Directed Evolution of Enzymes as Catalysts in Synthetic Organic Chemistry

Manfred T. Reetz

Updated August 2020

The following ongoing account posted at the website of the author is not intended to be a comprehensive description of our research in directed evolution of selective enzymes as catalysts in organic chemistry and biotechnology (and certainly not of the whole field). Rather, it focuses on the basic principles and methods that my group has developed during the past 25 years, while also highlighting the different types of lessons that we have been learned in this transdisciplinary research. The following catalytic parameters can be controlled by directed evolution, which are highlighted here:

Stereoselectivity, Regioselectivity, Substrate Scope and Activity

The concept: A fundamentally new source of catalysts for asymmetric catalysis

Enzymes have been used as catalysts under mild conditions in synthetic organic chemistry for more than a century,[1a] but biocatalysis has often suffered from the following traditional limitations:

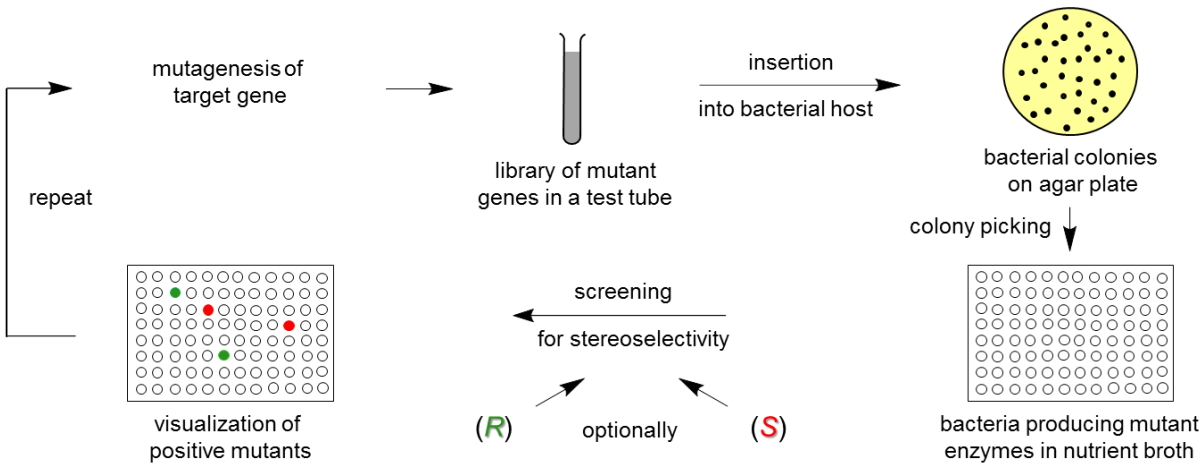
- Narrow substrate scope (activity)
- Insufficient or wrong stereoselectivity
- Insufficient or wrong regioselectivity
- Insufficient robustness under operating conditions

The development of protein engineering methods for enhancing enzyme robustness has hardly attracted the attention of organic chemists, because immobilization can often solve this problem. In contrast, *stereo- and regioselectivity as well as substrate scope are central issues if enzymes as ecologically and economically viable alternatives to man-made catalysts are to be used in synthetic organic chemistry*. During the past 25 years, the Reetz group pioneered the solution to these central problems by invoking directed evolution, a protein engineering method that mimics natural evolution. A Perspective Article[1a] covering this research area including biocatalysis in general appeared in 2013, followed in 2016 by a monograph focusing specifically on directed

evolution of selective enzymes.[1b] *The newest review with a wealth of information on methodology development appeared in 2020*.[1c] Our interest in this exciting endeavor goes back to the mid-1990s when my group proposed a *fundamentally new approach to asymmetric catalysis, namely directed evolution of stereoselective enzymes as catalysts in synthetic organic chemistry and biotechnology*.[1,2] Keeping the societal value of chiral pharmaceuticals, natural products, plant-protecting agents and fragrances in mind, we thought that *"evolution in the test-tube"* could be complementary to the development of chiral man-made synthetic catalysts for asymmetric transformations. If successful, this would provide a prolific and inexhaustible source of novel catalysts for a multitude of different asymmetric transformations under mild and environmentally friendly conditions.

The gene (DNA segment) which encodes a particular unselective wild-type (WT) enzyme is subjected to known molecular biological methods of random gene mutagenesis (e.g., error prone PCR, saturation mutagenesis, mutator strains, etc.), used previously by other groups, especially the Arnold-group*, to enhance the stability of proteins* (see Section below on directed evolution of protein stability). Subsequently the library of mutated genes are inserted into a bacterial host (expression system) such as *E. coli*, so that after a few simple steps bacterial colonies appear on agar plates. Following colony picking, the bacterial colonies are placed in the wells of 96- or 384-format microtiter plates where they grow in a nutrient broth, thereby producing mutant enzymes. In principle, the size of a given mutant library can vary between a few hundred to several million members or even more. Screening either for (*R*)- or (*S*)-selective mutants (or both!) follows, the gene of an improved mutant then being used as the template to initiate a subsequent cycle, and the process is continued until the desired degree of stereoselectivity has been achieved (Scheme 1).[1,2] Unlike the development of synthetic transition metal catalysts and organocatalysts, this concept relies on the evolutionary pressure exerted in each cycle. Therefore, in a sense it is perhaps the most rational way to develop stereoselective catalysts. Since protein sequence space is essentially endless, the problem boils down to finding a "hit" among millions and billions of mutants, which in turn raises the question of how to achieve high-throughput ee-screening, or ideally, how to reduce it as much as possible.[3] Indeed, screening is the bottleneck of directed

evolution. *Consequently, the generation of small but smart mutant libraries is essential, a conclusion that we made early on.*

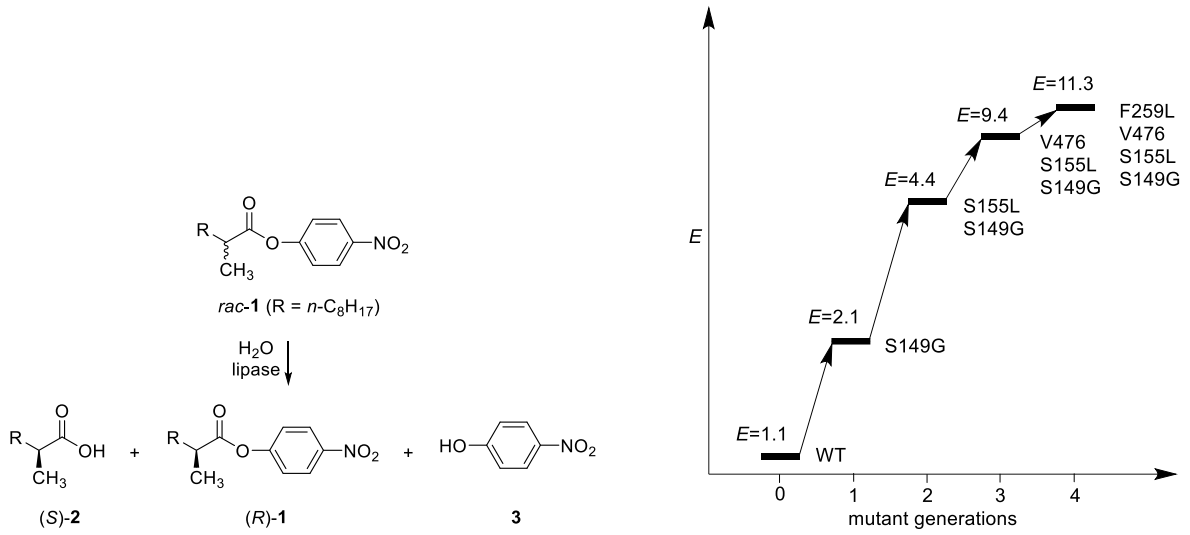


Scheme 1. General principle of laboratory evolution of stereoselective enzymes.[1,2] The green and red dots on the bottom left 96-well microtiter plate indicate improved (*R*)- and (*S*)-selective mutants, respectively, the genes of which are subsequently used as templates in the next cycle.

Since a given substrate may react very slowly or not at all, an on-plate pretest for activity can be included,[3] which means directed evolution of measurable substrate acceptance (rate) *and* stereoselectivity is ensured.

Proof-of-principle

In a proof-of-principle study using four cycles of epPCR, which we published in 1997, it was possible to increase the selectivity factor *E* (relative rates of the two enantiomers) of the hydrolytic kinetic resolution of *rac*-**1** in favor of (*S*)-**2**, catalyzed by the lipase from *Pseudomonas aeruginose* (PAL), from *E* =1.1 to *E*= 11.3 (Scheme 2).[4] In this early work, the enzyme PAL was obtained from the microbiologist K.-E. Jaeger at nearby Bochum University in a collaborative effort. *The excitement in the Reetz-group upon achieving this seminal result was unbounded!* In the overall task, the first high-throughput ee-assay had to be developed, in this case a UV/Vis-based system monitoring the time-dependent appearance of *p*-nitrophenolate (**3)** at 405 nm which is released from (*R*)- and (*S*)-**1** in parallel measurements on 96-well microtiter plates. Later we and others developed further ee-screening systems for different types of enzymes.[3]



Scheme 2. First case of directed evolution of a stereoselective enzyme, involving the lipase PAL as the enzyme and the hydrolytic kinetic resolution of *rac*-**1** as the model reaction.[4]

In subsequent efforts, the fifth epPCR based cycle resulted in only minor improvement (*E* = 13), which indicated that recursive epPCR is insufficient and that methodology development was necessary. It was also clear that efficacy in research demanded that mutagenesis, screening and organic chemical work should all be performed in our lab at the Max-Planck-Institut für Kohlenforschung in Mülheim. But we were actually not sure whether our concept (Scheme 1) was general. *Indeed, some protein engineers at international conferences were contending that the levelling-off effect that we had not kept secret was the major limitation of an otherwise novel approach!* Nevertheless, we took a bold step forward and decided not to give up, while investing at the Max-Planck-Institut für Kohlenforschung a great deal of money for establishing a gene mutagenesis system and a comprehensive robotic platform. We also hired molecular biologists, biochemists and synthetic organic chemists with the ability to work with enzymes. *It was an adventurous scientific endeavor and a considerable financial risk!*

In a follow-up study, saturation mutagenesis focused at a 4-residue site lining the binding pocket was performed, leading to *E* = 30, *which was the first case of saturation mutagenesis at an active site in the quest to manipulate stereoselectivity*.[5] Later we generalized this approach and gave it a name to distinguish it from saturation mutagenesis at remote sites for other purposes, namely ***C***ombinatorial ***A***ctive-Site ***S***aturation ***T***est (CAST) (see below). In the

same study, the combination of epPCR, saturation mutagenesis and DNA shuffling (Stemmer technique) with generation of a highly (*S*)-selective mutant (*E* = 51) provided even better results.[5] This mutant is characterized by six point mutations (D20N, S53P, S155M, L162G, T180I and T234S) which are mainly on the surface of the enzyme (Fig. 1), only mutation L162G being directly at the active site. This surprised us, and at the time we nevertheless hinted that close mutations are more important than those possibly exerting a remote effect. After all, the above mentioned single saturation mutagenesis experiment at the 4-residue site next to the binding pocket had led to notably enhanced enantioselectivity.[5] In all of these early efforts, a total of 50,000 transformants had to be screened, an enormously laborious effort.

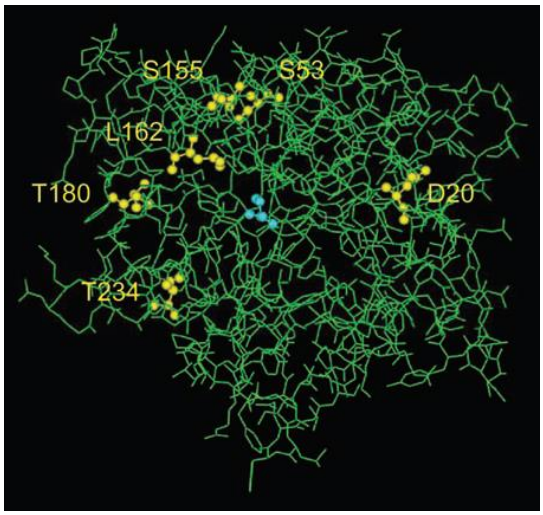
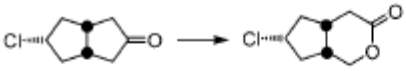
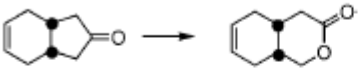
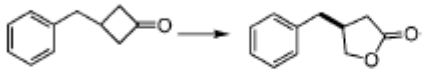
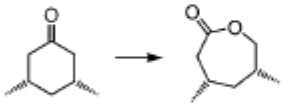
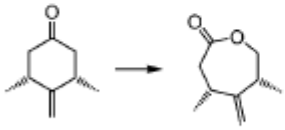
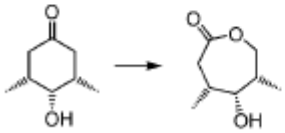
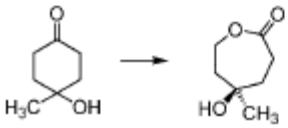


Fig. 1. X-ray structure of the WT lipase from *P. aeruginosa* featuring catalytically active serine at position 82 (blue) (B. W. Dijkstra, et al, J. Mol. Biol. 2000, 275, 31219-31225), in which the six point mutations (yellow) of the best mutant are marked in yellow, only L162G being close to the active site.[5,6]

We also succeeded in inverting enantioselectivity, i.e., an (*R*)-selective mutant lipase was evolved. Progress up to 2004 was summarized in a review, which includes inputs of other academic and industrial research groups who had joined efforts in generalizing the concept by using the same basic strategies but other enzyme types.[6] *One of the questions that arose early on was whether a mutant evolved for a given substrate is also effective as a catalyst for structurally related and/or different substrates.* Today we know that this is generally the case. An early example from our group concerns a mutant of the Baeyer-Villiger monooxygenase CHMO evolved for the desymmetrization of 4-hydroxycyclohexanone, which functions extremely well in the oxidative

desymmetrization of a number of structurally different prochiral ketones with yields of >85% and generally excellent stereoselectivity (Table 1).[7]

Table 1. Oxidative desymmetrization catalyzed by the CHMO mutant Phe432Ser using air as the sole oxidant in a whole-cell process.[7]

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

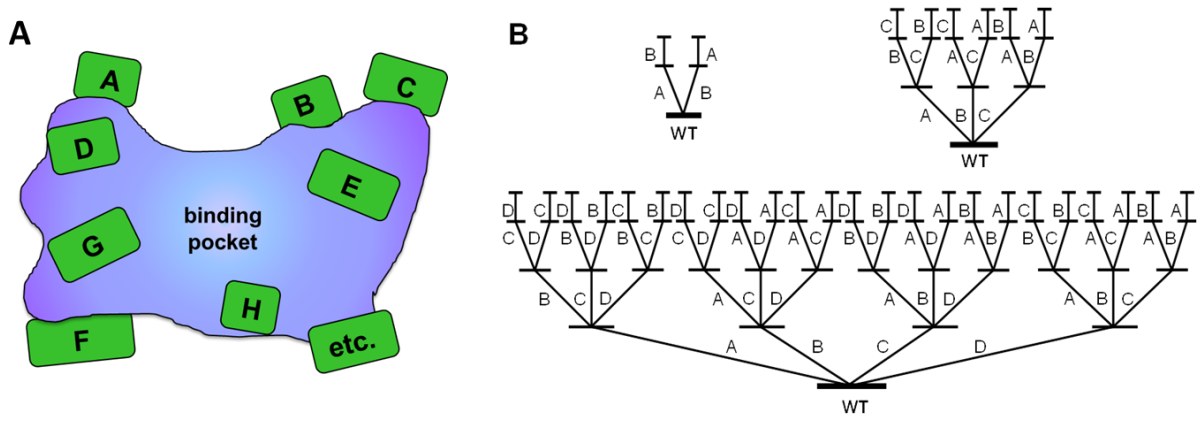
A very different and likewise intriguing aspect of this research concerns the study of structure/selectivity relationships. *We coined the expression "learning from directed evolution."* In collaboration with Walter Thiel, a mechanistic and QM/MM-based theoretical analysis of the best (*S*)-selective lipase (PAL) mutant for substrate *rac*-**1** uncovered the source of enhanced enantioselectivity as involving a "relay" effect, and also led to the prediction that only two of the six mutations are actually necessary.[8] In that study the predicted double mutant was generated and found to be even better (*E* = 63)! This initially "surprising" result suggested that the epPCR-based strategy has a tendency to introduce superfluous point mutations, which also explains the presence of most of the

remote amino acid substitutions.[8] *This was a triumph of theory, but it also clearly demonstrated that our strategy for probing protein sequence space was not as efficient as it could be, and that advanced methodology development in the area of directed evolutions is crucial for further progress*.

Methodology development: Away from blind directed evolution!

Since screening[3] is the bottleneck of directed evolution studies,[1,2] we decided in 2005 to intensify our activities in methodology development. Recalling our previous experiment involving saturation mutagenesis at a 4-residue site next to the binding pocket of the lipase PAL as catalyst in the hydrolytic kinetic resolution of *rac*-**1** which led to a selectivity factor of $E = 30$,[5] we re-considered this type of mutagenesis method, this time in a systematic manner. Sites labeled A, B, C, D, etc. around the binding pocket of an enzyme according to the *Combinatorial Active-Site Saturation Test (CAST)*[9] are first identified using X-ray data or homology models, supported by simple docking computations. Each site comprises one or more amino acid positions (residues) (Scheme 3a). The term CAST[9] is simply a convenient acronym for a process that we had already described in 2001 with a focus on stereoselectivity.[5] It distinguishes it from saturation mutagenesis at remote sites for protein stabilization that J. A. Wells had published back in 1985.

Following screening of the initial saturation mutagenesis libraries at sites A, B, C, D, etc., the best mutants serve as templates for randomization at the other sites, and the evolutionary procedure is continued until the desired degree of catalyst improvement has been achieved. The overall process was termed *Iterative Saturation Mutagenesis (ISM)* and was first illustrated using an epoxide hydrolase.[10,11] Scheme 3b features ISM systems for 2-, 3- and 4-site systems correlating with 2, 6 and 24 upward pathways (trajectories). This approach to directed evolution is knowledge-driven, requiring either a crystal structure or a homology model. *In view of Emil Fischer's lock-and-key hypothesis, CAST/ISM is the most logical strategy for reshaping the binding pockets of enzymes*.



Scheme 3. CAST/ISM approach to directed evolution in the quest to manipulate stereo- and regioselectivity as well as activity and substrate scope. A: CAST sites for saturation mutagenesis. B: Two-, three- and four-site ISM systems.[10,11]

In the original study,[10] statistical analysis regarding oversampling for a given degree of library coverage was not considered. For this purpose a user-friendly computer aid (CASTER) based on the Patrick/Firth algorithm was later developed, accessible free of charge at the Reetz-homepage www.kofo.mpg.de/en/research/biocatalysis. Since the number of transformants that have to be screened for, e.g., 95% library coverage, increases astronomically with increasing number of residues in a randomization site, *the use of reduced amino acid alphabets ranging from 12 down to 5 members in CAST/ISM was proposed and implemented experimentally*.[11,12] Since the amount of screening can be reduced drastically by this measure, it constitutes a significant advancement in laboratory evolution. Of course, structural diversity in the mutant libraries is also reduced, but it was shown that this is over-compensated by higher library quality.

The mathematics behind this approach is illustrated in Table 2 which features the number of transformants needed for 95% library coverage as a function of the amino acid alphabet, in this case NNK encoding all 20 canonical amino acids versus NDT encoding only 12 (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Ar, Ser, Gly). The respective numbers can be calculated by CASTER for any reduced amino acid alphabet that the researcher may want to use. We have successfully applied NDT codon degeneracy in the directed evolution of several enzyme types, enhancing and reversing stereoselectivity being the goals.[1,12] Moreover, our group has also used much smaller amino acid alphabets requiring even less

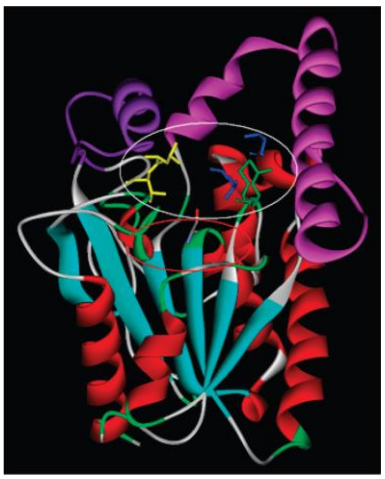
screening when applying saturation mutagenesis, e.g., as hypothesized by a bioinformatics guide in the case of a stereoselective Baeyer-Villiger monoxygenase (BVMO).[13] In all cases the optimal choice of the reduced amino acid alphabet should be guided by structural and mechanistic data. Previous data regarding successful amino acid substitutions induced by directed evolution techniques or by rational design can also be used to make optimal decisions concerning the nature of amino acids as building blocks. *Overall, our idea of utilizing reduced amino acid alphabets has proven to be extremely fruitful, and today it is a central pillar in protein engineering.*

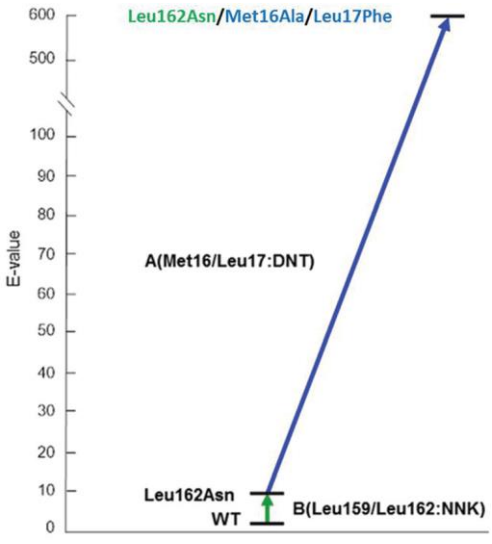
Table 2. Oversampling necessary for 95% coverage as a function of NNK and NDT codon degeneracy calculated on the basis of Patrick/Firth statistics (see CASTER computer aid free of charge on Reetz-homepage www.kofo.mpg.de/en/research/biocatalysis for any other codon degeneracy that may be of interest).

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	32 768	98 163	1728	5175
4	1 048 576	3 141 251	20 736	62 118
5	33 554 432	100 520 093	248 832	745 433
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	$>1.1\times10^{15}$	$>3.4\times10^{15}$	$>6.1\times10^{10}$	$>1.9\times10^{11}$

Following these advances, we and others applied ISM in the directed evolution of a number of other enzyme types, in all cases using small mutant libraries (generally only a few thousand transformants or less).[1,12] Because ISM appeared to be so efficient, it was logical and indeed necessary to re-visit our original model system of 1997 involving the hydrolytic kinetic resolution of *rac*-**1** catalyzed by the lipase PAL, this time employing the new method.[14] No other

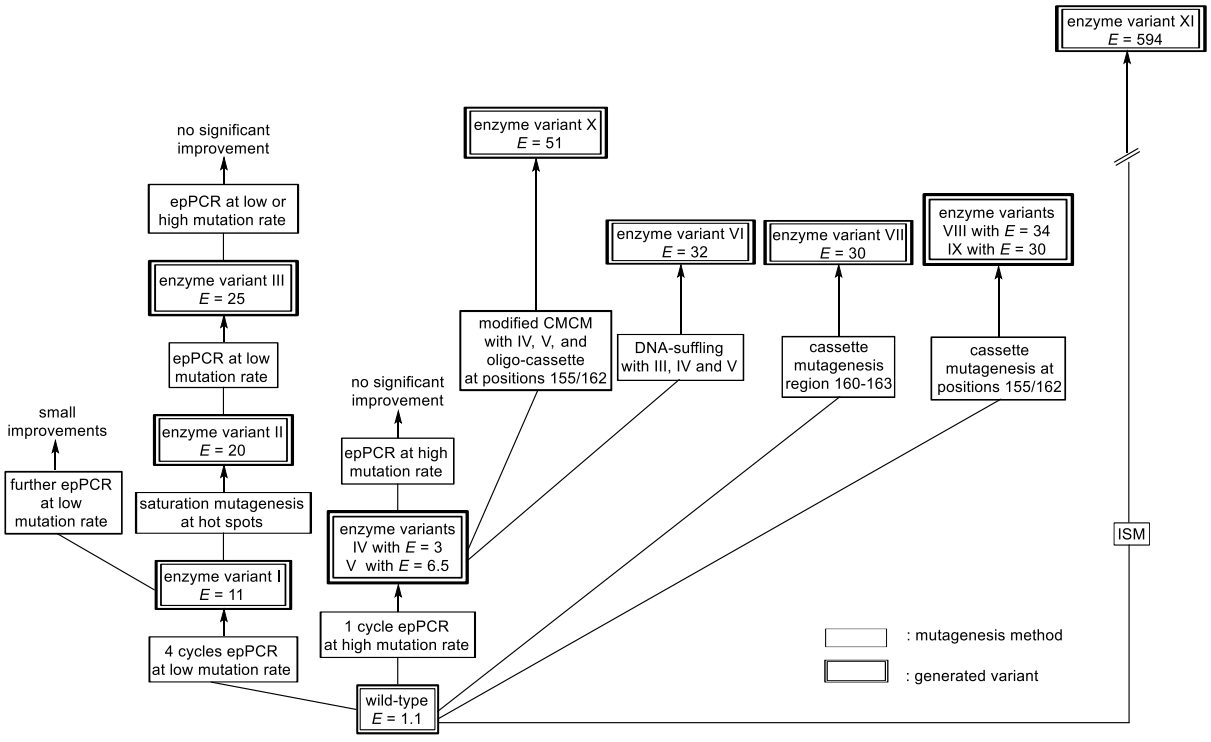
enzyme has been studied so systematically using essentially all major mutagenesis methods and strategies. *Systematic comparison of the different approaches in directed evolution is crucial in identifying the optimal strategy (but such assessments are rarely made!).* Guided by the X-ray structure of PAL, three 2-residue sites A, B and C were defined (Scheme 4, left), which means that 6 possible upward pathways are possible. Only two of them were explored, because the second one, WT → B → A, already provided a triple mutant Leu162Asn/Met16Al/Leu17Phe showing extremely high enantioselectivity with *E* = 594 (Scheme 4, right).[14a] Visiting site C was also not necessary. A total of less than 10,000 transformants had to be screened in exploring two pathways. It can be seen that ISM is by far superior to all previously practiced approaches. Later, second generation CAST/ISM was developed, which proved to be even more efficient requiring even less screening (see below). Furthermore, partial deconvolution experiments showed that none of the mutations are superfluous, and that a dramatic cooperative effect acting between point mutation Leu162Asn and the set of mutations Met16Al/Leu17Phe occurs. [14a] If one assumes traditional mutational additivity (A. Fersht, J. A. Wells), it would seem that the first mutation contributes very little, but this is not the case! The second set Met16Al/Leu17Phe alone has an *E*-value of only 2.6!!! Cooperative effects of this kind explain the efficacy of ISM. See also Section below entitled "Learning from Directed Evolution". This does not mean that mutational additivity never occurs in protein engineering, but non-additivity (synergistic or deleterious) are much more common than originally thought (see Section below on Learning from Directed Evolution).





Scheme 4. Left: Three sites in the lipase PAL considered on the basis of the PAL X-ray structure for iterative saturation mutagenesis (ISM), namely A (Met16/Leu17, green), B (Leu159/Leu162, blue), and C (Leu231/Val322, yellow); active site is Ser82 (stick representation in gray and red). Right: Experimental ISM results concerning PAL as the catalyst in the hydrolytic kinetic resolution of *rac*-**1** with preferential formation of (*S*)-**2** .[14a]

It is instructive to summarize all mutagenesis experiments that were performed with this enzyme system over the years (Scheme 5), which allows for important conclusions: The combination of epPCR and DNA shuffling is successful, as is the use of initial mutant libraries generated by saturation mutagenesis, but CAST/ISM is by far the most efficient strategy requiring the least degree of screening for reaching truly high stereoselectivity. Experience in our group with other enzyme types point in the same direction, although no other system in directed evolution has been studied as systematically as the lipase-catalyzed hydrolytic kinetic resolution of *rac*-**1**.



Scheme 5. Summary of all mutagenesis experiments using PAL as the enzyme and the hydrolytic kinetic resolution of *rac*-**1** as the model reaction with preferential formation of (*S*)-**2** .[14]

Our efforts in methodology development did not end with the above insights, since several questions remained to be addressed. These are ongoing efforts which continue to constitute the heart of our research. Recent refinements and

new developments led to notable advancements (second generation CAST/ISM). The interested reader is referred to the following key publications:

• *General use of reduced amino acid alphabets when applying CAST/ISM including tips on how to group individual residues into multi-residue sites*.[1,11-14]

• *Triple code saturation mutagenesis (TCSM) in a reduced amino acid alphabet utilizing three amino acids as combinatiorial building blocks*.[15]

• *Simultaneous directed evolution of thermostability, activity and stereoselectivity*.[15c]

• *New PCR-based saturation mutagenesis method for difficult-to-amplify templates*.[16]

• *Bioinformatics guide when applying ISM*.[13]

• *Development of Quick Quality Control (QQC) of mutant libraries*.[11,17]

• *Development of pooling strategy for lowering the amount of screening*.[17]

• *In silico guide for ISM based on the ASRA-algorithm*.[18]

•*Elimination of amino acid bias in saturation mutagenesis when using the 20 canonical amino acids as building blocks in focused libraries*.[19]

•*How to escape from local minima in evolutionary pathways*.[20]

•*Combinatorial solid phase gene synthesis for library creation*.[21]

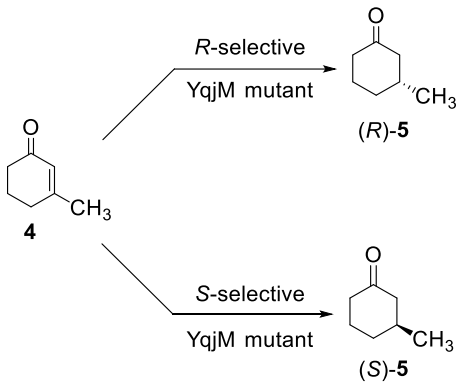
•*Techno-economical analysis of various saturation mutagenesis techniques*.[22]

An important issue that needed to be addressed concerns the question as to which upward pathway should be chosen in a given ISM system. In dozens of ISM-based studies, experience in our lab and in other groups has shown that arbitrarily chosen pathways lead to success. In one study featuring a 4-site ISM system we went to the trouble of exploring all 4! = 24 upward pathways.[20] *All of the 24 ISM pathways proved to be successful with creation of highly enantioselective mutants, which explains statistically why so many arbitrarily chosen pathways in various previous studies had resulted in satisfactorally enhanced (or reversed) stereoselectivity*. But some trajectories proved to be more successful than others, which is not surprising.[20]

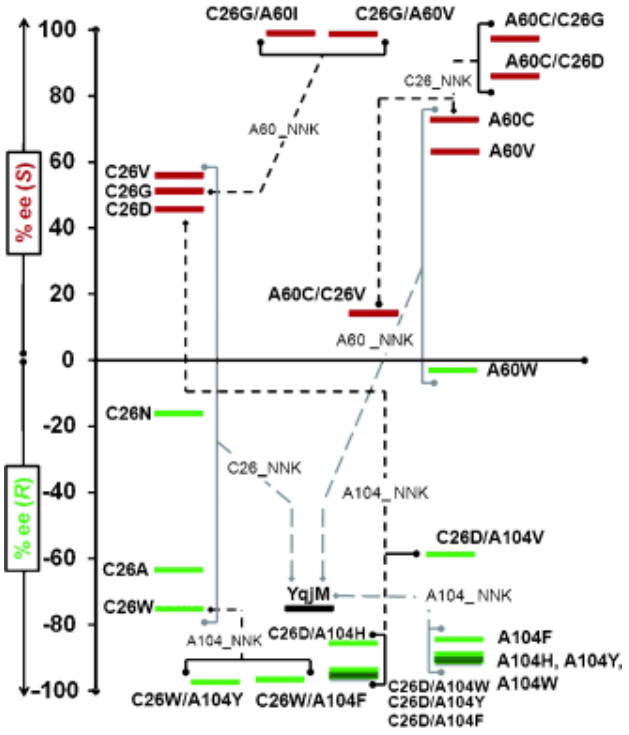
Along a different line of methodology development, a CAST/ISM approach to *simultaneously evolving three enzyme properties was recently reported, namely thermostability, enantioselectivity and activity* of an epoxide hydrolase.[15c] This is different from attempting to evolve such parameters sequentially in a traditional manner.

Selected examples of iterative saturation mutagenesis (ISM)

Judging by the increasing number of studies based on Iterative Saturation Mutagenesis (ISM) from our group and from other labs,[1,12] it appears that this method has emerged as the strategy of choice in directed evolution.[11] The important contributions by other groups using our ISM concept have been summarized in recent reviews,[1,12] and new examples continue to be reported. Only a few selected studies from our group are highlighted here without going into details. These include hydrolases such as the lipase CALB[23] and limonene epoxide hydrolase.[15a,24] Reductases also continue to be the subject of our research.[15b] The Old Yellow enzyme YqjM from *Bacillus subtilis* is a well-known enoate reductase which allows the conjugate reduction of enones and other activated olefins, but for many substrates such as 3-substituted 2-cyclohexenones enantioselectivity is poor to moderate. For example, in the reduction of 3-methyl-2-cyclohexenone (**4**), WT YqjM leads to only 76% ee in moderate favor of (*R*)-**5** (Scheme 6). Application of ISM provided both (*R*)- and (*S*)-selective mutants (>95% ee) (Scheme 7).[17] Other 3-alkyl substituted cyclohexenones also reacted with high enantioselectivity.

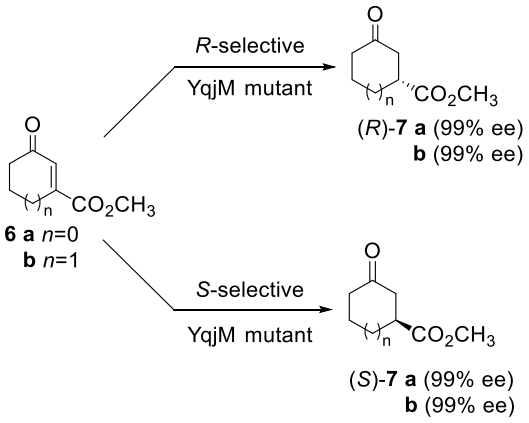


Scheme 6. Enoate reductase YqjM-catalyzed reduction of **4**.[17]



Scheme 7. Result of ISM-exploration in the YqjM-catalyzed reduction **4** → **5**.[17]

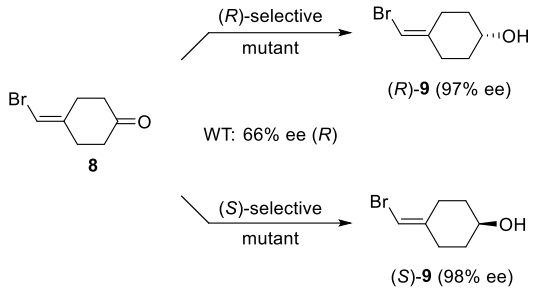
The best variants evolved for **4** also proved to be excellent catalysts in the conjugate reduction of keto-esters **6a-b** (Scheme 8).[17]



Scheme 8. Enoate reductase YqjM-catalyzed reduction of keto-esters **6a-b** → **7a-b**.[17]

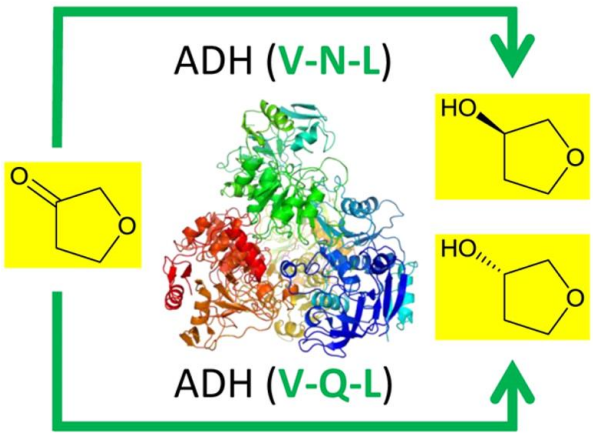
Another type of stereoselective reductive process, ketone reduction, can often be catalyzed by Noyori-Ru-catalysts with excellent enantioselectivity. An alternative is biocatalysis using alcohol dehydrogenases (ADHs). It is of particular synthetic value to focus on those cases in which the generally successful synthetic Ru-based chiral catalysts fail. One such example is the ADH-catalyzed reduction of ketones of the type **8** with formation of axially chiral alcohols **9**, in which the thermostable ADH from *Thermoethanolicus brockii* (TbSADH) served as the enzyme (Scheme 9).[25] The (*R*)- or (*S*)-products are key compounds in the

synthesis of a variety of other derivatives accessible by Pd-catalyzed carbonylation or Suzuki-coupling, *an example of the power of combining biocatalysis with transition metal catalysis*.



Scheme 9. ADH-catalyzed reduction of ketone **8** leading to axially chiral alcohols **9**.[25]

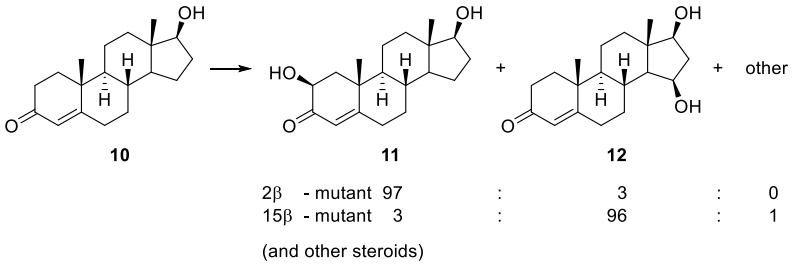
Another recent example of "difficult-to-reduce" ketones concerns TbSADH-catalyzed asymmetric reduction of ketones in which the α- and α'-groups flanking the carbonyl function are also sterically very similar (Scheme 10).[15b] In this case, Triple Code Saturation Mutagenesis (TCSM) at a large randomization site was applied, which was cut in half for (*R*)- and (*S*)-selectivity, respectively. Different reduced amino alphabets comprising only 3 amino acids were rationally chosen. In this case, it is the (*S*)-product which is highly desirable because it is the precursor of the therapeutic drug Amprenavir, a new HIV inhibitor.



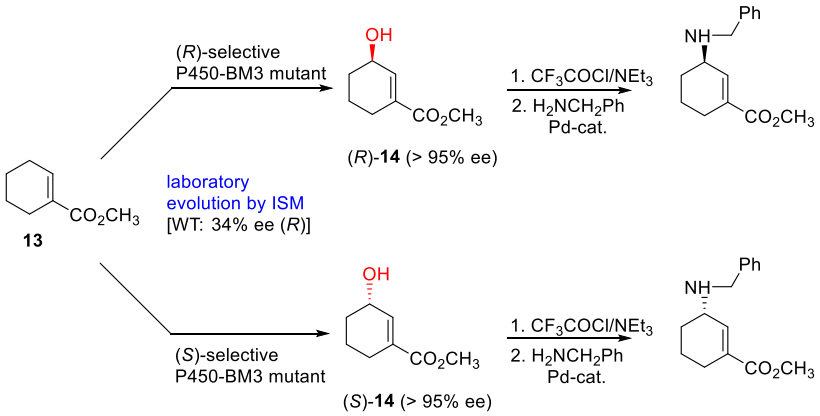
Scheme 10. ADH-catalyzed asymmetric reduction of difficult-to-reduce ketones using mutants evolved by TCSM (ee = ≥95% ee) either with (*R*)- *or* (*S*)-selectivity.[15b]

We have applied CAST/ISM to monooxygenases, including Baeyer-Villiger monooxygenases (BVMOs) for stereoselective oxidations of ketones[13,26] and

sulfoxidation of prochiral thio-ethers.[27] Cytochrome P450 monooxygenases have also been subjected to directed evolution by us and other groups, most often using CAST/ISM. Selected examples are shown in Scheme 11[28] and Scheme 12.[29] The starting P450-BM3(F87A) enzyme leads to a 1:1 mixture of steroid regioisomers, and essentially complete C2- and C15-regioselectivity was achieved with complete ß-diastereoselectivity (Scheme 11). The results depicted in Scheme 12 using olefin **13** as the substrate are the first examples of essentially complete regioselectivity *and* very high enantioselectivity (>95% ee) of *both* enantiomeric products in directed evolution.[29] The products were acylated and subsequently subjected to Pd-catalyzed regioselective amination with retention of configuration.[29] The respective amines are GABA-analogs and may be of interest to neuro- and/or pharmaceutical scientists. Other examples of P450-catalyzed regio- and stereoselective oxidative hydroxylation likewise involve products of pharmaceutical interest.[30] This includes *bioorthogonal enzymatic activation of caged compounds in living cells with release of an anti-cancer agent, catalyzed by regioselective P450-BM3 mutants in a proof-of-principle study*.[30c]

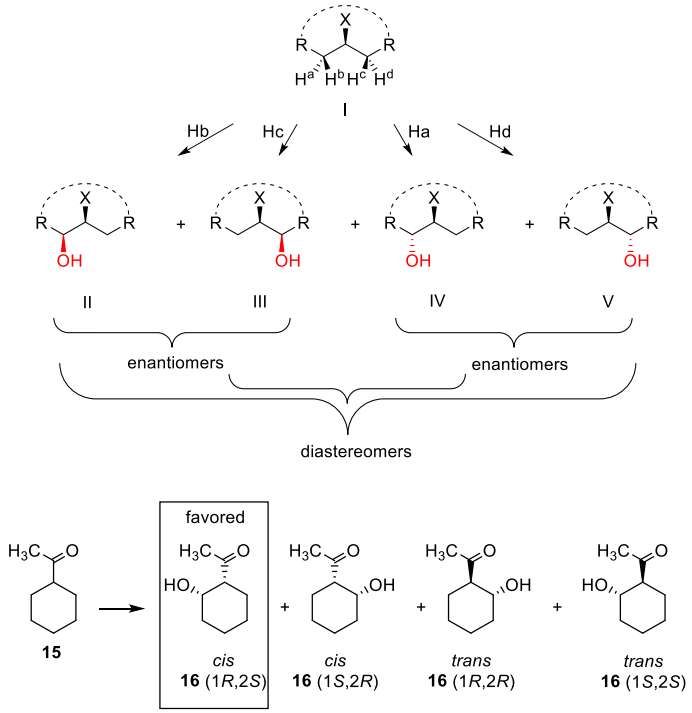


Scheme 11. P450-BM3 catalyzed regio- and stereoselective oxidative hydroxylation of testosterone (**10**).[28]



Scheme 12. P450-BM3 catalyzed regio- and enantioselective oxidative hydroxylation of ester **13** with 95% regioselectivity and >95% (*R*)- and (*S*)-enantioselectivity on an optional basis.[29]

Yet another study published by our group addresses the question whether a single CH-activating hydroxylation event can create not just one new center of chirality as in all previous cases, but simultaneously two chirality centers.[31] Such CH-activating processes convert simple starting materials in one step into value-added products of higher structural complexity. They constitute "dream reactions". Scheme 13 illustrates the general case of two new chirality centers, one of the eight published examples being shown here in which *an evolved P450-BM3 mutant ensures high regio-, diastereo- and enantioselectivity by focusing hydroxylation mainly on one of the four possible stereotopic H-atoms leading essentially only to product **16**-(1R,2S)*.[31] Eight different prochiral compounds showed a similar trend. Reviews covering these and other directed evolution studies of regio- and stereoselective P450 monooxygenases have appeared, which include references to works of L.-L. Wong, V. B. Urlacher, H. Zhao, F. H. Arnold, R. Bernhardt, R. Fasan, A. J. Mulholland and others.[32]



Scheme 13. Stereochemical consequences of CH-activating oxidative hydroxylation leading to the creation of two centers of chirality.[31]

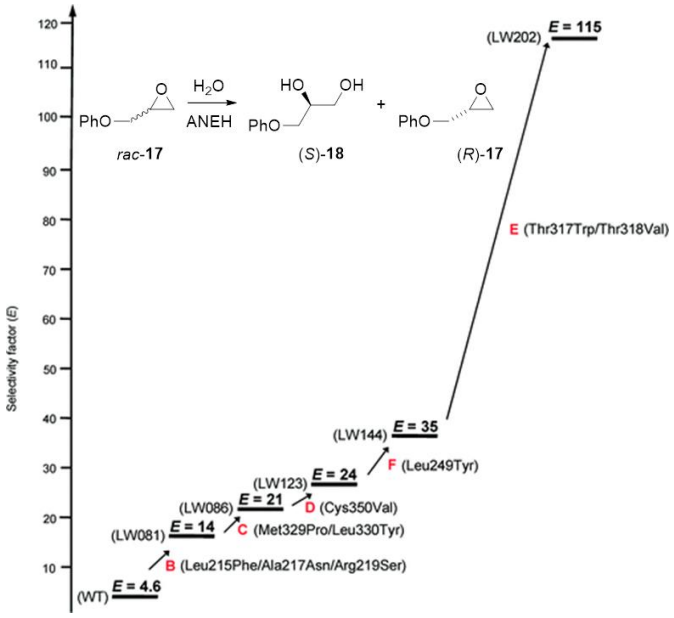
Learning from directed evolution

Two types of lessons can be learned from directed evolution as described herein, which require additional work. But the required efforts are well invested:

•*Uncovering the source of enhanced stereo- and/or regioselectivity by theoretical analyses based on docking computations, MD simulations and/or QM calculations, ideally flanked by X-ray structural analysis of the improved mutants.*

•*Performing deconvolution experiments in order to explore how point mutations and sets of point mutations interact with one another on a molecular level which allows for the construction of fitness pathway landscapes and the identification of epistatic effects.*

The first example of ISM, although at the time far from optimized, concerned the hydrolytic kinetic resolution of *rac*-**17** catalyzed by the epoxide hydrolases from *Aspergillus niger* (ANEH), WT being a poor catalyst with *E* = 4.6.[10] Five multi-residue CAST sites A, B, C, D and E were defined, and an arbitrarily chosen upward pathway B→ C→ D→ F→ E was transversed which provided a highly improved mutant LW202 with a selectivity factor of *E* = 115 in favor of (*S*)-**18** (Scheme 14). Enantioselectivity was already so high that the final site A was not visited. In this early first generation ISM study, 20,000 transformants were screened. The ISM technique has since been optimized with creation of higher-quality libraries requiring considerably less screening, generally just a few hundred transformants.[1,11,12]



Scheme 14. First example of iterative saturation mutagenesis (ISM).[10]

It was important to understand the reason(s) for enhanced enantioselectivity of mutant LW202 on a molecular level. The mechanism of ANEH involves substrate activation by H-bonding of two tyrosines to the oxygen-atom of the epoxide and

rate-determining S_N2-attack by Asp192 followed by rapid hydrolysis of the short-lived enzyme-ester intermediate. *An extensive structural and mechanistic study was undertaken which included kinetics, inhibition experiments, crystal structure of best mutant LW202 and molecular dynamics (MD) computations*.[33] The X-ray structures of WT ANEH and best mutant LW202 are essentially superimposable, which shows that the overall fold of the enzyme has not changed. However, upon inspecting the structural features at the respective active sites, it was discovered that the shape of ANEHs binding pocket has changed considerably.[33] One of several graphics as reported in the original study is reproduced here in Fig. 1. At the time, *X-ray structures of stereoselective mutants obtained by directed evolution were rare*.[1] Without such "hard" structural data, docking and MD computations alone would have remained somewhat speculative.

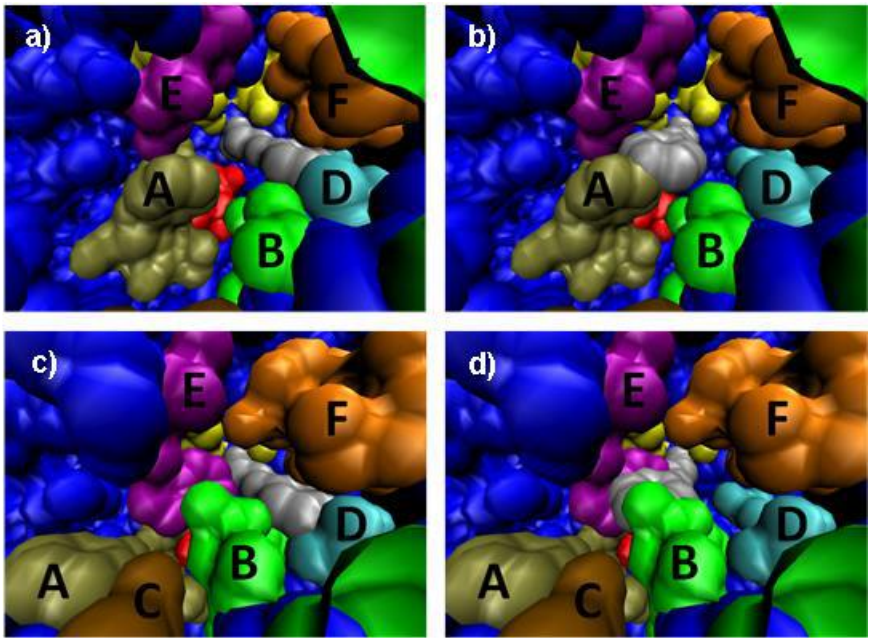


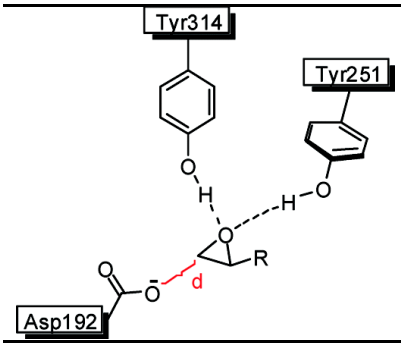
Fig. 1. a) Binding pocket of WT ANEH in which (*S*)-**17** (gray) has been docked; b) Binding pocket of WT ANEH in which (*R*)-**17** has been docked; c) Binding pocket of mutant LW202 in which (*S*)-**17** has been docked; d) Binding pocket of LW202 in which (*R*)-**17** has been docked.[33]

It can be seen that both enantiomers of the substrate (gray) are well accommodated in WT binding pocket, activation via hydrogen bonds from two tyrosines to the epoxide O-atom as well as close vicinity of the active nucleophile Asp192 ensuring smooth ring-opening reaction. In contrast, the geometric characteristics of the reshaped binding pocket of the evolved (*S*)-selective mutant LW2002 are quite different. (*S*)-**17** fits very well, while docking (*R*)-**17** leads to severe steric clashes. These and other insights[33] not only shed light on the source of enhanced enantioselectivity, but also on the intricacies of catalytic

machineries of enzymes, in this case the mechanistic details of type-1 epoxide hydrolases.

MD computations were likewise eye-opening. It became clear that smooth reaction can occur only if two prerequisites are fulfilled: The substrate must adopt a pose in the binding pocket in which 1) Tyr251 and Tyr314 activate the epoxide ring via H-bonding, and 2) The distance d between the O-atom of Asp192 and the C-atom of the epoxide should not be longer than ≈4 A (Table 3). It can be seen that as the evolutionary 5-step process proceeds, the distance d increases continually in the case of disfavored (*R*)-**17**, reaching a maximum (5.4 A) in the last cycle (mutant LW202). Therefore, its reaction has been shut down, which is in accord with the kinetic results.[33]

Table 3. Results of MD computations at each evolutionary stage in the directed evolution of *Aspergillus niger* epoxide hydrolase (ANEH) as the catalyst in the hydrolytic kinetic resolution of *rac*-**17**.[33]

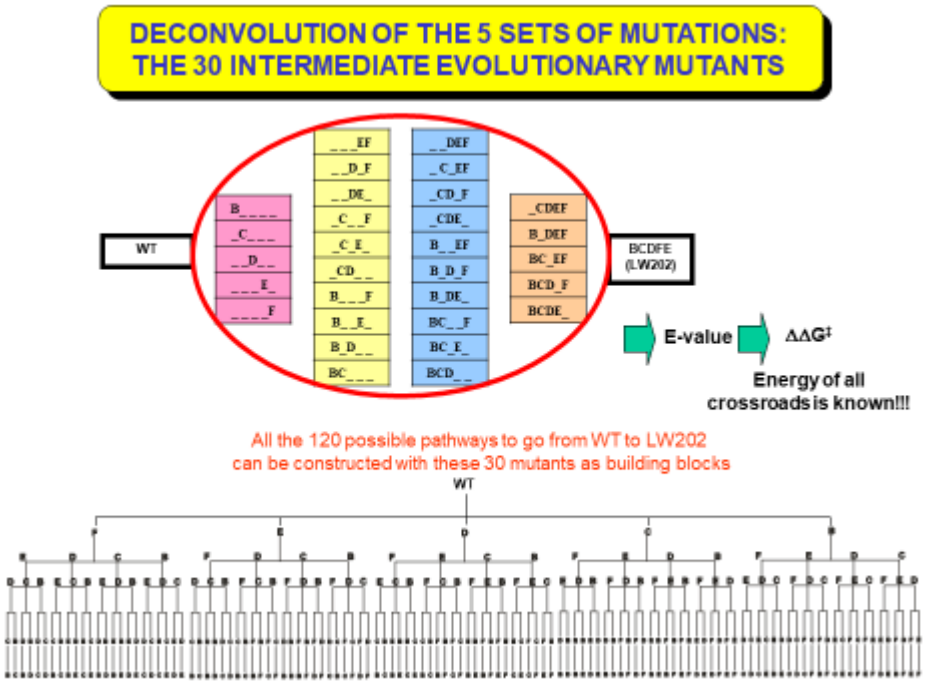


mutant	d_R	d_S	Δd_{R-S}	*E* (expl)
WT	4.3	3.5	0.8	4.6
LW081	4.8	4.0	0.8	14
LW086	4.9	4.0	0.9	21
LW123	5.1	4.0	1.1	24
LW44	5.1	3.9	1.2	35
LW202	5.4	3.8	1.6	115

Subsequently, limonene epoxide hydrolase (LEH), in which water is the attacking nucleophile (type-2 epoxide hydrolase), was subjected to directed evolution using the desymmetrization of cyclohexene oxide as the model reaction.[34] *The study includes several X-ray structures of mutants with enhanced and reversed enantioselectivity*, which in combination with docking computations shed light on the mechanistic intricacies of this type of epoxide hydrolases.[34] It is the first study in which crystal structures of both evolved enantiomers were reported, a textbook example illustrating the spirit of Emil Fischer's lock-and-key hypothesis.

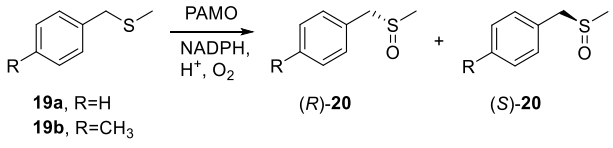
As already alluded to, *the second type of lesson to be learned from directed evolution is based on deconvolution of final multi-mutational variants*, especially when coupled with systematic theoretical analyses. When applying CAST/ISM (or any other directed evolution method), point mutations or sets of point mutations accumulate in each mutagenesis/screening cycle. Only the cumulative

effect is measured in the screening step, e.g., enantioselectivity, rate, or thermostability. The performance of each separate mutational increment in the overall accumulation is not known, with obvious exception of the initial point mutation or initial set of mutations. However, if one invests efforts by performing deconvolution with the creation of the respective variants corresponding to each new point mutation or each new set of mutations, then the separate "contributions" become accessible. *We have performed a number of complete and partial deconvolution studies focused on enantioselectivity, and made the surprising discovery that in most cases traditional mutational additivity is not operating, in contrast to cooperativity (more than additivity!)*.[35] Typical examples include an epoxide hydrolase[36] and a lipase[14a] (see Section above on Methodology Development). In both cases enantioselectivity was the catalytic parameter of interest. In the case of the epoxide hydrolase from *Aspergillus niger*, deconvolution led to 5! = 120 pathways from WT (E =) to the best mutant LW202 at the time for a hydrolytic kinetic resolution (*E* = 115).[36] This means that in order to construct the fitness pathway landscape, 30 mutants had to be generated by site directed mutagenesis as noted in Scheme 15. [36] It allows the assessment of all intermediate single, double and triple mutants and the experimental construction of all 120 pathways on the basis of $\Delta\Delta G^{\ddagger}$ values. Pronounced non-additive effects (meaning *more* than additivity!) were unveiled, which correlate with strong cooperative (synergistic) effects on a molecular level. In extreme cases, two different mutations separately may each favor (*R*)-selectivity, but together they act in concert to induce (*S*)-selectivity! In 2013 we published a short summary of our results originating from different studies and compared them with earlier studies reporting strict additivity, including the important reports by A. Fersht and J. A. Wells.[35] On the basis of the emerging picture, I believe that *mutational effects in proteins often constitute non-linear stochastic systems which are more prevalent than expected,* which is of fundamental importance in enzymology.



Scheme 15. Complete deconvolution of the best CAST/ISM-obtained mutant of the epoxide hydrolase from *Aspergillus niger* as the catalyst in the hydrolytic kinetic resolution of a racemic epoxide, the respective fitness pathway landscape being pictured in reference[36] .

An example of extreme cooperative mutational effects concerns the directed evolution of the Baeyer-Villiger monooxygenase phenylacetone monooxygenase (PAMO) as a catalyst in asymmetric sulfoxidation with thioether **19b** serving as the model compound (Scheme 16).[27] WT PAMO favors the formation of (*S*)-**20b** with 90% ee. The purpose of this ISM-based directed evolution project was to enhance (*S*)-selectivity and to invert enantioselectivity with preferential formation of (*R*)-**20b.**



Scheme 16. PAMO-catalyzed asymmetric sulfoxidation.[27]

Both enhanced (*S*)- and reversed (*R*)-selectivity were indeed achieved. The best (*R*)-selective variant ZGZ-2 (95% ee) proved to be the quadruple mutant I67Q/P440F/A442N/L443I. Thus, reversal of stereoselectivity in going from WT PAMO (90% ee, *S*) to mutant ZGZ-2 (95% ee, *R*) involves an unusually large energy change:

7.3 kJ/mol + 9.1 kJ/mol = 16.4 kJ/mol

Surprisingly, *deconvolution revealed that all four single mutants of ZGZ-2, namely I67Q, P440F, A442N and L443I are (S)-selective with ee-values of 69%, 97%, 69% and 98%, respectively!* In concert, the four point mutations induce dramatic synergistic (cooperative) effects on a molecular level leading to reversed (*R*)-selectivity. The synergistic mutational effects were explained by MD/docking computations.[27] All of the designed respective double and triple mutants which are combinatorially possible using the four point mutations of variant ZGZ-2 were also generated and tested in the model reaction. Here again unusual results were observed which do not correspond to traditional additive mutational effects. In terms of free energies, application of Eq. 1 reveals either traditional additive or non-additive mutational effects.[35]

$$\boldsymbol{\Delta G_{ij}^{\ddagger} = \Delta\Delta G_{exp}^{\ddagger} - (\Delta\Delta G_{i}^{\ddagger} + \Delta\Delta G_{j}^{\ddagger})} \qquad \text{Eq. 1}$$

where

$\boldsymbol{\Delta G_{ij}^{\ddagger}}$ is the free energy of interaction between two mutions or sets of mutations,

and $\boldsymbol{\Delta\Delta G_{i}^{\ddagger}}$, $\boldsymbol{\Delta\Delta G_{j}^{\ddagger}}$, and $\boldsymbol{\Delta\Delta G_{exp}^{\ddagger}}$ are the experimentally determined differences in activation energies for the formation of both enantiomers using mutants *i*, *j*, and the binary combination of *i* and *j*, respectively.

Upon generating all combinatorial combinations of double and triple mutants, it was possible to construct a fitness pathway landscape leading from WT PAMO to the best (*R*)-selective mutant ZGZ-2.[27] This involves 4! = 24 upward pathways (trajectories). The experimental results are shown in Figure 2:

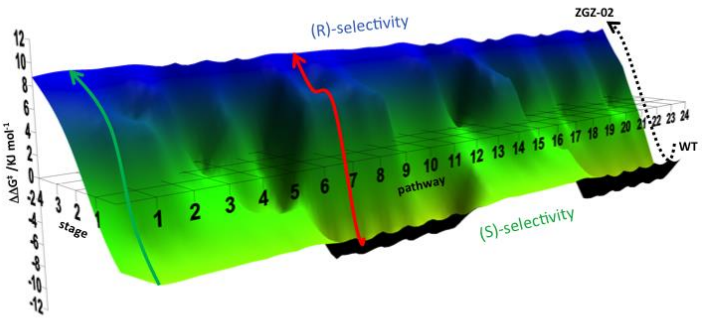
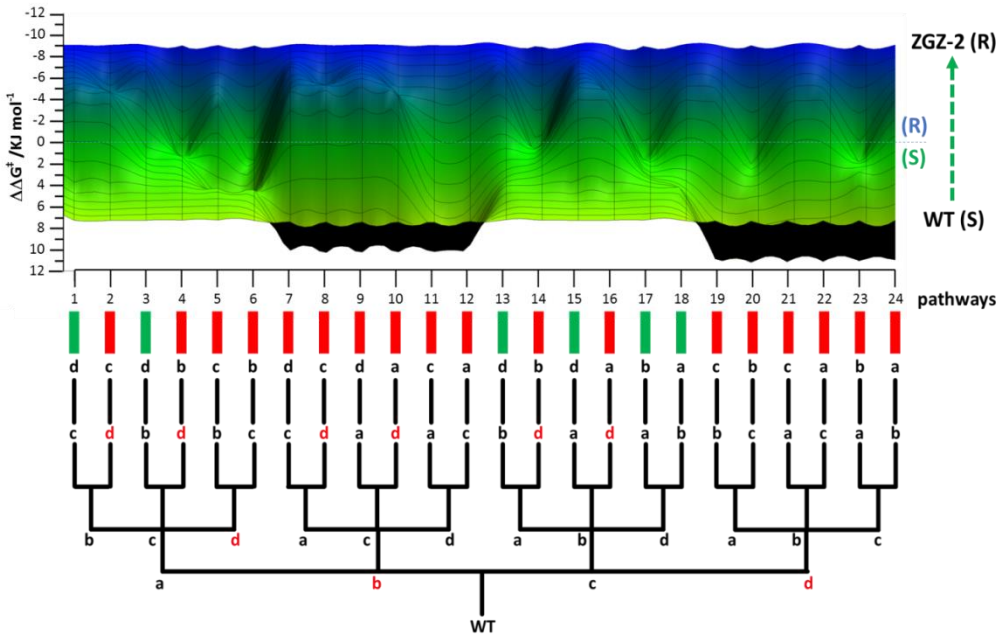


Fig. 2. Fitness pathway landscape showing the 24 pathways leading from WT PAMO displaying high (*S*)-selectivity (bottom) to best (*R*)-selective variant ZGZ-2, a typical trajectory lacking local minima (green pathway) and one having local minima (red) being featured.[27]

The complete profile of all 24 pathways is shown in Scheme 17. It can be seen that 6 pathways are energetically "favored", lacking any local minima (green trajectories), whereas 18 are energetically "disfavored" because local minima exist (red trajectories). Of course, all 24 pathways end up at the final mutant ZGZ-2, because this is an automatic consequence of the experimental setup. This simply means that strong cooperative effects acting between point mutations and sets of mutations must be occurring (see below).



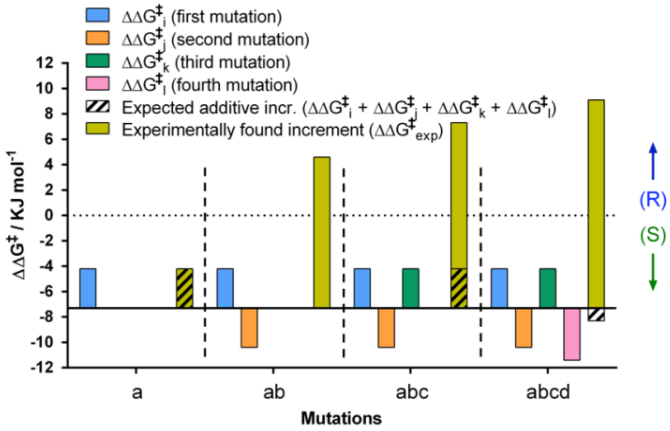
Scheme 17. Fitness pathway landscape in the frontal view of Fig. 2 featuring all 24 trajectories leading from WT PAMO to variant ZGZ-2 characterized by four point mutations.[27] Green notations indicate energetically favored pathways, whereas red notations stand for disfavored trajectories having local minima. A letter in red in the dendrogram denotes a local minimum after the introduction of this mutation.

This type of analysis of the empirical results allows the detailed study of additive versus non-additive (synergistic/cooperative/deleterious) mutational effects in all 24 pathways. For simplicity, we denote the four mutations as a, b, c, and d:

I67Q = a; P440F = b; A442N = c; L443I = d.

Only one of the 24 trajectories is analyzed here, namely a→b→c→d (Scheme 18). It can be seen that strong cooperative (non-additive) mutational effects are operating. For example, the second and third mutations (orange and pink, respectively) alone are contra-productive in enhancing (*R*)-selectivity, but in

concert with the other mutations a pronounced positive (synergistic) effect sets is!



Scheme 18. Thermodynamic cycle (Eq. 1) highlighting the interaction of point mutations and sets of mutations involved at every stage along the energetically favored pathway a→b→c→d toward variant ZGZ-2 as a catalyst in the sulfoxidation of substrate **19b**.[27]

The results teach us a lot about mutational effects in enzymology which have not been revealed previously in this branch of science. The effects were explained on a molecular level.[27] *As in the case of the epoxide hydrolase (Scheme 15), the lessons learned from such fitness pathway landscapes (Fig. 2 and Scheme 17) provide much more insight than other so-called fitness landscapes*.[37] Laboratory evolution experiments of this kind cannot be used to study phenomena in real Darwinian evolution, but the question remains: How often does Nature make use of cooperative effects in evolution?

This type of constrained fitness pathway landscape is different from non-constrained fitness pathway landscapes that result when going through all theoretically possible ISM pathways leading in each trajectory to a different mutant. *To date only one study has been published describing such a systematic CAST/ISM procedure in which all pathways were explored experimentally*.[20a] The epoxide hydrolase system in Scheme 13 was employed using only four ISM sites, meaning 24 pathways. All of them led to different mutants showing high enantioselectivity.[20a] In some of the pathways, local minima were encountered, meaning the absence of improved mutants in a given library. In those cases, it was possible to escape from such “dead ends” by choosing an inferior mutant as

the template in the subsequent saturation mutagenesis experiment. The problem of local minima occurs in all directed evolution systems irrespective of the mutagenesis method used. *The strategy delineated here, possibly related to the notion of neutral drift (D. Tawfik) or quasi species (M. Eigen), is likely to work in a general way in laboratory evolution.*

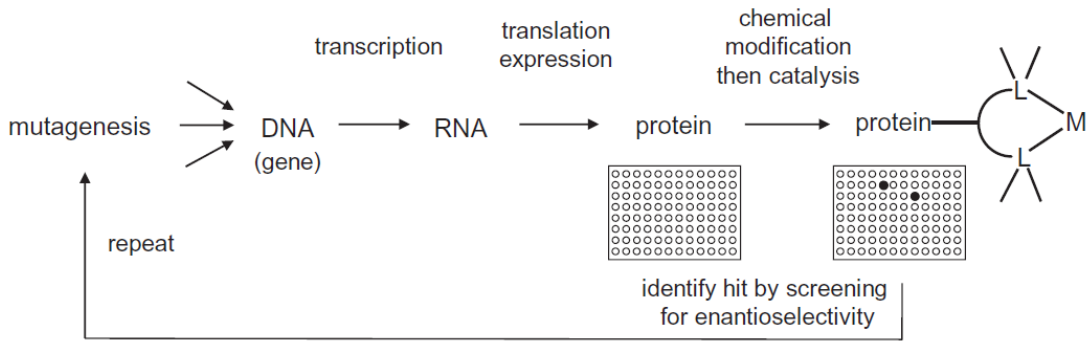
Directed evolution of thermostable enzymes: The B-FIT method

Directed evolution (or rational design) of enzymes for enhanced robustness constitutes an important research area that was already progressing well when we initiated research in engineering stereoselectivity in the mid-1990s.[1] The problem had been basically solved by the combined efforts of V. G. H. Eijsink, R. Hageman, J. A. Wells and F. H. Arnold. Nevertheless, years later when developing iterative saturation mutagenesis (ISM) for manipulating substrate scope, activity, stereoselectivity and/or regioselectivity, we posed *the question whether ISM can also be used for enhancing protein robustness.* In this endeavor, a criterion was needed for choosing appropriate randomization sites. CAST was certainly not appropriate. We therefore postulated that those residues displaying the highest B-factors available from X-ray data should be chosen. Those are the ones displaying highest flexibility, and we speculated that rigidification at these sites by appropriate mutagenesis would increase protein robustness. Accordingly, the B-FIT method was developed and illustrated using a lipase.[38] The B-FITTER computer aid provides a user-friendly guide for performing protein thermostabilization, available free of charge at www.kofo.mpg.de/en/research/biocatalysis (Reetz-homepage). The method is one of several successful approaches for enhancing thermostability of enzymes and robustness in organic solvents. In order to determine which strategy is fastest and most efficient, comparative studies are needed.

Directed evolution of hybrid catalysts (artificial metalloenzymes)

The concept of directed evolution of enantioselective enzymes is widely used by academic and industrial groups, as for example in the synthesis of therapeutic drugs.[1,12] One of many examples, published by Codexis in 2010, concerns reductive transaminase-catalyzed amination of a prochiral ketone with formation of Sitagliptin for the treatment of diabetes, in which CAST/ISM but

also epPCR and DNA shuffling were employed. However, no enzyme can mediate such transition metal catalyzed reactions as hydroformylation, hydrovinylation, allylic substitution, olefin metathesis, etc. We therefore proposed back in 2001/2002 *a concept which goes far beyond conventional directed evolution, namely laboratory evolution of selective hybrid catalysts*.[39] Promiscuous enzymes can be evolved by this approach. Accordingly, transition metal centers are anchored covalently (or non-covalently) to a suitable robust protein which functions as a host (its natural enzyme property is of no interest). However, this is not just performed once with the WT protein (which would not be new), but with thousands of mutants producing thousands of hybrid catalysts, later called artificial metalloenzymes (Scheme 19).[39-41] The best one displaying the highest activity/selectivity is then identified by a screening systems. Several such rounds comprising mutagenesis/chemical modification/screening exert an evolutionary pressure on the system, allowing for a fundamentally new type of ligand (catalyst) tuning. A number of difficult technical problems had to be solved before the concept could be generalized. Nevertheless, we were able to present proof-of-principle for the first time in a study utilizing iterative saturation mutagenesis (ISM).[41] It employed the Whitesides system of anchoring a biotinylated Rh-diphosphine complex non-covalently to (strept)avidin and using this artificial metalloenzyme as the catalyst in asymmetric olefin hydrogenation. We were able to increase enantioselectivity in a model reaction from 35%ee stepwise to 65%ee by applying CAST/ISM.[41] Thus, *a principally new approach to asymmetric catalysis was established in which the genetic techniques of molecular biology are used to tune a synthetic transition metal catalyst by exerting evolutionary pressure!*

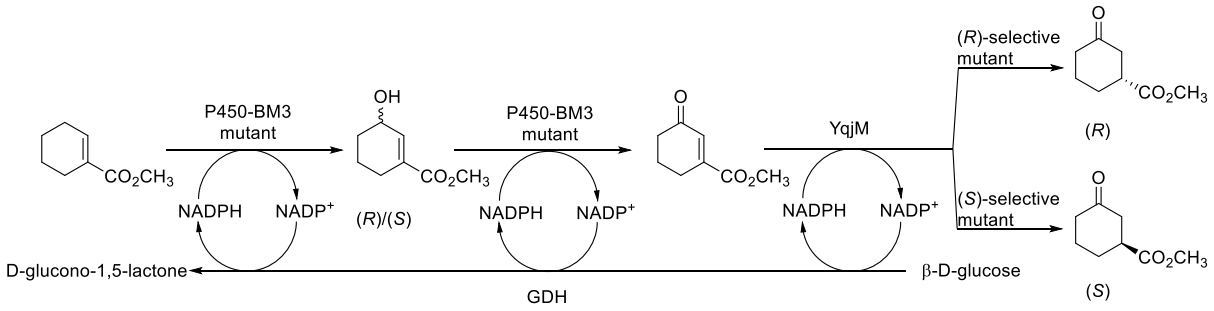


Scheme 19. Directed evolution of hybrid catalysts as artificial metalloenzymes.[39-41]

The technical difficulties in this particular system have been delineated in a recent review, which also covers other approaches to artificial metalloenzymes.[40a] It needs to be noted critically that more work is necessary in this research area. This would immensely expand the "application space" of proteins and allow for a wide range of stereoselective reactions unknown in nature (promiscuity), as an alternative to natural metalloenzymes as such. Advanced research is necessary, especially in designing systems that display notably *enhanced activity* relative to the respective transition metal catalyst not anchored to a host protein. A similar strategy is possible using achiral organocatalysts anchored to appropriate protein cavities. Here again, increasing activity is crucial, not just stereoselectivity.

Cascade reactions in designer cells enabled by directed evolution

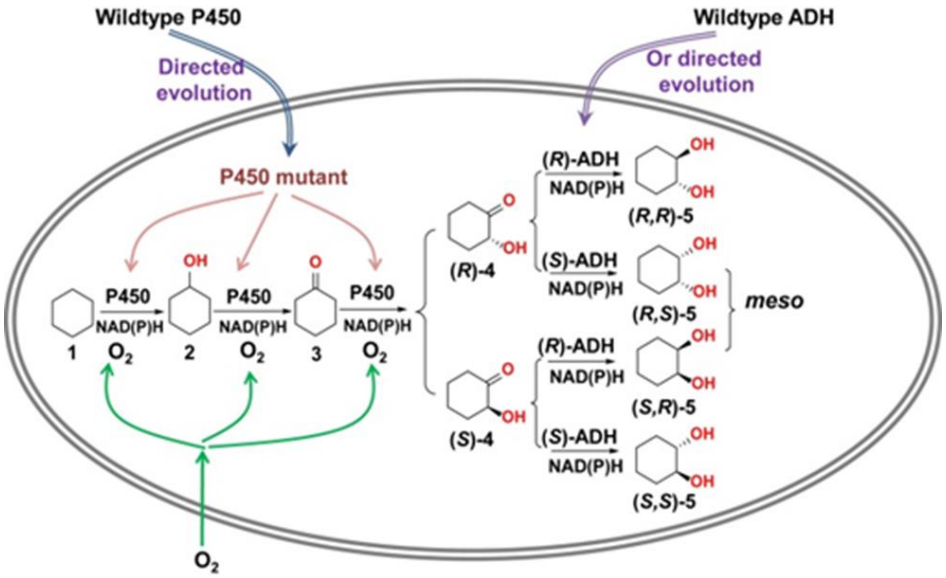
Cascade reactions catalyzed by man-made catalysts or enzymes have several advantages, including only a single workup which saves resources.[42] We have focused on whole cell biocatalyzed processes in which the mutant enzymes are used. *We applied this approach for constructing designer cells featuring regio- and stereoselective cascade reaction*. [43,44] For example, two different *E. coli* strains were engineered using P450-BM3 mutants and enoate reductase YqjM mutants that allow the conversion of cyclohexene carboxylic acid ester as feedstock into the (*R*)- or (*S*)-configurated keto-ester products on an optional basis (*99% ee in each case*)[43] (Scheme 20):



Scheme 20. Three-step cascade reactions using engineered (*R*)- and (*S*)- *E. coli* strains.[43] The P450-BM3 and YqjM mutants were evolved by CAST/ISM.

Another example from our group is the construction of *E. coli* designer cells that derive energy by consuming otherwise inert cyclohexane with formation of either (*R*)- or (*S*)-2-hydroxycyclohexanone, followed by diastereoselective ADH-

catalyzed reduction leading to the three stereoisomers of cyclohexanediol (Scheme 21).[44] The cascade can also be stopped after the first three steps with formation of the (*R*)- or (*S*)-acyloins on an optional basis.



Scheme 21. Designer *E. coli* cells enabled by directed evolution of P450-BM3 as the catalyst in the first three steps followed by the action of appropriate WT ADHs.[44]

In another development in collaboration with D. Belder, we reported the analysis of enantioselective *biotransformations catalyzed by enzyme mutants using only a few hundred cells on an integrated microfluidic chip*.[45]

Update of developments 2017-2020

During the final phase of our research, the following studies are particularly worthy of mention:

• *Method for simultaneous focused epPCR and saturation mutagenesis.* [46]

• *Simultaneous engineering of an enzyme's entrance tunnel and binding pocket.*[47]

• *A promiscuous redox-mediated Kemp eliminase.*[48]

• *Inducing high activity of a thermophilic enzyme at room temperature*.[49]

• *Structural and computational insight into mechanism of limonene epoxide hydrolase in stereoselective transformations.* [50]

• *Beating bias in directed evolution by combining high-fidelity on-chip solid-phase gene synthesis with efficient gene assembly for combinatorial library construction.* [51]

• *Targeted P450-catalyzed regio- and stereoselective steroid hydroxylation by mutability landscaping.* [52]

• *Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy*.[53]

• *Overriding traditional electronic effects in biocatalytic Baeyer-Villiger reactions*.[54]

• *A machine learning approach for reliable prediction of amino acid interactions and application in the directed evolution of enantioselective enzymes*.[55]

• *Chemo- and regioselective dihydroxylation of benzene and arbutin synthesis enabled by P450 mutants*.[56]

• *Utility of B-factors in protein science for interpreting rigidity, flexibility and internal motion and engineering thermostability*.[57]

• *Focused Rational Iterative Site-specific Mutagenesis (FRISM) as an efficient fusion of rational design and directed evolution*.[58]

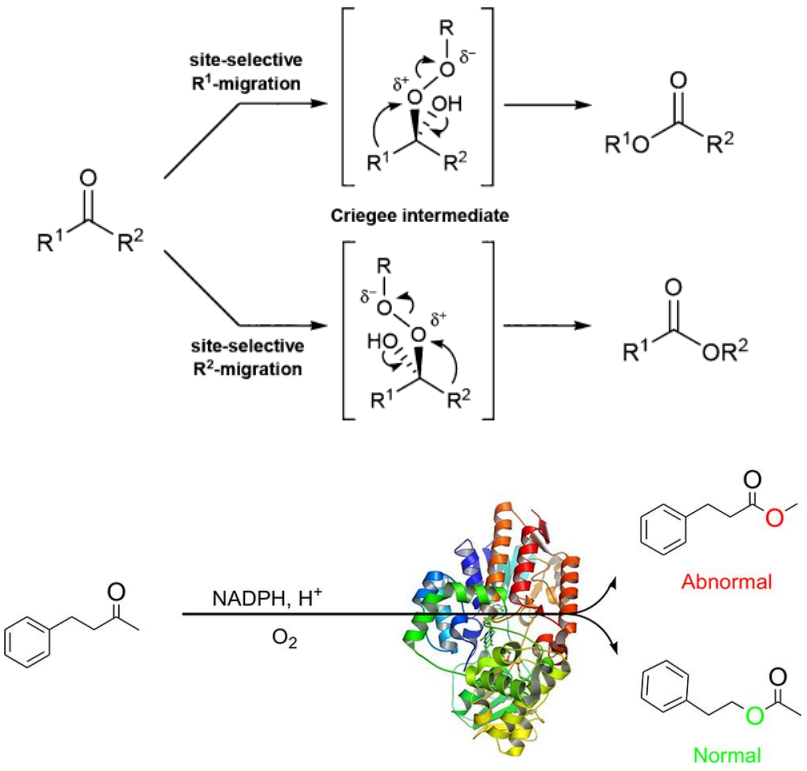
• *Exploiting designed oxidase-peroxygenase mutual benefit system for asymmetric cascade reactions*.[59]

• *Artificial cysteine-lipases with high activity and alter mechanism created by directed evolution*.[60]

• *P450-BM3 catalyzed sulfoxidation versus hydroxylation occurs via a common catalytic species as shown by directed evolution*.[61]

• *C7ß-selective hydroxylation of steroids by evolved P450-BM3 mutants with formation of therapeutic drugs of immense interest in the pharmaceutical community as neuroprotective and anti-inflammatory agents for the treatment of stroke, brain trauma and cerebral ischemia*.[62]

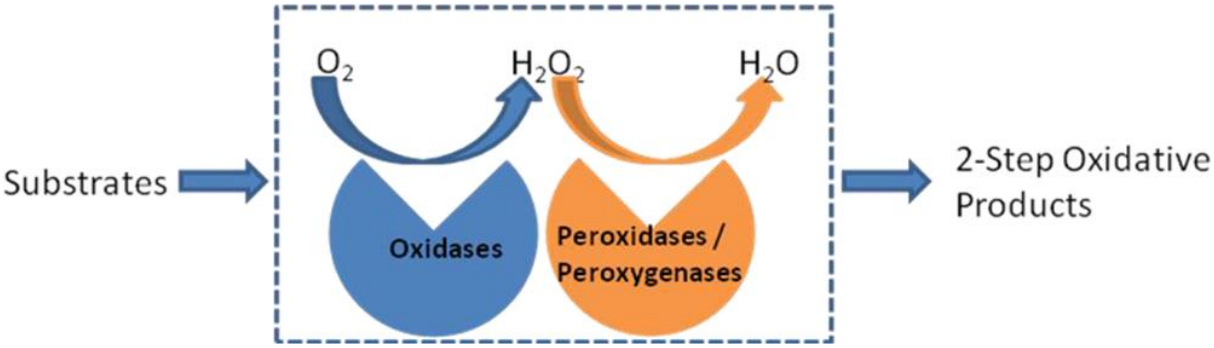
Only a few of these studies will be highlighted here, for the others the interested reader is referred to the original literature. One of them concerns a study in collaboration with K. N. Houk on directed evolution of a Baeyer-Villiger monooxygenase, specifically PAMO, for overriding traditional electronic effects so that only the abnormal product is formed (Scheme 22).[54] In addition to the model ketone shown in Scheme 22, several other structurally different ketones were shown to provide complete abnormal reaction modes, or enhanced normal products; in the case of benzyl ethyl ketone, mutants were generated that let either the benzyl or the ethyl group migrate! *These manipulated reactivity patterns are not possible using man-made catalysts.*



Scheme 22. Normal versus abnormal Baeyer-Villiger (BV) oxidations. Top: In organic chemistry long-known stereoelectronic effects in BV reactions. Bottom: Performance of an evolved Baeyer-Villiger monooxygenase for reversing regioselectivity in favor of the abnormal product.[54]

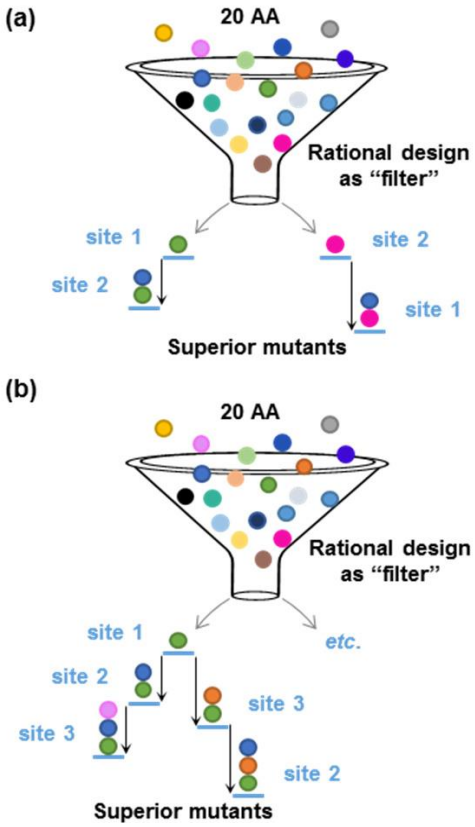
Another advance to be detailed here is the construction of a unique P450 monooxygenase-peroxygenase mutual benefit system (Scheme 23).[59] It was applied in a model cascade sequence leading from 3-phenyl propionic acid to (*R*)-phenyl glycol with high enantioselectivity. Accordingly, the monooxygenase

P450-BM3 and $OleT_{JE}$ as the P450 peroxygenase functioned as catalysts for the cascade steps, while also ensure internal an in situ H_2O_2 recyle mechanism which avoids its toxic accumulation leading to a breakdown of the catalytic system.



Scheme 23. Concept of a designed oxidase-peroxygenase mutual benefit system for asymmetric cascade reactions.[59]

Finally, a highly promising approach to fusing directed evolution and rational design is highlighted here, namely *Focused Rational Iterative Site-specific Mutagenesis (FRISM)*.[58] It was inspired by the success of CAST/ISM in so many studies,[9-13] the advantage of FRISM being that the generation and screening of saturation mutagenesis libraries is not necessary.[1c,58] FRISM requires the following steps: First, on the basis of the crystal structure of an enzyme to be worked on, CAST residues are identified. Then the usual techniques known in rational design, including mechanistic knowledge, consensus data, previous mutational effects, MD and other computational tools, are invoked to make 3-4 predictions for single amino acid exchange events at each CAST residue. This is followed by testing these few mutations in a screening process for activity and stereoselectivity, typically using GC or HPLC. The best mutant at one residue is then used as a template for the same procedure at another CAST residue, and so on. More than one pathway can be envisioned, depending upon the number of CAST residues chosen. The final best mutant can be expected to be different in each pathway, even if only 2 or 3 residues are considered in the iterative process, respectively (Scheme 24a/24b).[1c,58] This is so due to possible synergistic epistatic mutational effects (more than additive).



Scheme 24. Representation of Focused Rational Iterative Site-specific Mutagenesis (FRISM). [1c,58] A highly limited set of rationally chosen amino acids (AAs) are individually introduced at the chosen 1-residue CAST sites 1, 2, and 3 by site-specific mutagenesis in an iterative manner. a) The case of two mutagenesis sites involving only two pathways 1→2 and 2→1; b) The case of three sites, illustrated by two of the six possible pathways, namely 1→2→3 and 1→3→2.

Main conclusions and perspectives

On the basis of hundreds of academic and industrial studies, it was already clear back in 2015 or earlier that *our concept of directed evolution of stereoselective enzymes utilizing CAST/ISM* [4,6,12] *with extension to regioselectivity,* [28,29] *activity and substrate scope constitutes a general, reliable, prolific and unceasing source of catalysts for highly selective transformations in organic chemistry and biotechnology*.[1,2,6,12,27-32,62] Since its original inception in 1997,[4] it has been widely adopted by many academic and industrial groups. This approach to creating selective catalysts is complementary to chiral synthetic transition metal catalysts and organocatalysts. *Since CAST/ISM is structure-based and is intelligently guided by bioinformatics, MD simulations and/or machine learning, it has emerged as the most rational and reliable method in directed evolution*. Many more examples from academic and industrial labs can be expected in the

future, while further fine-tuning of ISM will probably continue. This also applies to synthetic metalloenzymes.[39,40] *Importantly, our work has popularized enzyme catalysis as a particularly sustainable, ecologically and economically viable approach for the production of products needed in modern society*.

Inspired by CAST/ISM, *the currently most effective fusion of rational design and directed evolution is Focused Rational Iterative Site-specific Mutagenesis (FRISM)*. [1c,58] It remains to be seen if it can replace the reliable performance of second generation CAST/ISM.

A currently important issue concerns the following question: As originally proposed,[21] *can combinatorial solid-phase gene synthesis of mutant libraries in directed evolution replace the traditional PCR-based technique of saturation mutagenesis? Recent work subsequent to our initial attempt*[21] *has clearly demonstrated that the quality of such synthetic libraries is distinctly superior (thus requiring less screening)*.[51] This was proven by massive sequencing of such a library and of the respective traditional saturation mutagenesis counterpart. While the former revealed the presence of 97% of the expected mutants on DNA level, the traditional library harbored only 56% of the expected mutants. Thus, by using the approach of synthetic gene libraries, the second law of directed evolution emerges: *"You get what you designed!"* The current prices of such commercial libraries are fairly high, generally too high for most academic groups. It remains to be seen whether the costs will continue to tumble in the coming years. In our eye-opening study, we collaborated with the biotech company Twist Bioscience,[51] but other companies also offer this commercial service. The potential customer needs to make sure that the high quality of synthetic saturation mutagenesis libraries (CAST) is actually ensured by the company with the corresponding data. *If so, then in the future, combinatorial solid-phase gene synthesis of the designed mutants for library creation will dominate the field of directed evolution*.[51]This gives the experimenter more time to concentrate on the essentials, namely to design the focused CAST mutant libraries or FRISM mutants optimally.

Directed evolution is also being increasingly applied in other areas such as the production of therapeutic antibodies and peptides as well as metabolic pathway engineering, in the latter case as an alternative to classical natural products

synthesis, among other goals.[63] *Other areas of applying CAST/ISM* already used by different groups refer to vaccine generation[64] and potential universal blood production,[65] while application in plant biology can be envisioned.

I thank all coworkers and collaborators who have participated in 25 years of efforts in developing the concept of directed evolution of selective enzymes as catalysts in organic and pharmaceutical chemistry. Without their creative and persisting inputs, the area of directed evolution will not have received the wide recognition of the chemical and pharmaceutical communities, and it continues to the present day!

References (complete list of all publications with names of coworkers and collaborators can be found on the Reetz-homepage: www.kofo.mpg.de/en/research/biocatalysis)

[1] *Perspective on biocatalysis in organic chemistry:* a) M. T. Reetz, *J. Am. Chem. Soc*. **2013**, *135*, 12480-12496; b) *Monograph on directed evolution:* M. T. Reetz, *Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology*, Wiley-VCH, Weinheim, **2016**; c) *Comprehensive review of methodology development in directed evolution of selective enzymes:* G. Qu, A. Li, C. G. Acevedo-Rocha, Z. Sun, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59,* 13204-13231.

[2] *Early review of directed evolution of stereoselective enzymes[6,12]*: M. T. Reetz, in: *Enzyme Catalysis in Organic Synthesis*, (Eds.: K. Drauz, H. Gröger, O. May), Wiley-VCH, Weinheim, **2012**, Vol. 1, pp.119-190.

[3] *Reviews of high-throughput ee-assays*: a) M. T. Reetz, *Angew. Chem. Int. Ed*. **2001**, *40*, 284-310; b) M. T. Reetz, in: *Enzyme Assays - High-throughput Screening, Genetic Selection and Fingerprinting*, (Ed.: J.-L. Reymond), Wiley-VCH, Weinheim, **2006**, pp. 41-76; c) M. T. Reetz, *Methods Mol. Biol.* **2003**,*230*, 283-290; d) J. L. Reymond, *Enzyme assays: High-throughput screening, genetic selection and fingerprinting*, Wiley−VCH, Weinheim, **2006**; e) H. Xiao, Z. Bao, H. Zhao, *Ind. Eng. Chem. Res.* **2015**, *54*, 4011-4020; f) C. G. Acevedo-Rocha, R. Agudo, M. T. Reetz, *J. Biotechnol*. **2014**, *191*, 3-10; g) Y. V. Sheludko, W.-D. Fessner, *Curr. Opin Struct. Biol*. **2020**, *63*, 123-133; h) U. Schwaneberg, et al, *Chem. Soc. Rev*. **2020**, *49*, 233-262.

[4] *First case of directed evolution of an enantioselective enzyme*: a) M. T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, K.-E. Jaeger, *Angew. Chem. Int. Ed. Engl*. **1997**, *36*, 2830-2832; b) M. T. Reetz, *Pure Appl. Chem.* **1999**, *71*, 1503-109.

[5] M. T. Reetz, S. Wilensek, D. Zha, K.-E. Jaeger, *Angew. Chem. Int. Ed.* **2001**, *40*, 3589-3591.

[6] *Review of directed evolution of stereoselective enzymes up to 2004*: M. T. Reetz, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 5716-5722.

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

[8] 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.

[9] *Definition and first systematic use of CAST for stereoselectivity based on a previous study[5]*: M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem. Int. Ed.* **2005**, *44*, 4192-4196.

[10] *First example of ISM for improved enzyme catalysis*: a) M. T. Reetz, L.-W. Wang, M. Bocola, *Angew. Chem. Int. Ed.* **2006**, *45*, 1236-1241; b) M. T. Reetz, *Max-Planck-Society Yearbook* **2005**, pp. 327-331.

[11] *Practical guide to the use of ISM*: C. G. Acevedo-Rocha, S. Kille, M. T. Reetz, *Methods in Molecular Biology*, Humana Press, Totowa, **2014**, Vol 1179, pp 103-128.

[12] *Review of directed evolution of stereoselective enzymes with emphasis on ISM*: M. T. Reetz, *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174.

[13] a) M. T. Reetz, S. Wu, *Chem. Commun.* **2008**, 5499-5501; b) *Review of different strategies for applying saturation mutagenesis at the binding pocket in order to control stereoselectivity:* Z. Sun, Y. Wikmark, J.-E. Bäckvall, M. T. Reetz, *Chem. Eur. J*. **2016**, *22*, 5046-5054.

[14] a) M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.* **2010**, *132*, 9144-9152; b) S. Prasad, M. Bocola, M. T. Reetz, *ChemPhysChem* **2011**, *12*, 1550-1557.

[15] a) Sun, R. Lonsdale, L. Wu, G. Li, A. Li, J. Wang, J. Zhou, M. T. Reetz, *ACS Catal*. **2016**, *6*, 1590-1597; b) Z. Sun, R. Lonsdale, A. Ilie, G. Li, J. Zhou, M. T. Reetz, *ACS Catal*. **2016**, *6*, 1598-1605; c) G. Li, H. Zhang, Z. Sun, X. Liu, M. T. Reetz, *ACS Catal*. **2016**, *6*, 3679-3687; d) A. Li, A. Ilie, Z. Sun, R. Lonsdale, J.-H. Xu, M. T. Reetz*, Angew. Chem. Int. Ed.* **2016**, *55*, 12026-12029; e) A. Li, G. Qu, Z. Sun, M. T. Reetz, *ACS Catal*. **2019**, *9*, 7769-7778.

[16] 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.

[17] a) D. J. Bougioukou, S. Kille, A. Taglieber, M. T. Reetz, *Adv. Synth. Catal.* **2009**, *351*, 3287-3305; b) *QM/MM study of the ene-reductase YqjM*: R. Lonsdale, M. T. Reetz, *J. Am. Chem. Soc*. **2015**, *137*, 14733-14742.

[18] X. Feng, J. Sanchis, M. T. Reetz, H. Rabitz, *Chem.-Eur. J.* **2012**, *18*, 5646-5654.

[19] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z.-G. Zhang, J. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol*. **2013**, *2*, 83-92.

[20] a) Y. Gumulya, J. Sanchis, M. T. Reetz, *ChemBioChem* **2012**, *13*, 1060-1066; see also b) Y. Gumulya, M. T. Reetz, *ChemBioChem* **2011**, *12*, 2502-2510.

[21] S. Hoebenreich, F. E. Zilly, C. G. Acevedo Rocha, M. Zilly, M. T. Reetz, *ACS Synth. Biol.* **2015**, *4*, 317-331.

[22] C. G. Acevedo-Rocha, M. T. Reetz, Y. Nov, *Sci. Reports*, **2015**, *5*: 10654.

[23] Q. Wu, P. Soni, M. T. Reetz, *J. Am. Chem. Soc.* **2013**, *135*, 1872-1881.

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

[25] R. Agudo, G.-D. Roiban, M. T. Reetz, *J. Am. Chem. Soc.* **2013**, 135, 1665-1668.

[26] *Studies focusing on directed evolution of Baeyer-Villiger monooxygenases*: a) M. T. Reetz, S. Wu, *J. Am. Chem. Soc.* **2009**, *131*, 15424-15432; b) S. Wu, J. P. Acevedo, M. T. Reetz, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2775-2780; c) Z.-G. Zhang, G.-D. Roiban, J. P. Acevedo, I. Polyak, M. T. Reetz, *Adv. Synth. Catal.* **2013**, *355*, 99-106; d) *Review of directed evolution of Baeyer-Villiger monooxygenases*: Z.-G. Zhang; L. P. Parra, M. T. Reetz, *Chem. Eur. J.* **2012**, *18*, 10160-10172.

[27] Z.-G. Zhang, R. Lonsdale, J. Sanchis, M. T. Reetz, *J. Am. Chem. Soc.* **2014**, *136*, 17262-17272.

[28] S. Kille, F. E. Zilly, J. P. Acevedo, M. T. Reetz, *Nat. Chem.* **2011**, *3*, 738-743.
[29] a) R. Agudo, G.-D. Roiban, M. T. Reetz, *ChemBioChem* **2012**, *13*, 1465-1473; b) *Review of protein engineering of regioselectivity:* J. Wang, G. Li, M. T. Reetz, *Chem. Commun*. **2017**, *53*, 3916-3928.
[30] a) G.-D. Roiban, R. Agudo, A. Ilie, R. Lonsdale, M. T. Reetz, *Chem. Commun.* **2014**, *50*, 14310-14313; b) R. Agudo, G.-D. Roiban, R. Lonsdale, A. Ilie, M. T. Reetz, *J. Org. Chem.* **2015**, *80*, 950-956; c) C. Ritter, N. Nett, C. G. Acevedo-Rocha, R. Lonsdale, K. Kräling, F. Dempwolff, S. Hob, P. L. Graumann, M. T. Reetz, E. Meggers, *Angew. Chem. Int. Ed*. **2015**, *54*, 13440-13443.
[31] G.-D. Roiban, R. Agudo, M. T. Reetz, *Angew. Chem. Int. Ed*. **2014**, *53*, 8659-8663.
[32] *Reviews of directed evolution of regio- and stereoselective P450 monooxygenases:* G.-D. Roiban, M. T. Reetz, *Chem. Comm*. **2015**, *51*, 2208-2224; b) Z. Sun, M. T. Reetz, Controlling the Regio- and Stereoselectivity of Cytochrome P450 Monooxygenases by Directed Evolution, in *Dioxygen-dependent Heme Enzymes* (Eds. M. Ikeda-Saitao, E. Raven), RSC, London, **2018**.
[33] 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.
[34] Z. Sun, R. Lonsdale, X.-D. Kong, J.-H. Xu, J. Zhou, M. T. Reetz, *Angew. Chem. Int. Ed.* **2015**, *54*, 12410-12415.
[35] *Review of additive versus non-additive mutational effects in protein engineering*: M. T. Reetz, *Angew. Chem. Int. Ed*. **2013**, *52*, 2658-2666.
[36] M. T. Reetz, J. Sanchis, *ChemBioChem* **2008**, *9*, 2260-2267.
[37] J. D. Bloom, Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9995-10000.
[38] a) M. T. Reetz, J. D. Carballeira, *Nature Protoc.* **2007**, *2*, 891-903; b) M. T. Reetz, J. D. Carballeira, A. Vogel, *Angew. Chem. Int. Ed.* **2006**, *45*, 7745-7751**.**
[39] a) M. T. Reetz, patent: "Optimization of Synthetic Catalysts by Means of Directed Evolution", DE-A 101 29 187.6 (**19. June 2001**); PCT/PT 02/05596; b) M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, C*himia* **2002**, *56*, 721-723.
[40] *Reviews of directed evolution of hybrid catalysts (artificial metalloenzymes)*: a) M. T. Reetz, *Acc. Chem. Res.* **2019**, *52*, 336-344; b) 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; c) M. T. Reetz, *Chem. Rec.* **2012**, *12*, 391-406; d) A. Ilie, M. T. Reetz, *Isr. J. Chem*. **2015**, *55*, 51-60.
[41] M. T. Reetz, J. J.-P. Peyralans, A. Maichele, Y. Fu, M. Maywald, *Chem. Commun.* **2006**, 4318-4320.
[42] *Reviews of biocatalytic cascade reactions*: a) Faber, K.; W. D. Fessner, W. D.; Turner, N. J. *Adv. Synth. Catal.* **2015**, *357*, 1565–1566; b) Turner, N. J. *ACS Catal*. **2017**, *7*, 710–724; c) Kroutil, W. *Chem. Rev.* **2018**, *118*, 270–348; d) Bornscheuer, U. T. *Nat. Catal*. **2018**, *1*, 12–22; e) Wu, S.; Li, Z. *ChemCatChem* **2018**, *10*, 2164–2178.
[43] R. Agudo, M. T. Reetz, *Chem. Comm*. **2013**, *49*, 10914-10916.
[44] A. Li, A. Ilie, Z. Sun, R. Lonsdale, J.-H. Xu, M. T. Reetz, *Angew. Chem. Int. Ed*. **2016**, *55*, 12026-12029.
[45] K. M. Krone, R. Warias, C. Ritter, C. G. Acevedo-Rocha, M. T. Reetz, D. Belder, *J. Am. Chem. Soc*. **2016**, *138*, 2102-2105.
[46] J. P. Acevedo, M. T. Reetz, J. A. Asenjof, L. P. Parra, *Enzyme Microbiol. Technol*. **2017**, *100*, 60-70.
[47] G. Li, P. Yao, R. Gong, J. Li, P. Liu, R. Lonsdale, Q. Wu, J. Lin, D. Zhu, M. T. Reetz, *Chem. Sci*. **2017**, *8*, 4093-4099.
[48] A. Li, B. Wang, A. Ilie, K. D. Dubey, G. Bange, I. V. Korendovych, S. Shaik,

M. T. Reetz, *Nature Comm*. **2017**, *8*: 14876.
[49] G. Li, M. A. Maria-Solano, A. Romero-Rivera, S. Osuna, M. T. Reetz, *Chem. Comm*. **2017**, *53*, 9454-9457
[50] Z. Sun, L. Wu, M. Bocola, H. C. Chan, R. Lonsdale, X.-D. Kong, S. Yuan, J. Zhou, M. T. Reetz, *J. Am. Chem. Soc.* **2018**, *140*, 310-318.
[51] *Key paper and review of chemical solid-phase gene synthesis for library construction in directed evolution*: a) A. Li, C. G. Acevedo-Rocha, Z. Sun, T. Cox, J. (L) Xu, M. T. Reetz, *ChemBioChem* **2018**, *19*, 221-228; b) A. Li, Z. Sun, M. T. Reetz, *ChemBioChem*. **2018**, *19*, 2023-2032.
[52] C. G. Acevedo-Rocha, C. Gamble, R. Lonsdale, A. Li, N. Nett, S. Hoebenreich, J. B. Lingnau, C. Wirtz, C. Fares, H. Hinrichs, A. Deege, A. J. Mulholland, Y. Nov, D. Leys, K. J. McLean, A. W. Munro, M. T. Reetz, *ACS Catal*. **2018**, *8*, 3395-3410.
[53] A. Li, C. G. Acevedo-Rocha, M. T. Reetz, *Appl. Microbiol. Biotechnol*. **2018**, *102*, 6095-610
[54] G. Li, M. Garcia-Borras, M. Fürst, A. Ilie, M. Fraaije, K. N. Houk, M. T. Reetz, *J. Am. Chem. Soc*. **2018**, *140*, 10464-1047
[55] *Key paper and mini-review of machine learning used in directed evolution of stereoselective enzymes*: a) F. Cadet, N. Fontaine, G. Li, J. Sanchis, M. N. F. Chong, R. Pandjaitan, I. Vetrivel, B. Offmann, M. T. Reetz, *Sci. Reports* **2018**, *8*: 16757; b) G. Li, Y. Dong, M. T. Reetz, *Adv. Synth. Catal.* **2019**, *361*, 2377-2386.
[56] H. Zhou, B. Wang, F. Wang, X. Yu, L. Ma, A. Li, M. T. Reetz, *Angew. Chem. Int. Ed*. **2019**, *58*, 764-768.
[57] Z. Sun, Q. Liu, G. Qu, Y. Feng, M. T. Reetz, *Chem. Rev*. **2019**, *119*, 1626-1665
[58] a) J. Xu, Y. Cen, W. Singh, J. Fan, L. Wu, X. Lin, J. Zhou, M. Huang, M. T. Reetz, Q. Wu, *J. Am. Chem. Soc.* **2019**, *141*, 7934-7945; b) D. Li, Q. Wu, M. T. Reetz, *Methods in Enzymology*, **2020**; DOI: 10.1016/bs.mie.2020.04.055.
[59] D. Yu, J. Wang, M. T. Reetz, *J. Am. Chem. Soc*. **2019**, *141*, 5655-565.
[60] Y. Cen, W. Singh, M. Arkin, T. S. Moody, M. Huang, J. Zhou, Q. Wu, M. T. Reetz, *Nature Commun.* **2019**, *10*: 3198.
[61] J. Wang, Q. Huang, W. Peng, P. Wu, D. Yu, B. Chen, B. Wang, M. T. Reetz, *J. Am. Chem. Soc*. **2020**, *142*, 2068-2073.
[62] A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borras, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou, M. T. Reetz, *Angew. Chem. Int. Ed*. **2020**, *59*, 12499-12505.
[63] a) T. W. Johannes, H. Zhao, *Curr. Opin. Microbiol*. **2006**, *9*, 261-267; b) Y. Wang, X. Yu, H. Zhao, *AIChE J*. **2020**, *66*: e16716; c) S. Panke, et al, *Metabol. Eng*. **2020**, 59, 15-23.
[64] J. Ihssen, et al, *Open Biol*. **2015**, *5*: 140227.
[65] S. G. Withers, et al, *J. Am. Chem. Soc*. **2015**, *137*, 5695-5705.