH-regenerating thermostable secondary alcohol dehydrogenase and isopropanol as the reductant in an in vitro process. Two hits were identified, with the more active one being mutant Gln93Asn/Pro94Asp, which was then utilized for all further studies. This double mutant displays a selectivity factor of $E = 50$ in the model reaction of *rac*-**40b** in favor of (*S*)-**41b**. Remarkably, the substrate scope of PAMO has been extended considerably, since all other substrates in Scheme 36 and 37 are also accepted, generally with good to excellent enantioselectivity (Table 6 and Scheme 37).[73] Further improvements in activity can be envisioned by invoking ISM, DNA shuffling, or combinations thereof.

The catalytic potential of PAMO mutant Gln93Asn/Pro94Asp was also tested in the desymmetrization reactions of 4-substituted cyclohexanone derivatives **42a–c** with formation of lactones **43a–c** (Scheme 37). It was found that

Table 6: Oxidative kinetic resolution of ketones **38a,c** and **40a–i** by using PAMO mutant Gln93Asn/Pro94Asp.[73] Note that the *R*/*S* assignment changes in accordance with the Cahn–Ingold–Prelog convention; the absolute configuration is as shown in the formulas.

				Residual ketone		Lactone		
Entry	Substrate	Specific activity [U mg^{-1}]	Conv. [%]	*ee* [%]	Abs. conf.	*ee* [%]	Abs. conf.	*E* value
1	*rac*-**38a**	64	45	38	*S*	95	*R*	92
2	*rac*-**38c**	40	17	2.7	*S*	18	*R*	1.5
3	*rac*-**40a**	35	15	46	*R*	91	*S*	25
4	*rac*-**40b**	49	21	25	*R*	95	*S*	50
5	*rac*-**40c**	42	37	37	*R*	95	*S*	68
6	*rac*-**40d**	51	36	14	*R*	92	*S*	40
7	*rac*-**40e**	24	23	19	*R*	94	*R*	42
8	*rac*-**40f**	39	15	17	*R*	>99	*R*	>200
9	*rac*-**40g**	48	42	43	*S*	86	*R*	25
10	*rac*-**40h**	58	43	58	*S*	98	*R*	>200
11	*rac*-**40i**	52	11	17	*S*	96	*R*	55



42a R = Methyl
b R = Ethyl
c R = *n*-Butyl

43a *ee* = 98% (*R*)
b *ee* = 98% (*R*)
c *ee* = 97% (*R*)

Scheme 37. Desymmetrization of prochiral ketones **42a–c** by using PAMO variant Gln93/Pro94Asp.[73]

excellent enantioselectivity was achieved in all cases. It is interesting to note that the absolute configuration of the lactone products is uniformly opposite to what is observed when using the thermolabile cyclohexanone monooxygenase (CHMO) as the catalyst.

Importantly, the thermostability of WT PAMO, which is crucial for industrial applications, is maintained in the double mutant Gln93Asn/Pro94Asp. Deconvolution experiments showed that the respective single mutants Gln93Asn and Pro94Asp failed to show any measurable activity as catalysts in the model reaction of *rac*-**40b**.[73] This finding proves that a strong cooperative effect is operating in the double mutant. If a different strategy had been chosen in this case, namely to target sites 93 and 94 separately in the hope of finding hits for ISM, then such an approach would have failed.

Molecular dynamics (MD) simulations of WT PAMO and of the superior double mutant Gln93Asn/Pro94Asp were carried out to uncover the source of enhanced activity and thus check our original postulate regarding possible allosteric communication induced by mutations. Space does not allow a detailed account of the results and analysis thereof.[73] Suffice it to say that significant domain movement occur, which lead to a very different shape of the binding pocket. A comparison of the partially covered binding pocket of WT PAMO with the more exposed and reshaped binding site of the double mutant is shown in Figure 13.[73] Interestingly, these conformational differences are similar to the structural differences

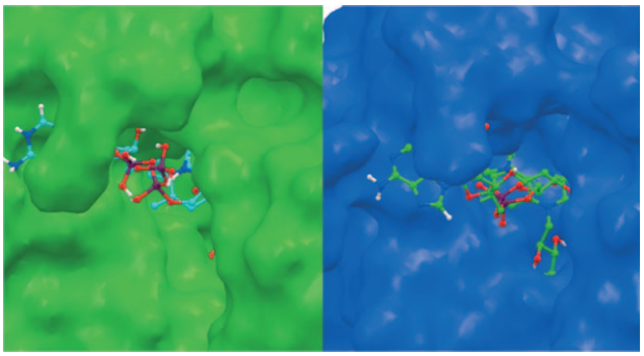


Figure 13. Close-up of the binding pocket of WT-PAMO (left) and of superior mutant Gln93Asn/Pro94Asp derived from MD simulations.[73]

between two separately determined crystalline forms of a CHMO from an environmental *Rhodococcus* strain with bound FAD and $NADP^+$.[117] The two reported conformers correspond to a closed and a more open state. The interconversion results in domain shifts with concomitant sliding of the cofactor $NADP^+$, much like the mutation-induced allosteric effect in our system. MD simulations showed that the combined action of the two point mutations in the double mutant Gln93Asn/Pro94Asp not only orchestrates allosterically distinct structural changes globally, but also locally at the binding pocket.[73]

Finally, covariance maps were constructed of the double mutant and of WT PAMO to check the conclusions derived from the experimental results and the MD simulations.[73] Such maps are known to be useful indicators of correlated and anticorrelated motions in enzymes and are consequently constructed to study the dynamics of proteins.[116,118] In our case, the covariance map of the double mutant Gln93Asn/Pro93Asp indeed showed increased correlation between the mutation site and the FAD segment as well as between the two cofactors, relative to the situation in the respective covariance map of WT PAMO. This finding supports the original model regarding mutationally induced allostery. Moreover, significant structural changes are indicated in the two covariance maps at the respective active sites—at the hinges and at the NADP-binding domains—which is also in line with our mechanistic conclusions.[73]

In conclusion, this study shows that the shape of a binding pocket can be altered significantly by focusing saturation mutagenesis at a remote site which can be expected to induce allosteric communication within the enzyme in the absence of an effector molecule. This approach is very different from the usual and more straightforward CASTing procedure. Similar strategies can be envisioned for other types of enzymes, for example, P450s.

5. ISM in the Quest To Enhance Thermostability of Enzymes

As delineated in the preceding sections, enantioselectivity and/or substrate acceptance can be controlled by saturation mutagenesis in single or preferably iterative modes, with the proper choice of the randomization sites being guided by structural data. A very different, yet equally important enzyme property is thermostability.[119] Since ISM had emerged as an efficient method to control enantioselectivity and substrate scope, it was of interest to devise an analogous iterative procedure to enhance the thermostability of proteins. This challenge required the development of a criterion for choosing appropriate randomization sites, which can be expected to be different from CAST sites. We suggested sites displaying the highest B factors, which are known to reflect maximal smearing of atomic electron densities relative to equilibrium positions as a result of thermal motion and positional disorder. Since it was well known that hyperthermally stable enzymes are more rigid than mesophilic analogues,[119] we surmised that focused randomization at sites displaying the highest B factors should result in mutations that increase the rigidity and therefore stabilize the enzyme.[20,120] For this purpose we developed a computer aid dubbed B-FITTER,[20] and applied it to the thermostabilization of the lipase from *Bacillus subtilis*. A high degree of thermostabilization was achieved in five ISM steps by screening the supernatants.[20,120] The final mutant was purified by a short heat treatment to eliminate other proteins present; it showed an increase in the T_{50}^{60} value by 45 °C. Subsequently, the effect was attributed as being due to the creation of a communicating amino acid network on the surface of the enzyme.[121] More recently, the B-FIT approach was applied successfully to the thermostabilization of the epoxide hydrolase from *Aspergillus niger* (ANEH), with up to 21 °C stabilization being achieved.[77] B-FIT can also be applied when attempting the opposite, namely greater thermolability in a controlled manner, which in rare but nevertheless important cases is of practical significance.[122]

6. Lessons Learned from Directed Evolution of Enantioselective Enzymes

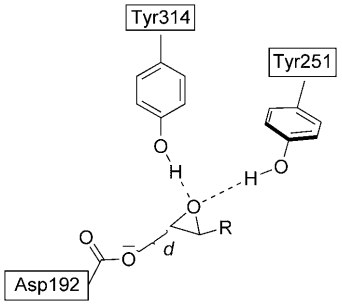
Two lessons can be learned from the studies highlighted in this Review. Firstly, the focus on methodology development in directed evolution allows one to compare mutagenesis methods and strategies. Evolving improved mutants is a practical goal, but understanding what makes one mutagenesis strategy more efficient than another is also important. In the case of ISM, it is the occurrence of strong cooperative effects between point mutations and sets of mutations along evolutionary pathways which makes this method so effective. Moreover, superfluous point mutations, which are a problem when using such methods as epPCR, appear not to occur with ISM.

The second lesson to be learned from any study on the evolution of enhanced enantioselectivity and/or substrate acceptance (rate) concerns the underlying factors responsible for improvement of the catalyst on a molecular level. This requires mechanistic studies, ideally including kinetics, X-ray data, and NMR investigations as well as theoretical analyses based on QM/MM calculations and MD simulations. One of the most thoroughly studied cases concerns the QM/MM investigation of the PAL mutant with six point mutations

which was evolved in our original work by epPCR and DNA shuffling.[32] A relay mechanism was proposed involving a remote mutation, which in the case of the preferred *S* substrate leads to additional stabilization of the respective oxyanion with a new hydrogen bridge, a situation that is not possible with the mirror image compound. Details can be found in the original publications[32] and in reviews.[58] Unfortunately, it was not possible to check this model by obtaining the crystal structure of the best mutant, a deficit which plagues all other studies concerning the directed evolution of enantioselective enzymes, with the exception of the following case.

A recent mechanistic and structural study of an improved enantioselective enzyme concerns the laboratory evolution of the epoxide hydrolase from *A. niger* as a catalyst for the hydrolytic kinetic resolution of *rac*-**19** (Section 3.2, Scheme 14).[76] WT ANEH is only slightly *S* selective ($E = 4.6$), while the best mutant LW202 evolved after five ISM steps leads to a selectivity factor of $E = 115$ (Scheme 15). Kinetic data, MD simulations, molecular modeling, inhibition experiments, and X-ray structural analyses of WT ANEH and of mutant LW202 uncovered the source of enhanced enantioselectivity at all five stages of the evolutionary process.[76] Earlier structural studies of WT ANEH, which included its X-ray structure in the absence of a substrate or inhibitor,[75] had defined the basic mechanistic features.[75,123] Two tyrosine residues at positions 251 and 314 at the end of a narrow tunnel-like binding pocket bind and concomitantly activate the substrate by forming hydrogen bonds with the epoxide O atom. This is followed by rate- and stereochemistry-determining nucleophilic attack of Asp192 at the sterically less hindered C atom, with subsequent rapid hydrolysis of the short-lived enzyme–ester intermediate (Scheme 38).

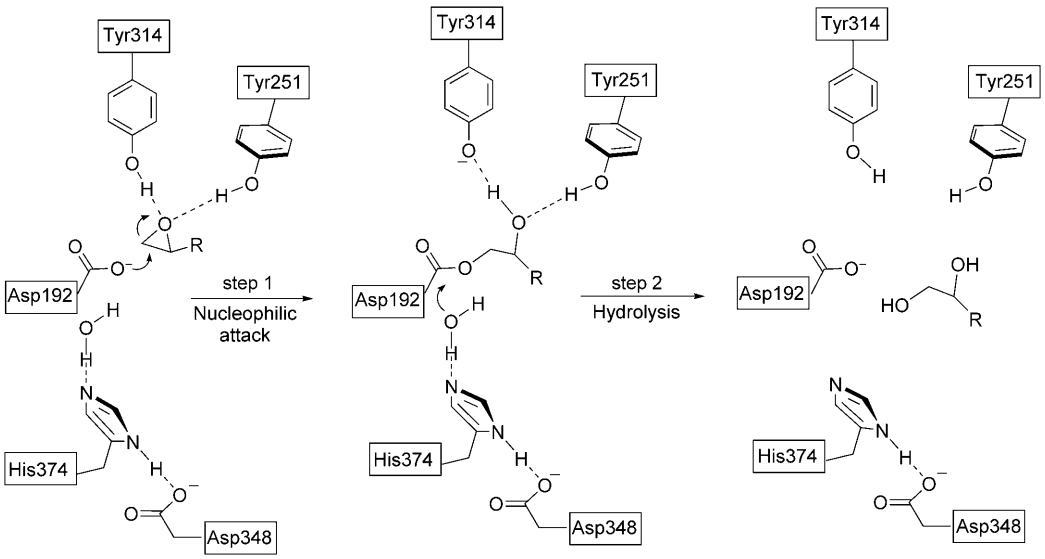
Michaelis–Menten kinetics of the superior mutant LW202 revealed that the mutational changes essentially shut down the reaction of the disfavored enantiomer (*R*)-**19**,[76] which is a close to ideal situation for kinetic resolution. This conclusion was corroborated by the experimentally determined k_{cat}/K_M values for WT ANEH and mutant LW202. We performed extensive MD simulations to study how the activated *R*- and *S*-configured substrates **19** are positioned in the narrow binding pocket, and to see if any substantial differences occur upon going from WT ANEH to LW202 along the five-step evolutionary pathway. The energetics of binding was expected to depend on the two hydrogen bonds originating from Tyr251 and Tyr314, as well as on other factors. We singled out the distance *d* between the attacking O atom of Asp192 and the epoxide C atom undergoing the S_N2 reaction as being crucial (Scheme 39).[76] A sufficiently small *d* value would mean that the substrate is ideally positioned for the rate- and stereoselectivity-determining reaction, perhaps in a near-attack conformer as discussed by Bruice in other enzyme-catalyzed reactions.[124]



Scheme 39. Definition of the distance *d* in the rate- and stereochemistry-determining step of ANEH-catalyzed reactions.[76]

The results of MD calculations using (*R*)- and (*S*)-**19** in separate simulations involving WT ANEH, with the four intermediate mutants LW081, LW086, LW123, and LW44 showing increased enantioelectivity as well as the final superior mutant LW202, point to a remarkable trend.[76] In the case of the preferred *S* substrate, the distance *d* remains more or less constant, whereas the corresponding value in the case of the *R* substrate increases as the evolutionary process proceeds, and reaches a maximum of $d = 5.4$ Å in LW202. This means that activation by Tyr251/Tyr314 is maintained, but the nucleophilic Asp192 is simply too far away for a smooth S_N2 reaction to occur. This analysis is in full accord with the kinetic data.

Fortunately, it was possible to obtain crystals of mutant LW202 suitable for an X-ray



Scheme 38. Reaction mechanism of ANEH.[75,122]

structural analysis.[76] Comparison with the crystal structure of WT ANEH showed that the gross features of the secondary and tertiary structures are essentially identical, but that the shape of the binding pocket of mutant LW202 has changed dramatically (Figure 14). In modeling studies, it was possible

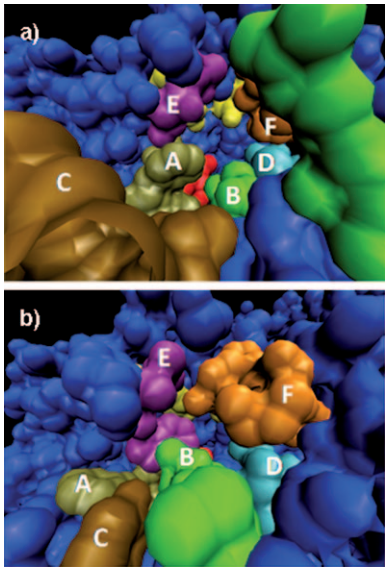


Figure 14. Results of X-ray structural analyses.[76] a) Empty binding pocket of WT-ANEH based on its X-ray structure. b) Empty binding pocket of mutant LW202 based on its X-ray structure. Catalytic Asp192: red, activating couple Tyr251/Tyr314: yellow; A–F indicate the respective randomization sites for CASTing (note that site A was not included in the evolutionary pathway B→C→D→F→E).[31]

to bind both (*R*)- and (*S*)-**19** into the WT binding pocket (Figure 14a) without causing any steric clashes, while maintaining hydrogen-bond activation by Tyr251/Tyr314. In sharp contrast, in the case of the genetically altered binding pocket corresponding to LW202 (Figure 14b), (*S*)-**19** binds again without any problems, whereas (*R*)-**19** causes severe steric clashes if activation by Tyr251/Tyr314 is maintained.

The model predicts that structurally different monosubstituted epoxides, but not *trans*-1,2-disubstituted analogues of the type *rac*-**21** (Scheme 19), should undergo enantioselective hydrolytic kinetic resolution catalyzed by mutant LW202. Indeed, 12 structurally different monosubstituted epoxides were shown to undergo kinetic resolution with good to excellent enantioselectivity ($E = 22$–90), whereas WT ANEH leads to poor results.[76] Directed evolution studies of this kind are valuable in a more general sense, because they contribute to a better understanding of how enzymes function in detail. In the present case, the gross mechanistic features of the epoxide hydrolase had been defined previously, but geometric details of substrate positioning and activation in the binding pocket have now become available.[76]

7. Conclusions and Perspectives

The application of biocatalysis in synthetic organic chemistry and biotechnology experienced a period of rapid expansion in 1970–1990,[2] but despite the advantages associated with high activity, mild conditions, and dispensability of protective groups, it has traditionally suffered from serious limitations. These include limited substrate scope and very often the poor enantioselectivity of enzymes, as well as in many cases insufficient protein stability under the operating conditions and sometimes product inhibition. During the last 15 years directed evolution has provided a means to solve these problems.[4] Following proof-of-principle studies concerned with increased stability toward denaturing organic solvents[7] and enhanced[8,9] or reversed[9b,33,45] enantioselectivity, further examples led to the generalization of this powerful form of protein engineering.[4] The laboratory evolution of stereoselective enzymes emerged as a fundamentally new approach to asymmetric catalysis, and was accompanied by numerous industrial processes.[58]

During the early phase of research on stereoselectivity, gene mutagenesis methods based primarily on epPCR, DNA shuffling, and saturation mutagenesis were applied,[8,9,33,58] but efficacy in probing protein sequence space was not the main focus of interest. This neglect, which originally characterized research in laboratory evolution,[7,119b,c] is understandable. During the last five years this picture has changed, because researchers began to consider "quality not quantity",[59g] which signaled the call for methodology development in directed evolution.[4,18,31,58–63] It is not just industry that is in need of "fast" and practical protein engineering methods, basic research leads the way.[4] Our contribution regarding this crucial goal is iterative saturation mutagenesis (ISM) as a means to control the stereoselectivity[18,31,58–60] and the substrate scope of enzymes,[18,58] as well as to increase the thermostability[20,77,120] of proteins in general. ISM has since been used to tune the binding properties of proteins,[99] to promote enzyme promiscuity,[97,104,107] and to modulate biosynthesis (pathway engineering) in green biotechnology.[106]

ISM is based on the formation of focused libraries that are generated by randomization at appropriately chosen sites in an iterative manner. We and other research groups had utilized saturation mutagenesis previously, with selected sites being chosen in a nonsystematic manner—in some cases going successively from one residue to another.[25,28,58,59,65,69] Thus, ISM is a systematization of saturation mutagenesis, which requires different criteria when choosing the correct randomization sites. We have proposed two different strategies in regard to enantioselectivity and/or substrate scope: Most often the systematic consideration of all the residues aligning the binding pocket (CASTing) is considered.[64] In some cases, so-called second-sphere residues not in direct contact with the binding pocket but spatially still relatively close can also be chosen.[20,42] The second approach involves randomization at distal residues that can be expected to induce allosteric effects.[73] ISM can also be employed when attempting to increase the stability of enzymes or proteins in general, in this case by restricting the randomization to residues showing high B factors (B-FIT method)[20,77,120]

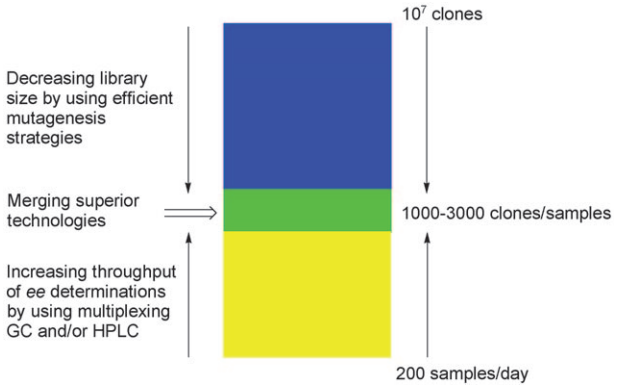
(Section 5). Of course, it is possible to utilize epPCR or DNA shuffling after ISM, or vice versa. Although all ISM studies thus far are characterized by high efficacy, this knowledge-driven protein engineering method has limitations in that X-ray structures or homology models need to be available; this prerequisite is fortunately fulfilled in most cases. If no such data is available, epPCR and DNA shuffling are the options of choice, or a bioinformatics approach is necessary to choose appropriate sites for saturation mutagenesis.

A highly useful tool when applying ISM concerns the use of reduced amino acid alphabets, as defined by the respective choice of codon degeneracy.[18,60,66] By using a statistical analysis with known algorithms[80b] we have shown that the degree of oversampling necessary to cover a certain percentage of a given mutant library depends upon the number of amino acids used as building blocks in the saturation mutagenesis: the less amino acids used, the less screening required.[60] Although a smaller amino acid library entails less structural diversity, the statistical advantages are substantial. Nevertheless, the opposite way by implementing an expanded genetic code with the introduction of unnatural amino acids may offer interesting perspectives in a different way.[125]

As an aid for designing saturation mutagenesis libraries, we have developed the computer programs CASTER for controlling the stereoselectivity and/or substrate scope and B-FITTER for increasing thermostability,[20,60] both are available free of charge on our homepage (http://www.mpi-muelheim.mpg.de/mpikofo_home.html). It is conceivable that ISM can be combined with other computer-aided approaches such as ProSAR[62] or MAP.[63d]

Understanding why a given mutagenesis strategy provides better results than an alternative constitutes a challenge in its own right. The source of the efficacy of ISM has been uncovered by a deconvolution strategy that enables the mapping of respective fitness landscapes. This allows analysis of all the epistatic interactions at all the stages of a multistep evolutionary process[60,72,121] (Section 3.2). This type of quality control has shown that ISM is not only characterized by additive effects, but is consistently accompanied by pronounced cooperativity (synergism) between single point mutations and between sets of mutations. This raises the question of how to group putative randomization residues into sites for saturation mutagenesis. Choosing single residue sites and working iteratively can be successful,[18,20] but double or triple residue sites are recommended,[31,60,72] because the probability of synergistic effects is then maximized. Superfluous sets of mutations generally do not occur, in contrast to other studies based on epPCR.[4,32b] The conventional and most improved methods for saturation mutagenesis are characterized by inherent amino acid bias. Several ways are possible to achieve a balanced ratio of amino acids, with the MAX-approach,[126] Sloning,[127] or "hand mixing" of designed oligonucleotides[127b] being some options.[77,69o] It remains to be seen whether the benefits outway the extra investment of time and financial means needed for the removal of redundancy in these approaches.

Some kind of a medium- or high-throughput *ee* assay has to be developed for each new directed evolution project,[10] some requiring expensive instrumentation. It is, therefore, our current goal to adapt the Trapp method[128] based on multiplexing GC (and HPLC) for the high-throughput analysis of chiral compounds.[77] We hope to reach throughputs of 1000–3000 samples or more per day. Since GC and HPLC belong to the standard equipment of chemical and biological laboratories, the screening problem would be solved in a general way once and for all, provided that smart mutagenesis strategies are applied as in ISM (Scheme 40).



Scheme 40. Merging efficient mutagenesis strategies for generating smaller but higher quality mutant libraries with increased *ee*-assay capacity on the basis of multiplexing GC and/or HPLC.

Two further points regarding directed evolution are worthy of note. In some cases enzymes cannot be expressed efficiently in the usual workhorses of molecular biology such as *E. coli*, which means that alternative expression systems have to be developed, with examples being horse radish peroxidase (yeast surface display)[57] and oxynitrilases (*Pichia pastoris*).[129]

Finally, ISM offers perspectives which go far beyond the control of substrate acceptance, stereoselectivity, and thermostability required in organic chemistry and white biotechnology.[99,106] An exciting future area of research is the application of the laboratory evolution of enantioselective enzymes to metabolic engineering. Influencing metabolic pathways[106,129] by altering the stereochemistry in a given step has the potential to access chiral compounds with different absolute configurations not previously possible, including the production of novel therapeutic drugs. ISM also appears to be a viable strategy for altering the binding properties of proteins[99] or peptides, which is of interest in a variety of areas including medical applications. ISM in the form of B-FIT[20,120] can also be applied to the thermostabilization of proteins needed in bionanotechnology or in organisms used in pollution control.

Note added in proof: ISM has been used to invert enantioselectivity of a P450 enzyme.[130]

I wish to express my sincere thanks to all past and present co-workers, whose creativity and perserverance have contributed to our research in directed evolution. Their names are listed in the references. Collaborations with K.-E. Jaeger, W. Thiel, M. Bocola, M. Mihovilovic, B. Dijkstra, H. Rabitz, W. Quax, S.

Mowbray, M. Arand, M. M. Kayser, and O. Trapp are also gratefully acknowledged. I also thank the MPG, DFG (Schwerpunkt 1170), German-Israeli Cooperation (DIP), and FCI for generous financial aid.

Received: February 10, 2010
Published online: August 16, 2010

[1] a) R. Noyori, *Angew. Chem.* **2002**, *114*, 2108 – 2123; *Angew. Chem. Int. Ed.* **2002**, *41*, 2008 – 2022; b) K. B. Sharpless, *Angew. Chem.* **2002**, *114*, 2126 – 2135; *Angew. Chem. Int. Ed.* **2002**, *41*, 2024 – 2032; c) P. J. Walsh, M. C. Kozlowski, *Fundamentals of Asymmetric Catalysis*, University Science Books, Sausalito, **2009**; d) H.-U. Blaser, E. Schmidt, *Asymmetric Catalysis on Industrial Scales*, Wiley-VCH, Weinheim, **2004**; e) A. Berkessel, H. Gröger, *Asymmetric Organocatalysis*, Wiley-VCH, Weinheim, **2004**; f) D. W. C. MacMillan, *Nature* **2008**, *455*, 304 – 308; g) B. List, *Angew. Chem.* **2010**, *122*, 1774 – 1779; *Angew. Chem. Int. Ed.* **2010**, *49*, 1730 – 1734.

[2] a) K. Faber, *Biotransformations in Organic Chemistry*, 5th ed., Springer, Berlin, **2004**; b) K. Drauz, H. Waldmann, *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook*, *Vol. I–III*, 2nd ed., Wiley-VCH, Weinheim, **2002**; c) A. Liese, K. Seelbach, C. Wandrey, *Industrial Biotransformations*, 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**.

[3] See, for example: a) E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, **1995**; b) K. C. Nicolaou, S. A. Snyder, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11929 – 11936; c) S. J. Keding, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11937 – 11942; d) E. A. Peterson, L. E. Overman, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11943 – 11948; e) B. M. Trost, *J. Org. Chem.* **2004**, *69*, 5813 – 5837; f) N. Z. Burns, P. S. Baran, R. W. Hoffmann, *Angew. Chem.* **2009**, *121*, 2896 – 2910; *Angew. Chem. Int. Ed.* **2009**, *48*, 2854 – 2867.

[4] Recent reviews of directed evolution:[58] a) T. W. Johannes, H. Zhao, *Curr. Opin. Microbiol.* **2006**, *9*, 261 – 267; b) S. Lutz, U. T. Bornscheuer, *Protein Engineering Handbook*, *Vol. 1–2*, 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. Biomol. Struct.* **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.

[5] Examples of rational and/or de novo design: a) F. Cedrone, A. Ménez, E. Quéméneur, *Curr. Opin. Struct. Biol.* **2000**, *10*, 405 – 410; b) T. Imanaka, H. Atomi in *Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook*, *Vol. I* (Eds.: K. Drauz, H. Waldmann), 2nd ed., Wiley-VCH, Weinheim, **2002**, pp. 67 – 95; c) D. Rotticci, J. C. Rotticci-Mulder, S. Denman, T. Norin, K. Hult, *ChemBioChem* **2001**, *2*, 766 – 770; d) T. Lanio, A. Jeltsch, A. Pingoud, *Protein Eng.* **2000**, *13*, 275 – 281; e) T. Ema, T. Fujii, M. Ozaki, T. Korenaga, T. Sakai, *Chem. Commun.* **2005**, 4650 – 4651; f) A. Glieder, R. Weis, W. Skranc, P. Poechlauer, I. Dreveny, S. Majer, M. Wubbolts, H. Schwab, K. Gruber, *Angew. Chem.* **2003**, *115*, 4963 – 4966; *Angew. Chem. Int. Ed.* **2003**, *42*, 4815 – 4818; g) S. Martí, J. Andrés, V. Moliner, E. Silla, I. Tuñón, *Chem. Soc. Rev.* **2008**, *37*, 2634 – 2643; h) J. A. Gerlt, P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2009**, *13*, 10 – 18; i) D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **2008**, *453*, 190 – 195; j) S. M. Lippow, B. Tidor, *Curr. Opin. Biotechnol.* **2007**, *18*, 305 – 311; k) A. Vardi-Kilshtain, M. Roca, A. Warshel, *Biotechnol. J.* **2009**, *4*, 495 – 500.

[6] a) T. Alber, J. A. Wozniak, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 747 – 750; b) M. Matsumura, S. Yasumura, S. Aiba, *Nature* **1986**, *323*, 356 – 358; c) B. C. Cunningham, J. A. Wells, *Protein Eng.* **1987**, *1*, 319 – 325; d) P. N. Bryan, M. L. Rollence, M. W. Pantoliano, J. Wood, B. C. Finzel, G. L. Gilliland, A. J. Howard, T. L. Poulos, *Proteins Struct. Funct. Genet.* **1986**, *1*, 326 – 334; e) M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 5685 – 5689; f) H. Liao, T. McKenzie, R. Hageman, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 576 – 580.

[7] K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5618 – 5622.

[8] M. T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem.* **1997**, *109*, 2961 – 2963; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 2830 – 2832.

[9] a) M. T. Reetz, *Pure Appl. Chem.* **1999**, *71*, 1503 – 1509; b) M. T. Reetz, *Pure Appl. Chem.* **2000**, *72*, 1615 – 1622; c) M. T. Reetz in *Evolutionary Methods in Biotechnology* (Eds.: S. Brakmann, A. Schwienhorst), Wiley-VCH, Weinheim, **2004**, pp. 113 – 141.

[10] a) M. T. Reetz in *Enzyme Assays—High-throughput Screening, Genetic Selection and Fingerprinting* (Ed.: J.-L. Reymond), Wiley-VCH, Weinheim, **2006**, pp. 41 – 76; b) J.-L. Reymond, *Enzyme Assays—High-throughput Screening, Genetic Selection and Fingerprinting*, Wiley-VCH, Weinheim, **2006**; c) H. Lin, V. W. Cornish, *Angew. Chem.* **2002**, *114*, 4580 – 4606; *Angew. Chem. Int. Ed.* **2002**, *41*, 4402 – 4425; d) Y. L. Boersma, M. J. Dröge, W. J. Quax, *FEBS J.* **2007**, *274*, 2181 – 2195; e) S. V. Taylor, P. Kast, D. Hilvert, *Angew. Chem.* **2001**, *113*, 3408 – 3436; *Angew. Chem. Int. Ed.* **2001**, *40*, 3310 – 3335; f) J.-L. Reymond, V. S. Fluxà, N. Maillard, *Chem. Commun.* **2009**, 34 – 46.

[11] a) D. W. Leung, E. Chen, D. V. Goeddel, *Technique* **1989**, *1*, 11 – 15; b) R. C. Cadwell, G. F. Joyce, *PCR Methods Appl.* **1994**, *3*, S136 – S140.

[12] a) P. C. Cirino, K. M. Mayer, D. Umeno, *Methods Mol. Biol.* **2003**, *230*, 3 – 9; b) T. S. Wong, D. Roccatano, M. Zacharias, U. Schwaneberg, *J. Mol. Biol.* **2006**, *355*, 858 – 871; c) 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; d) J. R. Cochran, Y.-S. Kim, S. M. Lippow, B. Rao, K. D. Wittrup, *Protein Eng. Des. Sel.* **2006**, *19*, 245 – 253; e) Y. Fujii, Y. Yamasaki, M. Matsumoto, H. Nishida, M. Hada, K. Ohkubo, *Biosci. Biotechnol. Biochem.* **2004**, *68*, 1722 – 1727; f) D. A. Drummond, B. L. Iverson, G. Georgiou, F. H. Arnold, *J. Mol. Biol.* **2005**, *350*, 806 – 816; g) A. Gratz, J. Jose, *Anal. Biochem.* **2008**, *378*, 171 – 176; h) T. Vanhercke, C. Ampe, L. Tirry, P. Denolf, *Anal. Biochem.* **2005**, *339*, 9 – 14.

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

[14] 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) R. Georgescu, G. Bandara, L. Sun in *Directed Evolution Library Creation*, *Vol. 231* (Eds.: F. H. Arnold, G. Georgiou), Humana Press, Totowa, **2003**, pp. 75 – 83; b) C. N. Dominy, D. W. Andrews in *Methods in Molecular Biology*, *Vol. 235* (Eds.: N. Casali, A. Preston), Humana Press, Totowa, **2003**,

pp. 209 – 223; c) M. A. Vandeyar, M. P. Weiner, C. J. Hutton, C. A. Batt, *Gene* **1988**, *65*, 129 – 133; d) R. D. Kirsch, E. Joly, *Nucleic Acids Res.* **1998**, *26*, 1848 – 1850; e) L. Zheng, U. Baumann, J.-L. Reymond, *Nucleic Acids Res.* **2004**, *32*, e115; f) W.-C. Tseng, J.-W. Lin, T.-Y. Wei, T.-Y. Fang, *Anal. Biochem.* **2008**, *375*, 376 – 378; g) A. Hidalgo, A. Schließmann, R. Molina, J. Hermoso, U. T. Bornscheuer, *Protein Eng. Des. Sel.* **2008**, *21*, 567 – 576.

[16] H. H. Hogrefe, J. Cline, G. L. Youngblood, R. M. Allen, *BioTechniques* **2002**, *33*, 1158 – 1165.

[17] J. Sanchis, L. Fernández, J. D. Carballeira, J. Drone, Y. Gumulya, H. Höbenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J.-P. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly, M. T. Reetz, *Appl. Microbiol. Biotechnol.* **2008**, *81*, 387 – 397.

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

[19] M. Z. Li, S. J. Elledge, *Nat. Methods* **2007**, *4*, 251 – 256.

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

[21] 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.

[22] a) M. T. Reetz, M. H. Becker, H.-W. Klein, D. Stöckigt, *Angew. Chem.* **1999**, *111*, 1872 – 1875; *Angew. Chem. Int. Ed.* **1999**, *38*, 1758 – 1761; b) W. Schrader, A. Eipper, D. J. Pugh, M. T. Reetz, *Can. J. Chem.* **2002**, *80*, 626 – 632.

[23] a) M. T. Reetz, K. M. Kühling, S. Wilensek, H. Husmann, U. W. Häusig, M. Hermes, *Catal. Today* **2001**, *67*, 389 – 396; b) M. T. Reetz, F. Daligault, B. Brunner, H. Hinrichs, A. Deege, *Angew. Chem.* **2004**, *116*, 4170 – 4173; *Angew. Chem. Int. Ed.* **2004**, *43*, 4078 – 4081.

[24] a) 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; b) K. M. Polizzi, C. U. Spencer, A. Dubey, I. Matsumura, J. H. Lee, M. J. Realff, A. S. Bommarius, *J. Biomol. Screening* **2005**, *10*, 856 – 864.

[25] K. Liebeton, A. Zonta, K. Schimossek, M. Nardini, D. Lang, B. W. Dijkstra, M. T. Reetz, K.-E. Jaeger, *Chem. Biol.* **2000**, *7*, 709 – 718.

[26] K. Miyazaki, F. H. Arnold, *J. Mol. Evol.* **1999**, *49*, 716 – 720.

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

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

[29] a) G. P. Horsman, A. M. F. Liu, E. Henke, U. T. Bornscheuer, R. J. Kazlauskas, *Chem. Eur. J.* **2003**, *9*, 1933 – 1939; b) S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45 – 54; c) R. J. Kazlauskas, *Chem. Listy* **2003**, *67*, 329 – 337.

[30] J. Paramesvaran, E. G. Hibbert, A. J. Russell, P. A. Dalby, *Protein Eng. Des. Sel.* **2009**, *22*, 401 – 411.

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

[32] a) M. Bocola, N. Otte, K.-E. Jaeger, M. T. Reetz, W. Thiel, *ChemBioChem* **2004**, *5*, 214 – 223; b) 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.

[33] D. Zha, S. Wilensek, M. Hermes, K.-E. Jaeger, M. T. Reetz, *Chem. Commun.* **2001**, 2664 – 2665.

[34] M. T. Reetz, B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier, M. M. Kayser, *Angew. Chem.* **2004**, *116*, 4167 – 4170; *Angew. Chem. Int. Ed.* **2004**, *43*, 4075 – 4078.

[35] a) C. T. Walsh, Y.-C. J. Chen, *Angew. Chem.* **1988**, *100*, 342 – 352; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 333 – 343; b) P. C. Brzostowicz, D. M. Walters, S. M. Thomas, V. Nagarajan, P. E. Rouvière, *Appl. Environ. Microbiol.* **2003**, *69*, 334 – 342.

[36] Reviews on BVMO-catalyzed asymmetric reactions: a) M. M. Kayser, *Tetrahedron* **2009**, *65*, 947 – 974; b) M. D. Mihovilovic, *Curr. Org. Chem.* **2006**, *10*, 1265 – 1287; c) R. Wohlgemuth, *Eng. Life Sci.* **2006**, *6*, 577 – 583.

[37] M. D. Mihovilovic, F. Rudroff, A. Winninger, T. Schneider, F. Schulz, M. T. Reetz, *Org. Lett.* **2006**, *8*, 1221 – 1224.

[38] a) M. W. Fraaije, J. Wu, D. P. H. M. Heuts, E. W. van Hellemond, J. H. L. Spelberg, D. B. Janssen, *Appl. Microbiol. Biotechnol.* **2005**, *66*, 393 – 400; b) F. Zambianchi, M. W. Fraaije, G. Carrea, G. de Gonzalo, C. Rodríguez, V. Gotor, G. Ottolina, *Adv. Synth. Catal.* **2007**, *349*, 1327 – 1331; c) C. Rodríguez, G. de Gonzalo, D. E. Torres Pazmiño, M. W. Fraaije, V. Gotor, *Tetrahedron: Asymmetry* **2008**, *19*, 197 – 203.

[39] E. Malito, A. Alfieri, M. W. Fraaije, A. Mattevi, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13157 – 13162.

[40] M. Bocola, F. Schulz, F. Leca, A. Vogel, M. W. Fraaije, M. T. Reetz, *Adv. Synth. Catal.* **2005**, *347*, 979 – 986.

[41] M. T. Reetz, S. Wu, *Chem. Commun.* **2008**, 5499 – 5501.

[42] M. T. Reetz, S. Wu, *J. Am. Chem. Soc.* **2009**, *131*, 15424 – 15432.

[43] a) E. Henke, U. T. Bornscheuer, *Biol. Chem.* **1999**, *380*, 1029 – 1033; b) M. Schmidt, D. Hasenpusch, M. Kähler, U. Kirchner, K. Wiggenhorn, W. Langel, U. T. Bornscheuer, *ChemBioChem* **2006**, *7*, 805 – 809.

[44] a) Y. Koga, K. Kato, H. Nakano, T. Yamane, *J. Mol. Biol.* **2003**, *331*, 585 – 592; b) W.-C. Suen, N. Zhang, L. Xiao, V. Madison, A. Zaks, *Protein Eng. Des. Sel.* **2004**, *17*, 133 – 140; c) Z. Qian, S. Lutz, *J. Am. Chem. Soc.* **2005**, *127*, 13466 – 13467.

[45] O. May, P. T. Nguyen, F. H. Arnold, *Nat. Biotechnol.* **2000**, *18*, 317 – 320.

[46] a) M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand, A. Cronin, Y. Genzel, A. Archelas, R. Furstoss, *Org. Lett.* **2004**, *6*, 177 – 180; b) B. van Loo, J. H. Lutje Spelberg, J. Kingma, T. Sonke, M. G. Wubbolts, D. B. Janssen, *Chem. Biol.* **2004**, *11*, 981 – 990; c) E. Y. Lee, M. L. Shuler, *Biotechnol. Bioeng.* **2007**, *98*, 318 – 327; d) L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *Appl. Environ. Microbiol.* **2005**, *71*, 3995 – 4003; e) M. Kotik, V. Štěpánek, P. Kyslik, H. Marešová, *J. Biotechnol.* **2007**, *132*, 8 – 15.

[47] G. DeSantis, K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz, M. J. Burk, *J. Am. Chem. Soc.* **2003**, *125*, 11476 – 11477.

[48] a) S. Fong, T. D. Machajewski, C. C. Mak, C.-H. Wong, *Chem. Biol.* **2000**, *7*, 873 – 883; b) M. Wada, C.-C. Hsu, D. Franke, M. Mitchell, A. Heine, I. Wilson, C.-H. Wong, *Bioorg. Med. Chem.* **2003**, *11*, 2091 – 2098; c) G. J. Williams, S. Domann, A. Nelson, A. Berry, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3143 – 3148; d) G. J. Williams, T. Woodhall, L. M. Farnsworth, A. Nelson, A. Berry, *J. Am. Chem. Soc.* **2006**, *128*, 16238 – 16247; e) S. Jennewein, M. Schürmann, M. Wolberg, I. Hilker, R. Luiten, M. Wubbolts, D. Mink, *Biotechnol. J.* **2006**, *1*, 537 – 548; f) N. Ran, J. W. Frost, *J. Am. Chem. Soc.* **2007**, *129*, 6130 – 6139; g) A. Bolt, A. Berry, A. Nelson, *Arch. Biochem. Biophys.* **2008**, *474*, 318 – 330.

[49] a) M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian, N. J. Turner, *Angew. Chem.* **2002**, *114*, 3309 – 3312; *Angew. Chem. Int. Ed.* **2002**, *41*, 3177 – 3180; b) R. Carr, M. Alexeeva, A. Enright, T. S. C. Eve, M. J. Dawson, N. J. Turner, *Angew. Chem.* **2003**, *115*, 4955 – 4958; *Angew. Chem. Int. Ed.* **2003**, *42*, 4807 – 4810; c) K. E. Atkin, R. Reiss, V. Koehler, K. R. Bailey, S. Hart, J. P. Turkenburg, N. J. Turner, A. M. Brzozowski, G. Grogan, *J. Mol. Biol.* **2008**, *384*, 1218 – 1231.

[50] A. Kirschner, U. T. Bornscheuer, *Appl. Microbiol. Biotechnol.* **2008**, *81*, 465 – 472.
[51] a) C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305 – 309; b) G. Matcham, M. Bhatia, W. Lang, C. Lewis, R. Nelson, A. Wang, W. Wu, *Chimia* **1999**, *53*, 584 – 589, and references therein.
[52] a) B. Lingen, D. Kolter-Jung, P. Dünkelmann, R. Feldmann, J. Grötzinger, M. Pohl, M. Müller, *ChemBioChem* **2003**, *4*, 721 – 726; b) B. J. Stevenson, J.-W. Liu, D. L. Ollis, *Biochemistry* **2008**, *47*, 3013 – 3025.
[53] a) M. Chen-Goodspeed, M. A. Sogorb, F. Wu, S.-B. Hong, F. M. Raushel, *Biochemistry* **2001**, *40*, 1325 – 1331; b) C. M. Hill, W.-S. Li, J. B. Thoden, H. M. Holden, F. M. Raushel, *J. Am. Chem. Soc.* **2003**, *125*, 8990 – 8991.
[54] a) M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **2003**, *125*, 13442 – 13450; b) M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, J.-E. Bäckvall, F. H. Arnold, *J. Am. Chem. Soc.* **2006**, *128*, 6058 – 6059; reviews of directed evolution of P450 enzymes: c) V. B. Urlacher, S. Lutz-Wahl, R. D. Schmid, *Appl. Microbiol. Biotechnol.* **2004**, *64*, 317 – 325; d) S. Kumar, J. R. Halpert, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 456 – 464; e) E. M. J. Gillam, *Chem. Res. Toxicol.* **2008**, *21*, 220 – 231; f) K. L. Tee, U. Schwaneberg, *Comb. Chem. High Throughput Screening* **2007**, *10*, 197 – 217.
[55] a) H. Asako, M. Shimizu, N. Itoh, *Appl. Microbiol. Biotechnol.* **2008**, *80*, 805 – 812; b) R. Machielsen, N. G. H. Leferink, A. Hendriks, S. J. J. Brouns, H.-G. Hennemann, T. Daußmann, J. van der Oost, *Extremophiles* **2008**, *12*, 587 – 594; c) J. Liang, J. Lalonde, B. Borup, V. Mitchell, E. Mundorff, N. Trinh, D. A. Kochrekar, R. N. Cherat, G. G. Pai, *Org. Process Res. Dev.* **2010**, *14*, 193 – 198; d) J. Liang, E. Mundorff, R. Voladri, S. Jenne, L. Gilson, A. Conway, A. Krebber, J. Wong, G. Huismann, S. Truesdale, J. Lalonde, *Org. Process Res. Dev.* **2010**, *14*, 188 – 192.
[56] a) M. Avi, R. M. Wiedner, H. Griengl, H. Schwab, *Chem. Eur. J.* **2008**, *14*, 11415 – 11422; see also b) R. Gaisberger, R. Weis, R. Luiten, W. Skranc, M. Wubbolts, H. Griengl, A. Glieder, *J. Biotechnol.* **2007**, *129*, 30 – 38.
[57] E. Antipov, A. E. Cho, K. D. Wittrup, A. M. Klibanov, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17694 – 17699.
[58] Reviews of directed evolution of enantioselective enzymes:[4g] a) M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5716 – 5722; b) M. T. Reetz, in *Advances in Catalysis*, *Vol. 49* (Eds.: B. C. Gates, H. Knözinger), Elsevier, San Diego, **2006**, pp. 1 – 69; c) M. T. Reetz, *J. Org. Chem.* **2009**, *74*, 5767 – 5778.
[59] Examples of comparative directed evolution studies:[4] a) M. R. Parikh, I. Matsumura, *J. Mol. Biol.* **2005**, *352*, 621 – 628; b) J. M. Joern, P. Meinhold, F. H. Arnold, *J. Mol. Biol.* **2002**, *316*, 643 – 656; c) T.-W. Wang, H. Zhu, X.-Y. Ma, T. Zhang, Y.-S. Ma, D.-Z. Wei, *Mol. Biotechnol.* **2006**, *34*, 55 – 68; d) M. Zaccolo, D. M. Williams, D. M. Brown, E. Gherardi, *J. Mol. Biol.* **1996**, *255*, 589 – 603; e) S.-Y. Tang, H. Fazelinia, P. C. Cirino, *J. Am. Chem. Soc.* **2008**, *130*, 5267 – 5271; review emphasizing quality not quantity when creating mutant libraries: f) S. Lutz, W. M. Patrick, *Curr. Opin. Biotechnol.* **2004**, *15*, 291 – 297.
[60] a) M. T. Reetz, D. Kahakeaw, J. Sanchis, *Mol. BioSyst.* **2009**, *5*, 115 – 122; b) M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem* **2008**, *9*, 1797 – 1804.
[61] G. Saab-Rincon, Y. Li, M. Meyer, M. Carbone, M. Landwehr, F. H. Arnold in *Protein Engineering Handbook* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, Weinheim, **2009**, pp. 481 – 492.
[62] 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.
[63] a) M. C. Saraf, A. R. Horswill, S. J. Benkovic, C. D. Maranas, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4142 – 4147; b) A. E. Firth, W. M. Patrick, *Nucleic Acids Res.* **2008**, *36*, W281 – W285; c) J. Damborsky, J. Brezovsky, *Curr. Opin. Chem. Biol.* **2009**, *13*, 26 – 34; d) J. F. Chaparro-Riggers, K. M. Polizzi, A. S. Bommarius, *Biotechnol. J.* **2007**, *2*, 180 – 191; e) T. S. Wong, D. Roccatano, U. Schwaneberg, *Environ. Microbiol.* **2007**, *9*, 2645 – 2659; f) S. A. Funke, N. Otte, T. Eggert, M. Bocola, K.-E. Jaeger, W. Thiel, *Protein Eng. Des. Sel.* **2005**, *18*, 509 – 514; g) G. L. Moore, C. D. Maranas, *AIChE J.* **2004**, *50*, 262 – 272; h) R. J. Hayes, J. Bentzien, M. L. Ary, M. Y. Hwang, J. M. Jacinto, J. Vielmetter, A. Kundu, B. I. Dahiyat, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15926 – 15931.
[64] 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.
[65] Early examples of saturation mutagenesis:[10e,26,28] a) D. K. Dube, L. A. Loeb, *Biochemistry* **1989**, *28*, 5703 – 5707; b) S. Climie, L. Ruiz-Perez, D. Gonzalez-Pacanowska, P. Prapunwattana, S.-W. Cho, R. Stroud, D. V. Santi, *J. Biol. Chem.* **1990**, *265*, 18776 – 18779; c) A. V. Teplyakov, J. M. van der Laan, A. A. Lammers, H. Kelders, K. H. Kalk, O. Misset, L. J. S. M. Mulleners, B. W. Dijkstra, *Protein Eng.* **1992**, *5*, 413 – 420; d) L. D. Graham, K. D. Haggett, P. A. Jennings, D. S. Le Brocque, R. G. Whittaker, P. A. Schober, *Biochemistry* **1993**, *32*, 6250 – 6258; e) M. S. Warren, S. J. Benkovic, *Protein Eng.* **1997**, *10*, 63 – 68; f) N. M. Antikainen, P. J. Hergenrother, M. M. Harris, W. Corbett, S. F. Martin, *Biochemistry* **2003**, *42*, 1603 – 1610; g) E. M. Gabor, D. B. Janssen, *Protein Eng. Des. Sel.* **2004**, *17*, 571 – 579; h) L. Rui, L. Cao, W. Chen, K. F. Reardon, T. K. Wood, *J. Biol. Chem.* **2004**, *279*, 46810 – 46817; i) A. R. Schmitzer, F. Lépine, J. N. Pelletier, *Protein Eng. Des. Sel.* **2004**, *17*, 809 – 819; j) T. S. Wong, K. L. Tee, B. Hauer, U. Schwaneberg, *Nucleic Acids Res.* **2004**, *32*, 26e.
[66] C. M. Clouthier, M. M. Kayser, M. T. Reetz, *J. Org. Chem.* **2006**, *71*, 8431 – 8437.
[67] J. D. Carballeira, P. Krumlinde, M. Bocola, A. Vogel, M. T. Reetz, J.-E. Bäckvall, *Chem. Commun.* **2007**, 1913 – 1915.
[68] S. Bartsch, R. Kourist, U. T. Bornscheuer, *Angew. Chem.* **2008**, *120*, 1531 – 1534; *Angew. Chem. Int. Ed.* **2008**, *47*, 1508 – 1511.
[69] Recent examples of saturation mutagenesis for a variety of purposes: a) N. U. Nair, H. Zhao, *ChemBioChem* **2008**, *9*, 1213 – 1215; b) M. Di Lorenzo, A. Hidalgo, R. Molina, J. A. Hermoso, D. Pirozzi, U. T. Bornscheuer, *Appl. Environ. Microbiol.* **2007**, *73*, 7291 – 7299; c) M. Zumárraga, C. Vaz Domínguez, S. Camarero, S. Shleev, J. Polaina, A. Martínez-Arias, M. Ferrer, A. L. De Lacey, V. M. Fernández, A. Ballesteros, F. J. Plou, M. Alcalde, *Comb. Chem. High Throughput Screening* **2008**, *11*, 807 – 816; d) H.-M. Li, L.-H. Mei, V. B. Urlacher, R. D. Schmid, *Appl. Biochem. Biotechnol.* **2008**, *144*, 27 – 36; e) X. Qi, Y. Chen, K. Jiang, W. Zuo, Z. Luo, Y. Wei, L. Du, H. Wei, R. Huang, Q. Du, *J. Biotechnol.* **2009**, *144*, 43 – 50; f) F. Shozui, K. Matsumoto, T. Sasaki, S. Taguchi, *Appl. Microbiol. Biotechnol.* **2009**, *84*, 1117 – 1124; g) K. Bajaj, P. C. Dewan, P. Chakrabarti, D. Goswami, B. Barua, C. Baliga, R. Varadarajan, *Biochemistry* **2008**, *47*, 12964 – 12973; h) S.-W. Oh, A.-U. Jang, C.-K. Jeong, H.-J. Kang, J.-M. Park, T.-J. Kim, *J. Microbiol. Biotechnol.* **2008**, *18*, 1401 – 1407; i) W. Wu, D. Zhu, L. Hua, *J. Mol. Catal. B* **2009**, *61*, 157 – 161; j) S. Wu, A. J. Fogiel, K. L. Petrillo, R. E. Jackson, K. N. Parker, R. DiCosimo, A. Ben-Bassat, D. P. O'Keefe, M. S. Payne, *Biotechnol. Bioeng.* **2008**, *99*, 717 – 720; k) G. J. Williams, R. D. Goff, C. Zhang, J. S. Thorson, *Chem. Biol.* **2008**, *15*, 393 – 401; l) A. Andreadeli, D. Platis, V. Tishkov, V. Popov, N. E. Labrou, *FEBS J.* **2008**, *275*, 3859 – 3869; m) A. Yep, G. L. Kenyon, M. J. McLeish, *Proc.*

Natl. Acad. Sci. USA **2008**, *105*, 5733 – 5738; n) D. Saboulard, V. Dugas, M. Jaber, J. Broutin, E. Souteyrand, J. Sylvestre, M. Delcourt, *BioTechniques* **2005**, *39*, 363 – 368.
[70] M. T. Reetz, J. D. Carballeira, J. J.-P. Peyralans, H. Höbenreich, A. Maichele, A. Vogel, *Chem. Eur. J.* **2006**, *12*, 6031 – 6038.
[71] M. T. Reetz in *Protein Engineering Handbook*, *Vol. 2* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, Weinheim, **2009**, pp. 409 – 439.
[72] M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260 – 2267.
[73] S. Wu, J. P. Acevedo, M. T. Reetz, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2775 – 2780.
[74] E. N. Jacobsen, *Acc. Chem. Res.* **2000**, *33*, 421 – 431.
[75] J. Y. Zou, B. M. Hallberg, T. Bergfors, F. Oesch, M. Arand, S. L. Mowbray, T. A. Jones, *Structure* **2000**, *8*, 111 – 122.
[76] 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.
[77] a) Y. Gumulya, projected Dissertation, Ruhr-Universität-Bochum, **2010**; b) S. Kille, Dissertation, Ruhr-Universität-Bochum, **2010**.
[78] a) D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, *Science* **2006**, *312*, 111 – 114; b) E. R. Lozovsky, T. Chookajorn, K. M. Brown, M. Imwong, P. J. Shaw, S. Kamchonwongpaisan, D. E. Neafsey, D. M. Weinreich, D. L. Hartl, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12025 – 12030.
[79] a) D. C. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, *Cell* **1984**, *38*, 835 – 840; b) J. A. Wells, *Biochemistry* **1990**, *29*, 8509 – 8517; c) A. Horovitz, *Curr. Biol.* **1996**, *6*, R121 – R126; d) A. S. Mildvan, *Biochemistry* **2004**, *43*, 14517 – 14520; e) C. M. Yuen, D. R. Liu, *Nat. Methods* **2007**, *4*, 995 – 997.
[80] a) L. Rui, Y. M. Kwon, A. Fishman, K. F. Reardon, T. K. Wood, *Appl. Environ. Microbiol.* **2004**, *70*, 3246 – 3252; b) W. M. Patrick, A. E. Firth, *Biomol. Eng.* **2005**, *22*, 105 – 112; c) A. D. Bosley, M. Ostermeier, *Biomol. Eng.* **2005**, *22*, 57 – 61; d) M. Denault, J. N. Pelletier in *Protein Engineering Protocols*, *Vol. 352* (Eds.: K. M. Arndt, K. M. Müller), Humana Press, Totowa, **2007**, pp. 127 – 154.
[81] a) K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem.* **2005**, *280*, 37742 – 37746; b) T. Li, K. Fan, J. Wang, W. Wang, *Protein Eng.* **2003**, *16*, 323 – 330; c) S. Akanuma, T. Kigawa, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 13549 – 13553; d) S. S. Sidhu, A. A. Kossiakoff, *Curr. Opin. Chem. Biol.* **2007**, *11*, 347 – 354.
[82] a) R. Stürmer, B. Hauer, M. B. Hall, K. Faber, *Curr. Opin. Chem. Biol.* **2007**, *11*, 203 – 213; b) J. F. Chaparro-Riggers, T. A. Rogers, E. Vazquez-Figueroa, K. M. Polizzi, A. S. Bommarius, *Adv. Synth. Catal.* **2007**, *349*, 1521 – 1531; c) A. Müller, R. Stürmer, B. Hauer, B. Rosche, *Angew. Chem.* **2007**, *119*, 3380 – 3382; *Angew. Chem. Int. Ed.* **2007**, *46*, 3316 – 3318; d) A. Fryszkowska, K. Fisher, J. M. Gardiner, G. M. Stephens, *J. Org. Chem.* **2008**, *73*, 4295 – 4298; e) S. K. Padhi, D. J. Bougioukou, J. D. Stewart, *J. Am. Chem. Soc.* **2009**, *131*, 3271 – 3280.
[83] a) S.-M. Lu, C. Bolm, *Chem. Eur. J.* **2008**, *14*, 7513 – 7516; b) T. Ohshima, H. Tadaoka, K. Hori, N. Sayo, K. Mashima, *Chem. Eur. J.* **2008**, *14*, 2060 – 2066; c) S.-M. Lu, C. Bolm, *Angew. Chem.* **2008**, *120*, 9052 – 9055; *Angew. Chem. Int. Ed.* **2008**, *47*, 8920 – 8923; d) C. J. Scheuermann née Taylor, C. Jaekel, *Adv. Synth. Catal.* **2008**, *350*, 2708 – 2714; e) J. G. de Vries, C. J. Elsevier, *Handbook of Homogeneous Hydrogenation*, *Vol. I–III*, Wiley-VCH, Weinheim, **2006**; f) B. H. Lipshutz, *Chem. Rev.* **2008**, *108*, 2916—2927.
[84] a) J. B. Tuttle, S. G. Ouellet, D. W. C. MacMillan, *J. Am. Chem. Soc.* **2006**, *128*, 12662 – 12663; b) N. J. A. Martin, B. List, *J. Am. Chem. Soc.* **2006**, *128*, 13368 – 13369.
[85] M. R. Martzen, S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, E. M. Phizicky, *Science* **1999**, *286*, 1153 – 1155.
[86] S. G. Peisajovich, D. S. Tawfik, *Nat. Methods* **2007**, *4*, 991 – 994.
[87] J. D. Bloom, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9995 – 10000.
[88] M. A. DePristo, *HFSP J.* **2007**, *1*, 94 – 98.
[89] M. Eigen, J. McCaskill, P. Schuster, *J. Phys. Chem.* **1988**, *92*, 6881 – 6891.
[90] S. Kurtovic, B. Mannervik, *Biochemistry* **2009**, *48*, 9330 – 9339.
[91] M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.* **2010**, *132*, 9144 – 9152.
[92] a) A. G. Sandström, K. Engström, J. Nyhlén, A. Kasrayan, J.-E. Bäckvall, *Protein Eng. Des. Sel.* **2009**, *22*, 413 – 420; b) K. Engström, J. Nyhlen, A. G. Sandström, J.-E. Bäckvall, *J. Am. Chem. Soc.* **2010**, *132*, 7038 – 7042.
[93] D. J. Ericsson, A. Kasrayan, P. Johansson, T. Bergfors, A. G. Sandström, J.-E. Bäckvall, S. L. Mowbray, *J. Mol. Biol.* **2008**, *376*, 109 – 119.
[94] a) J. Blanchet, J. Baudoux, M. Amere, M.-C. Lasne, J. Rouden, *Eur. J. Org. Chem.* **2008**, 5493 – 5506; b) M. Amere, M.-C. Lasne, J. Rouden, *Org. Lett.* **2007**, *9*, 2621 – 2624; c) J. T. Mohr, T. Nishimata, D. C. Behenna, B. M. Stoltz, *J. Am. Chem. Soc.* **2006**, *128*, 11348 – 11349.
[95] a) K. Miyamoto, H. Ohta, *Eur. J. Biochem.* **1992**, *210*, 475 – 481; b) Y. Terao, Y. Ijima, K. Miyamoto, H. Ohta, *J. Mol. Catal. B* **2007**, *45*, 15 – 20.
[96] 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.
[97] R. Hawwa, S. D. Larsen, K. Ratia, A. D. Mesecar, *J. Mol. Biol.* **2009**, *393*, 36 – 57.
[98] a) U. T. Bornscheuer, R. J. Kazlauskas, *Angew. Chem.* **2004**, *116*, 6156 – 6165; *Angew. Chem. Int. Ed.* **2004**, *43*, 6032 – 6040; b) K. Hult, P. Berglund, *Trends Biotechnol.* **2007**, *25*, 231 – 238; c) O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498 – 508; d) I. Nobeli, A. D. Favia, J. M. Thornton, *Nat. Biotechnol.* **2009**, *27*, 157 – 167.
[99] L. Liang, J. Zhang, Z. Lin, *Microb. Cell Fact.* **2007**, *6*, 36.
[100] a) P. Kötter, M. Ciriacy, *Appl. Microbiol. Biotechnol.* **1993**, *38*, 776 – 783; b) M. Jeppsson, O. Bengtsson, K. Franke, H. Lee, B. Hahn-Hägerdal, M. F. Gorwa-Grauslund, *Biotechnol. Bioeng.* **2006**, *93*, 665 – 673.
[101] B. Petschacher, S. Leitgeb, K. L. Kavanagh, D. K. Wilson, B. Nidetzky, *Biochem. J.* **2005**, *385*, 75 – 83.
[102] B. Henrissat, G. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637 – 644.
[103] M. Nishimoto, M. Kitaoka, *Biosci. Biotechnol. Biochem.* **2007**, *71*, 2101 – 2104.
[104] 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.
[105] a) R. A. Dixon, M. S. S. Reddy, *Phytochem. Rev.* **2003**, *2*, 289 – 306; b) J.-K. Weng, X. Li, N. D. Bonawitz, C. Chapple, *Curr. Opin. Biotechnol.* **2008**, *19*, 166 – 172.
[106] M.-W. Bhuiya, C.-J. Liu, *J. Biol. Chem.* **2010**, *285*, 277 – 285.
[107] R. M. Kelly, H. Leemhuis, L. Dijkhuizen, *Biochemistry* **2007**, *46*, 11216 – 11222.
[108] M. Ivancic, G. Valinger, K. Gruber, H. Schwab, *J. Biotechnol.* **2007**, *129*, 109 – 122.
[109] a) M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, *Chimia* **2002**, *56*, 721 – 723; b) M. T. Reetz, M. Rentzsch, A. Pletsch, A. Taglieber, F. Hollmann, R. J. G. Mondière, N. Dickmann, B. Höcker, S. Cerrone, M. C. Haeger, R. Sterner, *ChemBioChem* **2008**, *9*, 552 – 564; c) review of directed evolution of hybrid catalysts: M. T. Reetz, in *Topics in Organometallic Chemistry*, *Vol. 25* (Ed.: T. R. Ward), Springer, Heidelberg, **2009**, pp. 63 – 92.
[110] Reviews of synthetic enzymes comprising host proteins with anchored ligand/metal entities:[109c,111b] a) D. Qi, C.-M. Tann, D.

Haring, M. D. Distefano, *Chem. Rev.* **2001**, *101*, 3081 – 3111; b) Y. Lu, *Curr. Opin. Chem. Biol.* **2005**, *9*, 118 – 126.

[111] a) M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306 – 307; b) J. Steinreiber, T. R. Ward, in *Topics in Organometallic Chemistry*, *Vol. 25* (Ed.: T. R. Ward), Springer, Heidelberg, **2009**, pp. 93 – 112.

[112] M. T. Reetz, J. J.-P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun.* **2006**, 4318 – 4320.

[113] a) M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, P. Maiwald, J. J.-P. Peyralans, A. Maichele, Y. Fu, N. Jiao, F. Hollmann, R. Mondière, A. Taglieber, *Tetrahedron* **2007**, *63*, 6404 – 6414; b) see Ref. [109b].

[114] Select examples of remote mutations influencing enzyme activity:[4,25,28,33,58] a) S. Que, A. Okamoto, T. Yano, H. Kagamiyama, *J. Biol. Chem.* **1999**, *274*, 2344 – 2349; b) I. Axarli, A. Prigipaki, N. E. Labrou, *Biomol. Eng.* **2005**, *22*, 81 – 88; c) P. E. Tomatis, R. M. Rasia, L. Segovia, A. J. Vila, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13761 – 13766; d) L.-F. Wang, N. M. Goodey, S. J. Benkovic, A. Kohen, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15753 – 15758; e) see Ref. [49c].

[115] a) C.-J. Tsai, A. del Sol, R. Nussinov, *Mol. BioSyst.* **2009**, *5*, 207 – 216; b) Q. Cui, M. Karplus, *Protein Sci.* **2008**, *17*, 1295 – 1307; c) N. M. Goodey, S. J. Benkovic, *Nat. Chem. Biol.* **2008**, *4*, 474 – 482.

[116] a) T. Ichiye, M. Karplus, *Proteins Struct. Funct. Bioinf.* **2004**, *11*, 205 – 217; b) J. Liu, R. Nussinov, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 901 – 906; c) S. Hammes-Schiffer, S. J. Benkovic, *Annu. Rev. Biochem.* **2006**, *75*, 519 – 541.

[117] I. A. Mirza, B. J. Yachnin, S. Wang, S. Grosse, H. Bergeron, A. Imura, H. Iwaki, Y. Hasegawa, P. C. K. Lau, A. M. Berghuis, *J. Am. Chem. Soc.* **2009**, *131*, 8848 – 8854.

[118] T. Ichiye, M. Karplus, *Proteins Struct. Funct. Genet.* **1991**, *11*, 205 – 217.

[119] 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.

[120] M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem.* **2006**, *118*, 7909 – 7915; *Angew. Chem. Int. Ed.* **2006**, *45*, 7745 – 7751.

[121] M. T. Reetz, P. Soni, J. P. Acevedo, J. Sanchis, *Angew. Chem.* **2009**, *121*, 8418 – 8422; *Angew. Chem. Int. Ed.* **2009**, *48*, 8268 – 8272.

[122] M. T. Reetz, P. Soni, L. Fernández, *Biotechnol. Bioeng.* **2009**, *102*, 1712 – 1717.

[123] a) C. Morisseau, A. Archelas, C. Guitton, D. Faucher, R. Furstoss, J. C. Baratti, *Eur. J. Biochem.* **1999**, *263*, 386 – 395; b) M. Arand, H. Hemmer, H. Dürk, J. Baratti, A. Archelas, R. Furstoss, F. Oesch, *Biochem. J.* **1999**, *344*, 273 – 280.

[124] a) For a review of the theory of near-attack conformers in enzyme-catalyzed reaction, see T. C. Bruice, *Acc. Chem. Res.* **2002**, *35*, 139 – 148; b) for an MD study of epoxide hydrolases, see B. Schiøtt, T. C. Bruice, *J. Am. Chem. Soc.* **2002**, *124*, 14558 – 14570.

[125] a) J. K. Montclare, D. A. Tirrell, *Angew. Chem.* **2006**, *118*, 4630 – 4633; *Angew. Chem. Int. Ed.* **2006**, *45*, 4518 – 4521; b) P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112 – 4115; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052 – 4055; c) R. K. McGinty, C. Chatterjee, T. W. Muir, *Methods Enzymol.* **2009**, *462*, 225 – 243; d) A. L. Stokes, S. J. Miyake-Stoner, J. C. Peeler, D. P. Nguyen, R. P. Hammer, R. A. Mehl, *Mol. BioSyst.* **2009**, *5*, 1032 – 1038; e) E. Kaya, K. Gutsmiedl, M. Vrabel, M. Müller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858 – 2861.

[126] M. D. Hughes, D. A. Nagel, A. F. Santos, A. J. Sutherland, A. V. Hine, *J. Mol. Biol.* **2003**, *331*, 973 – 979.

[127] a) J. Van den Brulle, M. Fischer, T. Langmann, G. Horn, T. Waldmann, S. Arnold, M. Fuhrmann, O. Schatz, T. O'Connell, D. O'Connell, A. Auckenthaler, H. Schwer, *BioTechniques* **2008**, *45*, 340 – 343; b) P. Iyidogan, S. Lutz, *Biochemistry* **2008**, *47*, 4711 – 4720.

[128] O. Trapp, *Angew. Chem.* **2007**, *119*, 5706 – 5710; *Angew. Chem. Int. Ed.* **2007**, *46*, 5609 – 5613.

[129] Reviews of synthetic biology focusing on metabolic engineering: a) J. Kirby, J. D. Keasling, *Annu. Rev. Plant Biol.* **2009**, *60*, 335 – 355; b) A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 277 – 285; c) S. C. Wenzel, R. Müller, *Curr. Opin. Biotechnol.* **2005**, *16*, 594 – 606; d) A. Kern, E. Tilley, I. S. Hunter, M. Legiša, A. Glieder, *J. Biotechnol.* **2006**, *129*, 6 – 29; e) S. Blanchard, J. S. Thorson, *Curr. Opin. Chem. Biol.* **2006**, *10*, 263 – 271; f) C. Hertweck, *Angew. Chem.* **2009**, *121*, 4782 – 4811; *Angew. Chem. Int. Ed.* **2009**, *48*, 4688 – 4716; g) A. Das, C. Khosla, *Acc. Chem. Res.* **2009**, *42*, 631 – 639; h) C. H. Martin, D. R. Nielsen, K. V. Solomon, K. L. Jones Prather, *Chem. Biol.* **2009**, *16*, 277 – 286; i) E. J. Steen, Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, J. D. Keasling, *Nature*, **2010**, *463*, 559 – 563.

[130] W. L. Tang, Z. Li, H. Zhao, *Chem. Commun.* **2010**, *46*, 5461 – 5463.

