1
Introduction to Directed Evolution

1.1
General Definition and Purpose of Directed Evolution of Enzymes

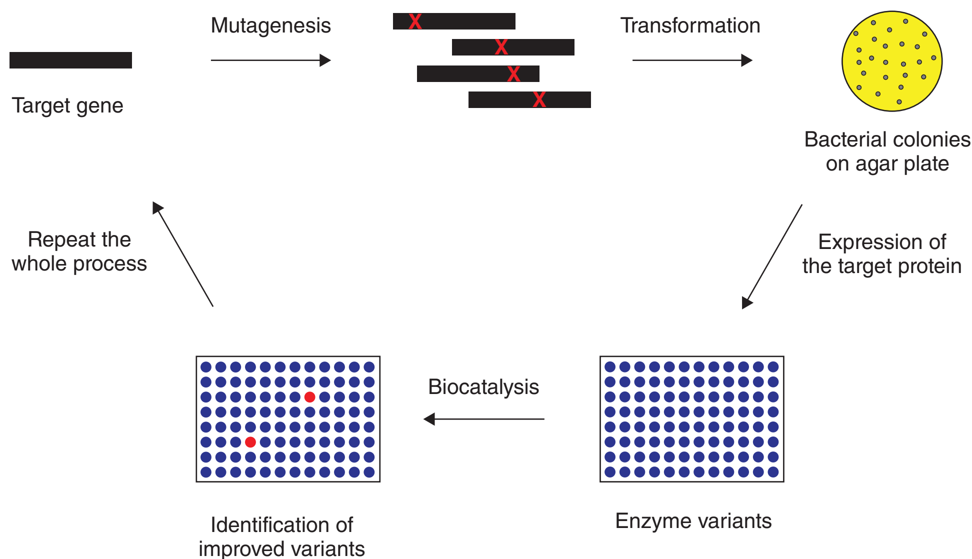
Enzymes have been used as catalysts in organic chemistry for more than a century [1a], but the general use of biocatalysis in academia and, particularly, in industry has suffered from the following often encountered limitations [1b–d]:

- Limited substrate scope
- Insufficient activity
- Insufficient or wrong stereoselectivity
- Insufficient or wrong regioselectivity
- Insufficient robustness under operating conditions.

Sometimes, product inhibition also limits the use of enzymes. All of these problems can be addressed and generally solved by applying directed evolution (or laboratory evolution as it is sometimes called) [2]. It mimics Darwinian evolution as it occurs in Nature, but it does not constitute real natural evolution. The process consists of several steps, beginning with mutagenesis of the gene encoding the enzyme of interest. The library of mutated genes is then inserted into a bacterial or yeast host such as *Escherichia coli* or *Pichia pastoris*, respectively, which is plated out on agar plates. After a growth period, single colonies appear, each originating from a single cell, which now begin to express the respective protein variants. Multiple copies of transformants as well as wild-type (WT) appear, which unfortunately decrease the quality of libraries and increase the screening effort. Colony harvesting must be performed carefully, because cross-contamination leads to the formation of inseparable mixtures of mutants with concomitant misinterpretations. The colonies are picked by a robotic colony picker (or manually using toothpicks), and placed individually in the wells of 96- or 384-format microtiter plates that contain nutrient broth. Portions of each well-content are then placed in the respective wells of another microtiter plate where the screening for a given catalytic property ensues. In some (fortunate) cases, an improved variant (hit) is identified in such an initial library, which fulfills all the requirements for practical application as defined by the experimenter. If this does not happen, which generally proves to be the

COPYRIGHTED MATERIAL

Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology, First Edition.
Manfred T. Reetz.
© 2017 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2017 by Wiley-VCH Verlag GmbH & Co. KGaA.



Scheme 1.1 The basic steps in directed evolution of enzymes. The rectangles represent 96 well microtiter plates that contain enzyme variants, the red dots symbolizing hits.

case, then the gene of the best variant is extracted and used as a template in the next cycle of mutagenesis/expression/screening (Scheme 1.1). This mimics "evolutionary pressure," which is the heart of directed evolution.

In most directed evolution studies further cycles are necessary for obtaining the optimal catalyst, each time relying on the Darwinian character of the overall process. A crucial feature necessary for successful directed evolution is the linkage between phenotype and genotype. If a library in a recursive mode fails to harbor an improved mutant/variant, the Darwinian process ends abruptly in a local minimum on the fitness landscape. Fortunately, researchers have developed ways to escape from such local minima ("dead ends") (see Section 4.3).

Directed evolution is thus an alternative to so-called "rational design" in which the researcher utilizes structural, mechanistic, and sequence information, possibly flanked by computational aids, in order to perform site-directed mutagenesis at a given position in a protein [3]. The molecular biological technique of site-specific mutagenesis with exchange of an amino acid at a specific position in a protein by one of the other 19 canonical amino acids was established by Michael Smith in the late 1970s [4a] which led to the Nobel Prize [4b]. The method is based on designed synthetic oligonucleotides and has been used extensively by Fersht [4c] as well as numerous other researchers in the study of enzyme mechanisms [4b]. This approach to protein engineering has also been fairly successful in thermostabilization experiments in which, for example, mutations leading to stabilizing disulfide bridges or intramolecular H-bridges are introduced "rationally" [5]. Nevertheless, in a vast number of other cases, directed evolution of protein robustness constitutes the superior

strategy [6]. Moreover, when aiming for enhanced or reversed enantioselectivity, diastereoselectivity, and/or regioselectivity, rational design is much more difficult [3], in which case directed evolution is generally the preferred strategy [7]. In some cases, researchers engaging in rational design actually prepare a set of mutants, test such a "library" and even combine the designed mutations, a process that resembles "real" laboratory evolution, as shown by Bornscheuer and coworkers who generated 28 rationally designed variants of a lipase, one of them showing an improved catalytic profile [8]. Other examples are listed in Table 5.1 in Chapter 5. However, this technique has limitations, and standard directed evolution approaches are more general and most reliable.

Directed evolution of enzymes is not as straightforward as it may appear to be at this point. The challenge in putting the above principles into practice has to do with the vastness of protein sequence space. High structural diversity is easily designed in mutagenesis, but the experimenter is quickly confronted by the so-called "numbers problem" which in turn relates to the screening effort (bottleneck). When mutagenizing a given protein, the theoretical number of variants N is described by Eq. (1.1), which is based on the use of all 20 canonical amino acids as building blocks [2]:

$$N = 19^M X!/[(X - M)!M!] \tag{1.1}$$

where M denotes the total number of amino acid substitutions per enzyme molecule and X is the total number of residues (size of protein in terms of amino acids). For example, when considering an enzyme composed of 300 amino acids, 5700 different mutants are possible if one amino acid is exchanged randomly, 16 million if two substitutions occur simultaneously, and about 30 billion if three amino acids are substituted simultaneously [2].

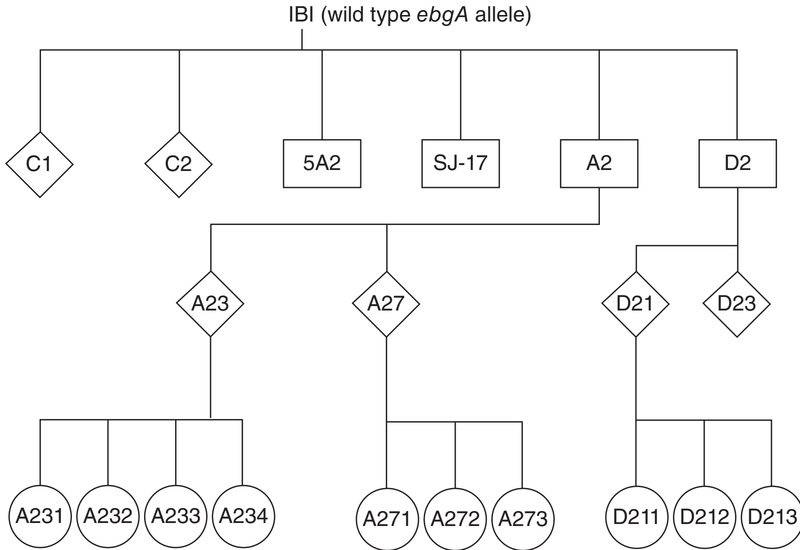
Such calculations pinpoint a dilemma that accompanies directed evolution to this day, namely how to probe the astronomically large protein sequence space efficiently. One strategy is to limit diversity to a point at which screening can be handled within a reasonable time, but excessive diversity reduction should be avoided because then the frequency of hits in a library diminishes and may tend toward zero in extreme cases. Finding the optimal compromise constitutes the primary issue of this monograph. A very different strategy is to develop selection systems rather than experimental platforms that require screening. In a selection system, the host organism thrives and survives because it expresses a variant having the catalytic characteristics that the researcher wants to evolve. A third approach is based on the use of various types of display systems, which are sometimes called "*selection systems*," although they are more related to screening. These issues are delineated in Chapter 2, which serves as a guide for choosing the appropriate system. Since it is extremely difficult to develop genuine selection systems or display platforms for directed evolution of stereo- and regioselective enzymes, researchers had to devise medium- and high-throughput screening systems (Chapter 2).

1.2
Brief Account of the History of Directed Evolution

Scientists have strived for a long time to "reproduce" or mimic natural evolution in the laboratory. In 1965–1967 Spiegelman and coworkers performed a "Darwinian experiment with a self-duplicating nucleic acid molecule" (RNA) outside a living cell [9]. It was believed that this mimics an early precellular evolutionary event. Later investigations showed that Spiegelman's RNA molecules were not truly self-duplicating, but his contributions marked the beginning of a productive new area of research on RNA evolution as fueled by such researchers as Szostak, Joyce, and others [10]. At this point, it should be noted that directed evolution at RNA level is a very different field of research with totally different goals, focusing on selection of RNA aptamers, selection of catalytic RNA molecules, or evolution of RNA polymerase ribozyme and of ribozymes by continuous serial transfer [10]. The history of directed evolution in this particular area has been reviewed [10b, 11]. The term "directed evolution" in the area of protein engineering was used as early as 1972 by Francis and Hansche, describing an *in vivo* system involving an acid phosphatase in *Saccharomyces cerevisiae* [12]. In a population of 10^9 cells, spontaneous mutations in a defined environment were continuously monitored over 1000 generations for their influence on the efficiency and activity of the enzyme at pH6. A single mutational event (M1) induced a 30% increase in the efficiency of orthophosphate metabolism. The second mutational event (M2 in the region of the structural gene) led to an adaptive shift in the pH optimum and in the enhancement of phosphatase activity by 60%. Finally, the third event (M3) induced cell clumping with no effect on orthophosphate metabolism [12].

In the 1970s, further contributions likewise describing *in vivo* directed evolution processes appeared sporadically. The contribution of Hall using the classical microbiological technique of genetic complementation constitutes a prominent example [13]. In one of the earliest directed evolution projects, new functions for the *ebgA* (*ebg* = evolved ß-galactosidase) were explored (Scheme 1.2) [13b]. Growth on different carbohydrates as the energy source was the underlying evolutionary principle. WT *ebgAo* is an enzyme showing very little or no activity toward certain carbohydrates such as the natural sugar lactose. It was shown, inter alia, that for an *E. coli* strain with *lac2* deletion to obtain the ability to utilize lactobionate as the carbon source, a series of mutations must be introduced in a particular order in the *ebg* genes. It was also found experimentally, when growing cells on different carbon sources, that in some cases old enzyme functions either remain unaffected or are actually improved.

Two decades later, the technique was extended by Kim and coworkers [14a]. It may have inspired other groups to study and develop new evolution experiments, for example, by Lenski and coworkers who investigated parallel changes in gene expression after 20 000 generations of evolution in bacteria [14b], and more recently by Liu and coworkers who implemented a novel technique for continuous evolution [14c] including a phage-assisted embodiment [14d].



Scheme 1.2 Pedigree of *ebgA* alleles in evolved strains [13b]. Strain 1B1 carries the wild type allele, *ebgAO*. Strains on line one have a single mutation in the *ebgA* gene; those in line two have two mutations in *ebgA*; those in line three have three mutations in *ebgA*. All strains are *ebgR*. Strains enclosed in rectangles were selected for growth on lactose; those enclosed in diamonds were selected for growth on lactulose; those in circles were selected for growth on lactobionate. This pedigree shows only the descent of the *ebgA* gene; that is, strains SJ-17, A2, 5A2, and D2 were not derived directly from IBI, but their *ebgA* alleles were derived directly from the *ebgA* allele carried in IBI. (Hall [13b]. Reproduced with permission of Genetic Society of America.)

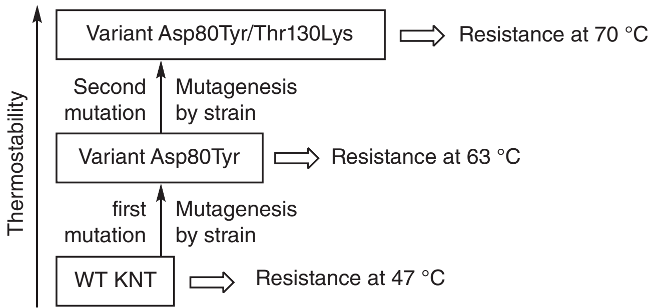
Although originally not specifically related to directed evolution, developments such as the Kunkel method of mutational specificity based on depurination [15] deserves mention because it was used two decades later in mutant library design based on error-prone rolling circle amplification (epRCA) [16]. These and many other early developments inspired scientists to speculate about the potential applications of directed evolution in biotechnology. In 1984, Eigen and Gardiner formulated these intriguing perspectives by emphasizing the necessity of self-replication in molecular *in vitro* evolution [17]. At that time the best self-replication system for the laboratory utilized the replication of single-stranded RNA by the replication enzyme of the coliphage Qf3. The logic of laboratory Darwinian evolution involving recursive cycles of gene mutagenesis, amplification, and selection was formulated schematically (Scheme 1.3), although the generation of bacterial colonies on agar plates for ensuring the genotype–phenotype relation (Scheme 1.1) as employed later by essentially all directed evolution researchers was not considered. It should be stated that in the early 1980s the polymerase chain reaction (PCR) for high-fidelity DNA amplification had not yet been developed. Following its announcement in the 1980s by Mullis [18], completely new perspectives emerged for many fields, including directed evolution.

10 START WITH SELECTED GENOTYPE
20 LET IT REPRODUCE, MUTATING OCCASIONALLY
30 FORCE DIFFERENT GENOTYPES TO COMPETE
40 NATURAL SELECTION OF QUASI-SPECIES AROUND BEST-ADAPTED GENOTYPE OCCURS
50 WHEN ADVANTAGEOUS MUTANT APPEARS – GO TO 10

Scheme 1.3 Logic of Darwinian evolution in the laboratory according to Eigen and Gardiner [17]. (Adapted from Eigen and Gardiner [17]. Reproduced with permission of De Gruyter.)

Parallel to these developments, researchers began to experiment with different types of mutagenesis methods in order to generate mutant libraries, which were subsequently screened or selected for an enzyme property, generally protein thermostability. Sometimes mutagenesis methods were introduced without any real applications at the time of publication. These and other early contributions, as summarized in a 1997 review article [19], paved the way to modern directed evolution [2]. Only a few early representative developments are highlighted here. In 1985, Matsumura and Aiba subjected kanamycin nucleotidyltransferase (cloned into a single-stranded bacteriophage M13) to hydroxylamine-induced chemical mutagenesis [20]. Following recloning of the mutagenized gene of the enzyme into the vector plasmid pTB922, the recombinant plasmid was employed to transform *Bacillus stearothermophilus* so that more stable variants could be identified by screening. About 12 out of 8000 transformants were suspected to harbor thermostabilized variants, the best one being characterized by a single point mutation and a stabilization of 6 °C. A number of other early papers concerning the robustness of T4 lysozyme by chemically induced random mutagenesis likewise contributed to directed evolution of protein thermostabilization, as summarized by Matthews and coworkers in a 2010 review article [21].

Today, many protein engineers maintain that the discovery of improved enzymes in an initial mutant library does not (yet) constitute an evolutionary process, and that at least one additional cycle of mutagenesis/expression/screening as shown in Scheme 1.1 is required before the term “directed evolution” applies [2]. The first example of two mutagenesis cycles was reported by Hageman and coworkers in 1986 in their efforts to enhance the thermostability of kanamycin nucleotidyltransferase by an evolutionary process based on a mutator strain [22]. Basically, this seminal study consisted of cloning the gene that encodes the enzyme from a mesophilic organism, introducing the gene into an appropriate thermophile and selecting for activity at the higher growth temperatures of the host organism (in this case *B. stearothermophilus*). The host organism is resistant to the antibiotic at 47 °C, but not at temperatures above 55 °C. Upon passing a shuttle plasmid through the *E. coli mutD5* mutator strain and introduction into *B. stearothermophilus,* a point mutation that led to resistance to kanamycin at 63 °C was identified, namely Asp80Tyr. Using this as a template, the second round was performed under higher selection pressure at 70 °C, leading to the accumulation of mutation Thr130Lys, the respective double mutant Asp80Tyr/Thr130Lys



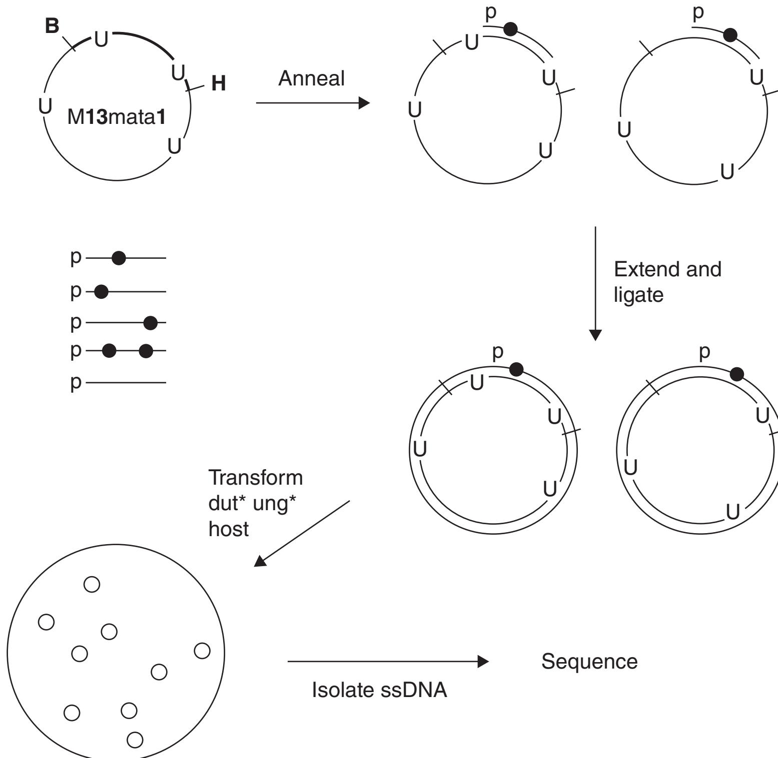
Scheme 1.4 Early example of directed evolution of thermostability with kanamycin nucleotidyltransferase (KNT) serving as the enzyme and a mutator strain as the random mutagenesis technique in an iterative manner [22].

showing even higher thermostability (Scheme 1.4) [22]. The Darwinian character of this approach to thermostabilization of proteins is self-evident.

The original site-specific mutagenesis established by Smith allows the specific exchange of any amino acid in a protein by any one of the other 19 canonical amino acids [4], but the generation of random mutations at a single residue or defined multi-residue randomization site was not developed until later. Early on, several variations of cassette mutagenesis based on the use of “doped” synthetic oligodeoxynucleotides were developed, allowing the combinatorial introduction of all of the 19 other canonical amino acids at a given position [23]. These and similar studies were performed for different reasons, not all having to do with enzyme catalysis. The study by Wells and coworkers is highlighted here, because it constitutes a clever combination of rational design and directed evolution for the purpose of increasing the robustness of the serine protease subtilisin (enhanced resistance to chemical oxidation) [24]. Focused random mutagenesis was induced by cassette mutagenesis (see Section 3.3 for the details of this and other saturation mutagenesis methods). At the time it was known that residue Met222 constitutes a site at which undesired oxidation occurs. Therefore, saturation mutagenesis was performed at this position, which led to several improved variants showing resistance to 1 M H_2O_2 as measured by the reaction of *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide, including mutants Met222Ser, Met222Ala, and Met222Leu [24].

As pointed out by Ner *et al.* in 1988, a disadvantage of cassette mutagenesis as originally developed is the fact that the synthetic oligodeoxynucleotides in form of a cassette have to be introduced between two restriction sites, one on either side of the to be randomized sequence [25]. Since the restriction sites had to be generated by standard oligodeoxynucleotide mutagenesis, additional steps were necessary prior to the actual randomization procedure. Therefore, an improved version was developed using a combination of the known primer extension procedure [26] and Kunkel’s method of strand selection [27]. The technique uses a mixed pool of oligodeoxynucleotides prepared by contaminating the monomeric nucleotides with low levels of the other three nucleotides so that the full-length oligonucleotide contains on average one to two changes/molecules.

It was employed in priming *in vitro* synthesis of the complementary strand of cloned DNA fragments in M13 or pEMBL vectors, the latter having been passed through the *E. coli* host. The method allows random point mutations as well as codon replacements. Scheme 1.5 illustrates the case of the *MAT*a1 gene from *S. cerevisiae* [25].



Scheme 1.5 Mixed oligonucleotide mutagenesis of the gene MATa1 from *Saccharomyces cerevisiae* [25]. (Ner *et al.* [25]. Reproduced with permission of Mary Ann Liebert, Inc.)

Further variations and improvements appeared in the late 1980s. These include the generation of mutant libraries using spiked oligodeoxyribonucleotide primers according to Hermes *et al.* [28]. The use of overlap extension polymerase chain reaction (OE-PCR) for site-specific mutagenesis constitutes a seminal contribution by Pease and coworkers at the Mayo Clinic, which has influenced directed evolution because it can be employed in saturation mutagenesis [29]. OE-PCR can also be used for insertion and deletion mutations [30].

In yet another contribution appearing in the 1980s, Dube and Loeb generated ß-lactamase mutants that render *E. coli* resistant to the antibiotic carbenicillin by replacing the DNA sequence corresponding to the active site with random nucleotide sequences without exchanging the codon encoding catalytically active

Ser70 [31]. The inserted oligonucleotide $Phe^{66}XXXSer^{70}XXLys^{73}$ contains 15 base pairs of chemically synthesized random sequences that code for 2.5 million amino acid exchanges. It should be noted that ß-lactamase is an ideal enzyme with which randomization-based protein engineering can be performed because a simple and efficient selection system is available (see Chapter 2).

Further variations and improvements of site-specific mutagenesis appeared in the 1990s (see Chapter 3 for details), which were extended to allow randomization at more than one residue site. Based on some of these developments, the so-called QuikChange™ protocol for saturation mutagenesis emerged in 2002 [32], which is described in detail in Section 3.3. Another important version of saturation mutagenesis is the "megaprimer" method of site-specific mutagenesis introduced by Kammann *et al.* [33] and improved by Sarkar and Sommer in 1990 [34]. The overall procedure is fairly straightforward and easy to perform, but it also has limitations as discussed in Section 3.3. These and other early developments of site-directed mutagenesis, which can also be used for randomization, were summarized by Reikofski and Tao in 1992 [35].

In 1989, a landmark study was published by Leung *et al.* describing error-prone polymerase chain reaction (epPCR) [36a], but it was not applied to enzymes until a few years later (see following text). It relies on *Taq* polymerase or similar DNA polymerases that lack proofreading ability (no removal of mismatched bases). In order to control the mutational rate, the reaction conditions need to be optimized by varying such parameters as the $MgCl_2$ or $MnCl_2$ concentrations and/or employing unbalanced nucleotide concentrations (see details in Section 3.3) [36b].

The first applications of epPCR are due to Hawkins *et al.* in 1992 [37], who reported *in vitro* selection and affinity maturation of antibodies from combinatorial libraries. The creation of large combinatorial libraries of antibodies was a new area of science at the time, as shown earlier by Lerner and coworkers using different techniques [38]. It should be noted that epPCR suffers from various limitations [39] that are discussed in Section 3.2. To this day, the technique continues to be employed, especially when X-ray structural data of the protein is not available. A different but seldom used molecular biological random mutagenesis method was developed and applied in 1992/1993 by Zhang *et al.* in order to increase the thermostability of aspartase as a catalyst in the industrially important addition reaction of ammonia to fumarate with formation of L-aspartic acid [40]. Unbalanced nucleotide amounts were used in a special way, but from today's perspective it is clear that diversity is lower than in the case of epPCR [40b].

In 1993, Chen and Arnold published a key paper describing the use of random mutagenesis in the quest to increase the robustness of the protease subtilisin E in aqueous medium containing a hostile organic solvent (dimethylformamide, DMF) [41]. First, the mutations of three variants obtained earlier by rational design were combined with formation of the respective triple mutant Asp60Asn/Gln103Arg/Asn218Ser to which was added a fourth point mutation Asp97Gly, leading to variant Asp60Asn/Gln103Arg/Asn218Ser/Asp97Gly ("4M variant"). The *Hin*dIII/*Bam*HI DNA fragment of 4M subtilisin E from

residue 49 to the C-terminus was then employed as the template for PCR-based random mutagenesis. Thus, this diverges a little from epPCR as originally developed by Leung *et al.* [36a] which addresses the whole gene. The PCR conditions were modified so that the mutational frequency increased (including the use of $MnCl_2$). An easy to perform prescreen for activity was developed using agar plates containing 1% casein, which upon hydrolysis forms a halo. The roughly identified active mutants were then sequenced and used as catalysts in the hydrolysis of *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Met-*p*-nitroanilide and *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide. Upon going through three cycles of random mutagenesis, the final best hit PC3 was identified as having a total of 10 point mutations. The catalytic efficiency of variant PC3 relative to WT subtilisin E in aqueous medium containing different amounts of DMF is shown in Figure 1.1 [41].

Upon generating 10 single mutants corresponding to the 10 point mutations that accumulated successively, it was discovered that they are not additive. All of the point mutations that influence activity in the presence of DMF were found to be on the surface of the enzyme, and none were found in the conserved α-helix and ß-sheet structures. Rather, they are located in the loops that interconnect the core secondary structures [41]. Another significant aspect of this work is the fact that not just initial mutant libraries were created as in most other studies of the 1980s, but that the protocol constitutes another example of more than one cycle of mutagenesis, expression, and screening as demonstrated earlier by Hageman and coworkers (Scheme 1.4) [22]. The use of recursive cycles clearly underscores the Darwinian nature of this procedure.

In 1996, the Arnold group applied conventional epPCR [36] in a study directed toward increasing the robustness and activity of subtilisin E in 30% aqueous DMF

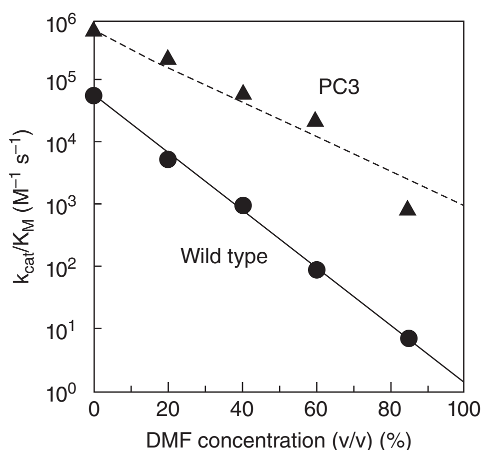
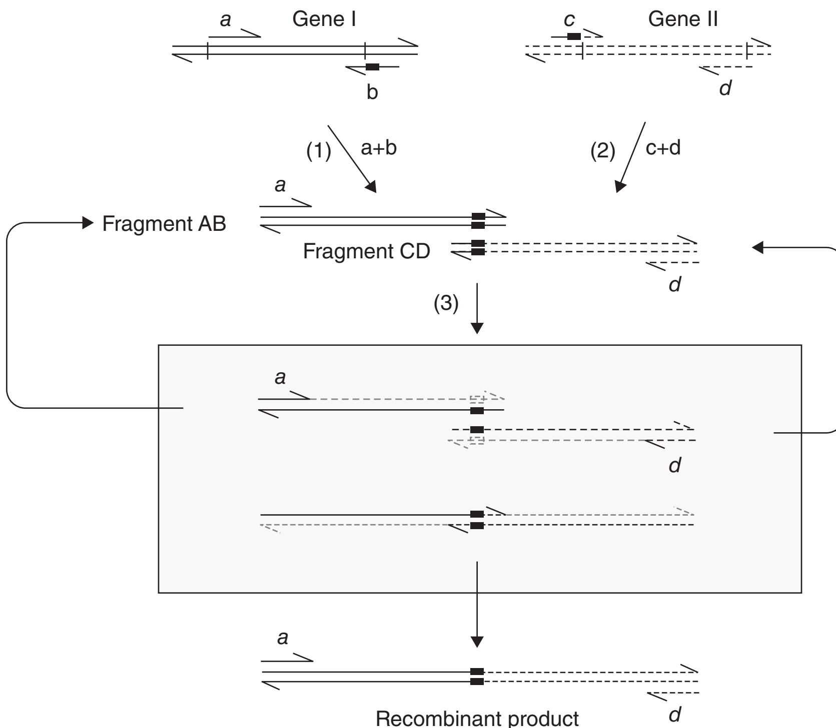


Figure 1.1 Catalytic efficiency of WT subtilisin E and variant PC3 as catalysts in the hydrolytic cleavage of *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Met-*p*-nitroanilide [41]. (Adapted from Chen and Arnold [41]. Reproduced with permission of National Academy of Sciences.)

as a catalyst in the hydrolysis of *p*-nitrophenyl esters [42]. Four cycles of epPCR were transversed, *p*-nitrophenylacetate serving as the model substrate that forms acetic acid and *p*-nitrophenol. The latter has a yellow color and can then be used conveniently in the UV/vis-based screening system, a well-known assay used in biochemistry for decades. The improved mutants were then tested successfully as robust catalysts in the hydrolysis of *p*-nitrobenzyl esters in 30% aqueous DM [42].

New methods promising practical applications were developed in the 1980s, a key study by Horton *et al.* being a prime example [43]. It is an extension of their earlier work on OE-PCR [29]. Fragments from two genes that are to be recombined are first produced by separate PCR, the primers being designed so that the ends of the products feature complementary sequences (Scheme 1.6). Upon mixing, denaturing, and reannealing the PCR products, those strands that have matching sequences at their 3′ ends overlap and function as primers for each other. Extension of the overlap by a DNA polymerase leads to products in which the original sequences are spliced together. This recombinant technique for producing chimeric genes was called splicing by overlap extension (SOE), which also allows the introduction of random errors (mutations). The technique was

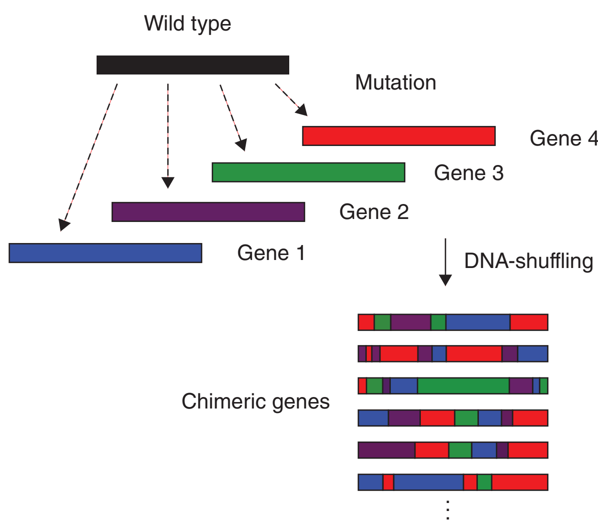


Scheme 1.6 Steps in the recombinant technique of splicing by overlap extension (SOE), illustrated here using two different genes [43]. (Adapted from Horton *et al.* [43]. Reproduced with permission of Elsevier.)

illustrated using two different mouse class-I major histo-compatible genes. However, at the time it was not exploited by the biotechnology community active in directed evolution [43].

The recombinant process of SOE can be considered to be a forerunner of DNA shuffling, an efficient and general recombinant technique introduced by Stemmer in 1994 [44]. Another forerunner of DNA shuffling was developed by Brown, who coined the term "*oligonucleotide shuffling*" in 1992 when evolving mutants of the *E. coli* phage receptor that displayed enhanced adhesion to iron oxide [45]. Libraries of randomized oligonucleotides were shuffled in a process reminiscent of exon shuffling [46].

DNA shuffling goes far beyond these forerunners. It is a process that simulates sexual evolution as it occurs in Nature. In the original study, ß-lactamase served as the enzyme, the selection system being based on the increased resistance to an antibiotic. DNA shuffling is illustrated here when starting with mutants of a given enzyme (Scheme 1.7). Family shuffling, introduced in 1998 Winter, is a variation which in many cases constitutes the superior approach [47] (see Section 3.4 for a description of this technique and other recombinant methods).



Scheme 1.7 DNA shuffling starting from a single gene encoding a given enzyme.

These seminal papers sparked a great deal of further research in the area of directed evolution in the 1990s. In many of the studies, recombinant and/or non-recombinant methods were applied in order to shed light on the mechanism of enzymes, but usually only initial mutant libraries were considered. To this day, directed evolution is often employed in the quest to study enzyme mechanisms rather than for the purpose of evolving altered enzymes for practical purposes. Contributions by Benkovic and coworkers [48] are prominent examples, as are the

studies by Hecht and coworkers concerning binary patterning [49]. In an informative overview by Lutz and Benkovic that appeared in 2002, many of these and other early developments in directed evolution were assessed [50]. For example, the invention of phage display by Smith in 1985 [51], although originally not intended for protein engineering, was employed by Winter *et al.* [52] and Benkovic and coworkers [53] for antibody selection, and by several groups for evolving catalytic profiles, including Fastrez and coworkers [54], Lerner and coworkers [55], Winter *et al.* [56], and Schultz and coworkers [57].

Phage display inspired the development of several other early display platforms such as ribosomal display by Szostak and coworkers [58] and yeast display in the same year by Boder and Wittrup [59], which set the stage for many exciting developments in directed evolution. Although flow cytometry had been developed at an early stage, it was not combined with fluorescence-activated cell sorter (FACS) technology for application in directed evolution until much later, as demonstrated by the early pioneering contributions of Georgiou and coworkers [60]. The water-in-oil emulsion technology, elegantly developed by Griffiths and Tawfik [61], likewise deserves mention. All of these selection platforms, which are really screening techniques [62], are useful in a number of protein engineering applications, but to this day their utilization in the laboratory evolution of stereo- and/or regioselective enzymes remains marginal (see Chapter 2).

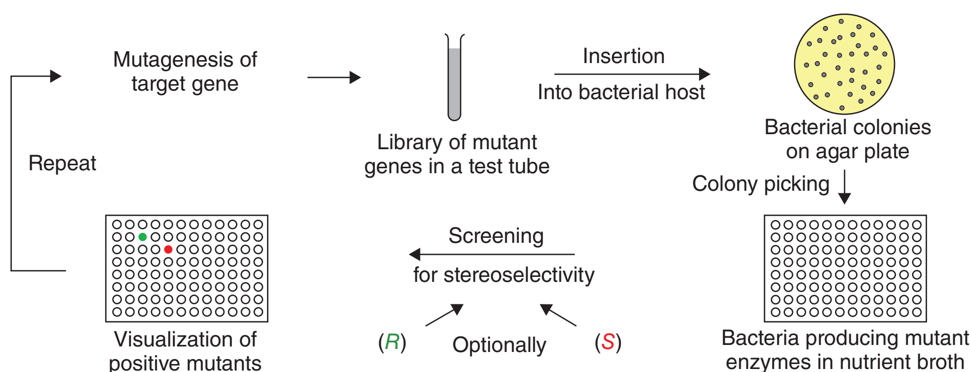
The distinction between selection and screening [63a] was recognized by Hilvert and coworkers in the 1990s, who consequently developed impressive selection systems in which the host organism experiences a growth advantage due to the generation of enzyme mutants displaying desired properties [63b]. Applying this to stereo- and/or regioselectivity remains a challenge [62], as delineated in Chapter 2.

The generation of selective catalytic monoclonal antibodies can be considered to be based on evolutionary principles, but despite impressive contributions [64], these biocatalysts have not entered a stage of practical applications in stereoselective organic chemistry or biotechnology. This appears to be because the immune system functions on the basis of binding, and not on catalytic turnover [64c].

In directed evolution of enzymes as catalysts in organic chemistry and biotechnology, an important early contribution by Patrick and Firth describing algorithms for designing mutant libraries based on statistical analyses has influenced the field to this day [65]. Ostermeier developed a similar metric [66], and Pelletier has extended these statistical models [67]. Later, these contributions led to further developments, for example, the incorporation of the Patrick/Firth algorithm in two other computer aids, CASTER for user-friendly design of saturation mutagenesis libraries for activity, stereo- and regioselectivity, and B-FITTER for designing libraries of mutants displaying improved thermostability [68], both available free of charge on the author's homepage (http://www.kofo.mpg.de/en/research/biocatalysis) [68], (see Section 3.3 for details).

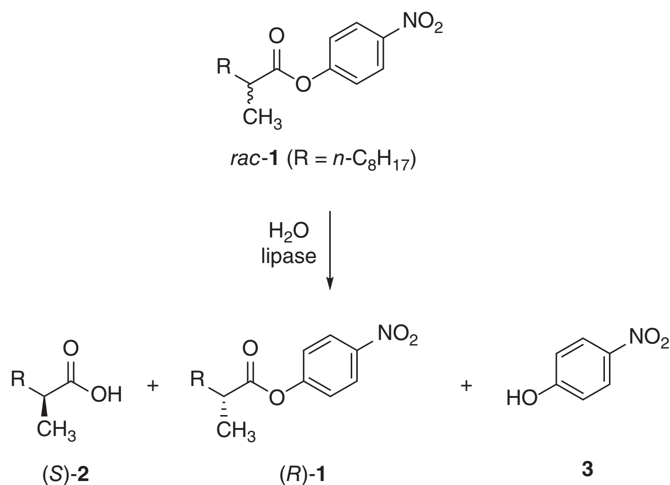
While the creation of enhanced enzyme thermostability paved the way for potential applications in biotechnology, realizing the potentially broad utility of directed evolution as a prolific source of selective catalysts in synthetic organic chemistry was still to come. In the mid-1990s the Reetz group became

interested in protein engineering because they wanted to develop a new approach to asymmetric catalysis: the directed evolution of stereoselective enzymes as catalysts in organic chemistry and biotechnology [69a]. As organic chemists we speculated that directed evolution could possibly be harnessed to enhance and perhaps even to invert enantioselectivity of enzymes (Scheme 1.8). Consequently, some of the traditional limitations of biocatalysis (Section 1.1) would be eliminated, thereby establishing a prolific and unceasing source of stereoselective biocatalysts for the major enzyme types including hydrolases (e.g., lipases, esterases, epoxide hydrolases), oxidases (e.g., P450-monooxygenases, Baeyer–Villiger monooxygenases), reductases (e.g., alcohol dehydrogenases, enoate-reductases), lyases (addition/elimination), isomerases (e.g., epimerization), and ligases (e.g., aldolases, oxynitrilases, benzoylformate decarboxylases). The underlying idea is very different from the traditional development of chiral synthetic transition metal catalysts or organocatalysts, because the stepwise increase in stereoselectivity can be expected to emerge as a consequence of the evolutionary pressure exerted in each cycle. Since stereoselectivity stands at the heart of modern synthetic organic chemistry, we reasoned that this complementary approach would enrich the toolbox of organic chemists (for a personal account of our entry into directed evolution, see [70]).

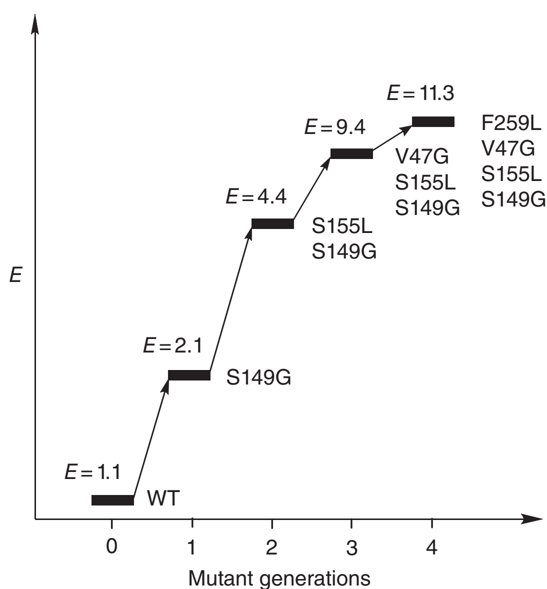


Scheme 1.8 Concept of directed evolution of stereoselective enzymes with (*R*)- or (*S*)-selective mutants being accessible on an optional basis [69]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

In a proof-of-principle study, the lipase from *Pseudomonas aeruginosa* (PAL) was used as the enzyme in the hydrolytic kinetic resolution of ester **1** (Scheme 1.9) [69a]. WT PAL is a poor catalyst in this reaction because the selectivity factor measuring the relative rate of reaction of (*R*)- and (*S*)-**1** amounts to only $E = 1.1$ with slight preference for (*R*)-**2**. Four cycles of epPCR at low mutation rate led to variant A showing notably enhanced enantioselectivity ($E = 11$). It is characterized by four point mutations S149G/S155L/V476/F259L, which accumulated in a step-wise manner (Scheme 1.10) [69]. Since even medium-throughput ee-assays were not available at the time and the first truly high-throughput ee-screening



Scheme 1.9 Hydrolytic kinetic resolution of *rac*-**1** catalyzed by the lipase from *Pseudomonas aeruginosa* (PAL) [69a]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)



Scheme 1.10 First example of directed evolution of a stereoselective enzyme [69a]. The model reaction involves the hydrolytic kinetic resolution of *rac*-**1** catalyzed by the lipase PAL, four rounds of epPCR being used as the gene mutagenesis method. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

system was not developed until 1999 [71], an on-plate pretest as well as a UV/vis-based screening system for identifying enantioselective lipase mutants (300–600 transformants/day) had to be developed first [69a] (see Chapter 2). Although a selectivity factor of $E = 11$ does not suffice for practical applications, this study set the stage for the rapid development of directed evolution of stereoselective enzymes in which we and many other groups participated (see Chapter 5). Progress up to 2004 covering several different enzyme types was summarized in two reviews [72]. At that time improved directed evolution strategies for the PAL-catalyzed asymmetric transformation of *rac*-**1** led to notable enhancement of the selectivity factor ($E = 51$), but it was also clear that further methodology development was necessary in order to promote genuine advances in the field of directed evolution (see Chapters 3–5).

1.3 Applications of Directed Evolution of Enzymes

Following the early groundbreaking studies of directed evolution (Section 1.2), this type of protein engineering has rapidly emerged as a major research area worldwide. Hundreds of studies appear each year describing the evolution of proteins featuring altered properties. In addition to the extensive area of evolved enzymes as catalysts in synthetic organic and pharmaceutical chemistry as well as biotechnology, applications extend into an array of very different areas, including:

- Metabolic pathway engineering [73]
- Engineered CRISPR-Cas9 nucleases [74]
- Vaccine production [75a–c]
- Potential universal blood generation [75d]
- Engineered antibodies [76]
- Genetic modification of plants for agricultural and medicinal purposes [77]
- Genetically modified yeasts in food industry [78]
- Photosynthetic CO_2 fixation [79]
- Engineered proteins in pollution control [80]
- Engineered enzymes in evolutionary biology for studying natural evolution [81]
- Engineered DNA polymerases for accepting synthetic nucleotides [82].

This monograph features primarily the laboratory evolution of enzymes as catalysts in synthetic organic chemistry and biotechnology, the focus being on the most important developments during recent years. Rather than being comprehensive, general principles, practical guidelines, and limitations are delineated. In this spirit, mutagenesis techniques and screening systems are described, followed by the analysis of selected case studies. Where possible, different approaches and strategies of directed evolution are critically compared.

The complementarity of enzymes and man-made synthetic transition metal catalysts and organocatalysts is emphasized where appropriate, as in recent perspectives on biocatalysis [1d, 7d]. With the establishment of directed evolution [2],

enzyme-based retrosynthetic analyses and, therefore, complex biocatalysis-based synthesis planning as put forth by Turner and O'Reilly [83] also constitute complementary strategies in synthetic organic chemistry. These developments include one-pot enzymatic cascade reactions, optionally in combination with man-made transition metal catalysts, processes that can be implemented with WT and/or evolved enzymes [84].

References

1. (a) Rosenthaler, L. (1908) Durch Enzyme bewirkte asymmetrische Synthesen. *Biochem. Z.*, **14**, 238–253; (b) Drauz, K., Gröger, H., and May, O. (eds) (2012) *Enzyme Catalysis in Organic Synthesis*, 3rd edn, Wiley-VCH Verlag GmbH, Weinheim; (c) Faber, K. (2011) *Biotransformations in Organic Chemistry*, 6th edn, Springer, Heidelberg; (d) Reetz, M.T. (2013) Biocatalysis in organic chemistry and biotechnology: past, present and future. *J. Am. Chem. Soc.*, **135**, 12480–12496; (e) Liese, A., Seeelbach, K., and Wandrey, C. (2006) *Industrial Biotransformations*, 2nd edn, Wiley-VCH Verlag GmbH, Weinheim.
2. Recent reviews of directed evolution of enzymes: (a) Bommarius, A.S. (2015) Biocatalysis, a status report. *Annu. Rev. Chem. Biomol. Eng.*, **6**, 319–345; (b) Denard, C.A., Ren, H., and Zhao, H. (2015) Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.*, **25**, 55–64; (c) Currin, A., Swainston, N., Day, P.J., and Kell, D.B. (2015) Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.*, **44**, 1172–1239; (d) Gillam, E.M.J., Copp, J.N., and Ackerley, D.F. (eds) (2014) Directed evolution library creation, in *Methods in Molecular Biology*, Humana Press, Totowa, NJ; (e) Widersten, M. (2014) Protein engineering for development of new hydrolytic biocatalysts. *Curr. Opin. Chem. Biol.*, **21**, 42–47; (f) Reetz, M.T. (2012) in *Enzyme Catalysis in Organic Synthesis*, 3rd edn (eds K. Drauz, H. Gröger, and O. May), Wiley-VCH Verlag GmbH, Weinheim, pp. 119–190; (g) Goldsmith, M. and Tawfik, D.S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.*, **22** (4), 406–412; (h) Bommarius, A.S., Blum, J.K., and Abrahamson, M.J. (2011) Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.*, **15** (2), 194–200; (i) Brustad, E.M. and Arnold, F.H. (2011) Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.*, **15** (2), 201–210; (j) Jäckel, C. and Hilvert, D. (2010) Biocatalysts by evolution. *Curr. Opin. Biotechnol.*, **21** (6), 753–759; (k) Lutz, S. and Bornscheuer, U.T. (eds) (2009) *Protein Engineering Handbook*, Wiley-VCH Verlag GmbH, Weinheim.
3. (a) Chica, R.A., Doucet, N., and Pelletier, J.N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.*, **16** (4), 378–384; (b) Ema, T., Nakano, Y., Yoshida, D., Kamata, S., and Sakai, T. (2012) Redesign of enzyme for improving catalytic activity and enantioselectivity toward poor substrates: manipulation of the transition state. *Org. Biomol. Chem.*, **10** (31), 6299–6308; (c) Pleiss, J. (2012) in *Enzyme Catalysis in Organic Synthesis*, 3rd edn (eds K. Drauz, H. Gröger, and O. May), Wiley-VCH Verlag GmbH, Weinheim, pp. 89–117; (d) Ma, B.-D., Kong, X.-D., Yu, H.-L., Zhang, Z.-J., Dou, S., Xu, Y.-P., Ni, Y., and Xu, J.-H. (2014) Increased catalyst productivity in α-hydroxy acids resolution by esterase mutation and substrate modification. *ACS Catal.*, **4** (3), 1026–1031; (e) Steiner, K. and Schwab, H. (2012) Recent advances in

rational approaches for enzyme engineering. *Comput. Struct. Biotechnol. J.*, **2**, e201209010.

4. (a) Smith, M. (1985) In vitro mutagenesis. *Annu. Rev. Genet.*, **19**, 423–462; (b) Smith, M. (1994) Synthetic DNA and biology (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.*, **33** (12), 1214–1221; (c) Fersht, A. (1999) *Structure and Mechanism in Protein Science*, 3rd edn, W. H. Freeman and Company, New York.
5. Reviews of rational design of protein thermostabilization: (a) Oshima, T. (1994) Stabilization of proteins by evolutionary molecular engineering techniques. *Curr. Opin. Struct. Biol.*, **4** (4), 623–628; (b) Ó'Fágáin, C. (2003) Enzyme stabilization—recent experimental progress. *Enzyme Microb. Technol.*, **33** (2-3), 137–149; (c) Eijsink, V.G.H., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B., and Vriend, G. (2004) Rational engineering of enzyme stability. *J. Biotechnol.*, **113** (1-3), 105–120; (d) Renugopalakrishnan, V., Garduno-Juarez, R., Narasimhan, G., Verma, C.S., Wei, X., and Li, P.Z. (2005) Rational design of thermally stable proteins: relevance to bionanotechnology. *J. Nanosci. Nanotechnol.*, **5** (11), 1759–1767; (e) Crespo, M.D. and Rubini, M. (2011) Rational design of protein stability: effect of (2*S*,4*R*)-4-fluoroproline on the stability and folding pathway of ubiquitin. *PLoS One*, **6** (5), e19425; (f) Tadokoro, T., Kazama, H., Koga, Y., Takano, K., and Kanaya, S. (2013) Investigating the structural dependence of protein stabilization by amino acid substitution. *Biochemistry*, **52** (16), 2839–2847.
6. Reviews of directed evolution of protein thermostabilization: (a) Arnold, F.H. (1998) Design by directed evolution. *Acc. Chem. Res.*, **31** (3), 125–131; (b) Eijsink, V.G.H., Gaseidnes, S., Borchert, T.V., and van den Burg, B. (2005) Directed evolution of enzyme stability. *Biomol. Eng*, **22** (1-3), 21–30; (c) Bommarius, A.S. and Broering, J.M. (2005) Established and novel tools to investigate biocatalyst stability. *Biocatal. Biotransform.*, **23** (3-4), 125–139.
7. Reviews of directed evolution of stereoselectivity [2f]: (a) Reetz, M.T. (2011) Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.*, **50** (1), 138–174; (b) Reetz, M.T., Wu, S., Zheng, H.B., and Prasad, S. (2010) Directed evolution of enantioselective enzymes: an unceasing catalyst source for organic chemistry. *Pure Appl. Chem.*, **82** (8), 1575–1584; (c) Reetz, M.T. (2010) in *Manual of Industrial Microbiology and Biotechnology*, 3rd edn (eds R.H. Baltz, A.L. Demain, J.E. Davies, A.T. Bull, B. Junker, L. Katz, L.R. Lynd, P. Masurekar, C.D. Reeves, and H. Zhao), ASM Press, Washington, DC, pp. 466–479; (d) Sun, Z., Wikmark, Y., Bäckvall, J.-E., and Reetz, M.T. (2016) New concepts for increasing the efficiency in directed evolution of stereoselective enzymes. *Chem. Eur. J.* **22**, 5046–5054.
8. Müller, J., Sowa, M.A., Fredrich, B., Brundiek, H., and Bornscheuer, U.T. (2015) Enhancing the acyltransferase activity of *Candida antarctica* lipase A by rational design. *ChemBioChem*, **16** (12), 1791–1796.
9. (a) Mills, D.R., Peterson, R.L., and Spiegelman, S. (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.*, **58** (1), 217–224; (b) Spiegelman, S. (1971) An approach to the experimental analysis of precellular evolution. *Q. Rev. Biophys.*, **4** (2 and 3), 213–253.
10. (a) Adamala, K., Engelhart, A.E., and Szostak, J.W. (2015) Generation of functional RNAs from inactive oligonucleotide complexes by non-enzymatic primer extension. *J. Am. Chem. Soc.*, **137** (1), 483–489; (b) Joyce, G.F. (2007) Forty years of in vitro evolution. *Angew. Chem. Int. Ed.*, **46** (34), 6420–6436; (c) Blain, J.C. and Szostak, J.W. (2014) Progress toward synthetic cells. *Annu. Rev. Biochem.*, **83**, 615–640; (d) Sun, H. and Zu, Y. (2015) Aptamers and their applications in nanomedicine. *Small*, **11** (20), 2352–2364; (e) Mayer, G., Ahmed, M.S., Dolf, A., Endl, E., Knolle, P.A.,

and Famulok, M. (2010) Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. *Nat. Protoc.*, **5** (12), 1993–2004.
11. Kim, E.-S. (2008) Directed evolution: a historical exploration into an evolutionary experimental system of nanobiotechnology, 1965–2006. *Minerva*, **46**, 463–484.
12. Francis, J.C. and Hansche, P.E. (1972) Directed evolution of metabolic pathways in microbial populations. I. Modification of acid-phosphatase pH optimum in *S. Cerevisiae*. *Genetics*, **70** (1), 59–73.
13. (a) Hall, B.G. (1977) Number of mutations required to evolve a new lactase function in *Escherichia coli*. *J. Bacteriol.*, **129** (1), 540–543; (b) Hall, B.G. (1978) Experimental evolution of a new enzymatic function. II. Evolution of multiple functions for *EBG* enzyme in *E. Coli*. *Genetics*, **89** (3), 453–465; (c) Hall, B.G. (1981) Changes in the substrate specificities of an enzyme during directed evolution of new functions. *Biochemistry*, **20** (14), 4042–4049.
14. (a) Hwang, B.Y., Oh, J.M., Kim, J., and Kim, B.G. (2006) Pro-antibiotic substrates for the identification of enantioselective hydrolases. *Biotechnol. Lett*, **28** (15), 1181–1185; (b) Cooper, T.F., Rozen, D.E., and Lenski, R.E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **100** (3), 1072–1077; (c) Esvelt, K.M., Carlson, J.C., and Liu, D.R. (2011) A system for the continuous directed evolution of biomolecules. *Nature*, **472** (7344), 499–503; (d) Leconte, A.M., Dickinson, B.C., Yang, D.D., Chen, I.A., Allen, B., and Liu, D.R. (2013) A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry*, **52** (8), 1490–1499.
15. Kunkel, T.A. (1984) Mutational specificity of depurination. *Proc. Natl. Acad. Sci. U.S.A.*, **81** (5), 1494–1498.
16. (a) Fujii, R., Kitaoka, M., and Hayashi, K. (2006) Error-prone rolling circle amplification: the simplest random mutagenesis protocol. *Nat. Protoc.*, **1** (5), 2493–2497; (b) Fujii, R., Kitaoka, M., and Hayashi, K. (2004) One-step random mutagenesis by error-prone rolling circle amplification. *Nucleic Acids Res.*, **32** (19), e145.
17. Eigen, M. and Gardiner, W. (1984) Evolutionary molecular engineering based on RNA replication. *Pure Appl. Chem.*, **56** (8), 967–978.
18. (a) Mullis, K.B. (1994) The polymerase chain-reaction (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.*, **33** (12), 1209–1213; (b) Glick, B.R., Pasternak, J.J., and Patten, C.L. (2010) *Molecular Biotechnology: Principles and Applications of Recombinant DNA*, ASM Press, Washington, DC.
19. Koltermann, A. and Kettling, U. (1997) Principles and methods of evolutionary biotechnology. *Biophys. Chem.*, **66** (2-3), 159–177.
20. Matsumura, M. and Aiba, S. (1985) Screening for thermostable mutant of kanamycin nucleotidyltransferase by the use of a transformation system for a thermophile. *Bacillus Stearothermophilus*. *J. Biol. Chem.*, **260** (28), 15298–15303.
21. Baase, W.A., Liu, L., Tronrud, D.E., and Matthews, B.W. (2010) Lessons from the lysozyme of phage T4. *Protein Sci.*, **19** (4), 631–641.
22. Liao, H., Mckenzie, T., and Hageman, R. (1986) Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc. Natl. Acad. Sci. U.S.A.*, **83** (3), 576–580.
23. (a) Matteuchi, M.D. and Heyneker, H.L. (1983) Targeted random mutagenesis: the use of ambiguously synthesised oligonucleotides to mutagenize sequences immediately 5′ of an ATG initiation codon. *Nucleic Acids Res.*, **11**, 3113–3121; (b) Hui, A., Hayflick, J., Dinkelspiel, K., and de Boer, H.A. (1984) Mutagenesis of the three bases preceding the start codon of the ß-galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO J.*, **3** (3), 623–629; (c) Dreher, T.W., Bujarski, J.J., and Hall, T.C. (1984) Mutant viral RNAs synthesized in vitro show altered aminoacylation and replicase template activities. *Nature*, **311**

(5982), 171–175; (d) Seeburg, P.H., Colby, W.W., Capon, D.J., Goeddel, D.V., and Levinson, A.D. (1984) Biological properties of human c-Ha-ras1 genes mutated at codon 12. *Nature*, **312** (5989), 71–75; (e) Schultz, S.C. and Richards, J.H. (1986) Site-saturation studies of beta-lactamase: production and characterization of mutant ß-lactamases with all possible amino acid substitutions at residue 71. *Proc. Natl. Acad. Sci. U.S.A.*, **83** (6), 1588–1592; (f) Derbyshire, K.M., Salvo, J.J., and Grindley, N.D. (1986) A simple and efficient procedure for saturation mutagenesis using mixed oligodeoxynucleotides. *Gene*, **46** (2-3), 145–152; (g) Reidhaar-Olson, J.F. and Sauer, R.T. (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science*, **241** (4861), 53–57; (h) Oliphant, A.R., Nussbaum, A.L., and Struhl, K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44** (2–3), 177–183.

24. Estell, D.A., Graycar, T.P., and Wells, J.A. (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.*, **260** (11), 6518–6521.
25. Ner, S.S., Goodin, D.B., and Smith, M. (1988) A simple and efficient procedure for generating random point mutations and for codon replacements using mixed oligodeoxynucleotides. *DNA*, **7** (2), 127–134.
26. Zoller, M.J. and Smith, M. (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res.*, **10** (20), 6487–6500.
27. Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U.S.A.*, **82** (2), 488–492.
28. Hermes, J.D., Parekh, S.M., Blacklow, S.C., Koster, H., and Knowles, J.R. (1989) A reliable method for random mutagenesis - the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene*, **84** (1), 143–151.
29. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain-reaction. *Gene*, **77** (1), 51–59.
30. Lee, J., Shin, M.K., Ryu, D.K., Kim, S., and Ryu, W.S. (2010) Insertion and deletion mutagenesis by overlap extension PCR. *Methods Mol. Biol.*, **634**, 137–146.
31. Dube, D.K. and Loeb, L.A. (1989) Mutants generated by the insertion of random oligonucleotides into the active-site of the ß-lactamase gene. *Biochemistry*, **28** (14), 5703–5707.
32. Hogrefe, H.H., Cline, J., Youngblood, G.L., and Allen, R.M. (2002) Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques*, **33** (5), 1158–1160.
33. Kammann, M., Laufs, J., Schell, J., and Gronenborn, B. (1989) Rapid insertional mutagenesis of DNA by polymerase chain-reaction (PCR). *Nucleic Acids Res.*, **17** (13), 5404.
34. Sarkar, G. and Sommer, S.S. (1990) The megaprimer method of site-directed mutagenesis. *Biotechniques*, **8** (4), 404–407.
35. Reikofski, J. and Tao, B.Y. (1992) Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.*, **10** (4), 535–547.
36. (a) Leung, D.W., Chen, E., and Goeddel, D.V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, **1**, 11–15; (b) Cadwell, R.C. and Joyce, G.F. (1994) Mutagenic PCR. *PCR Methods Appl.*, **3** (6), S136–S140.
37. Hawkins, R.E., Russell, S.J., and Winter, G. (1992) Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.*, **226**, 889–896.
38. (a) Huse, W., Sastry, L., Iverson, S., Kang, A., Alting-Mees, M., Burton, D., Benkovic, S., and Lerner, R. (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science*, **246** (4935), 1275–1281; (b) Barbas, C.F., Bain, J.D., Hoekstra, D.M., and Lerner, R.A. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the

diversity problem. *Proc. Natl. Acad. Sci. U.S.A.*, **89** (10), 4457–4461.

39. (a) Eggert, T., Reetz, M.T., and Jaeger, K.-E. (2004) in *Enzyme Functionality – Design, Engineering, and Screening* (ed. A. Svendsen), Marcel Dekker, New York, pp. 375–390; (b) Ruff, A.J., Dennig, A., and Schwaneberg, U. (2013) To get what we aim for -- progress in diversity generation methods. *FEBS J.*, **280** (13), 2961–2978; (c) Hanson-Manful, P. and Patrick, W.M. (2013) Construction and analysis of randomized protein-encoding libraries using error-prone PCR. *Methods Mol. Biol.*, **996**, 251–267; (d) Copp, J.N., Hanson-Manful, P., Ackerley, D.F., and Patrick, W.M. (2014) Error-prone PCR and effective generation of gene variant libraries for directed evolution. *Methods Mol. Biol.*, **1179**, 3–22.

40. (a) Zhang, H.Y., Zhang, J., Lin, L., Du, W.Y., and Lu, J. (1993) Enhancement of the stability and activity of aspartase by random and site-directed mutagenesis. *Biochem. Biophys. Res. Commun.*, **192** (1), 15–21; (b) Zhang, J., Li, Z.-Q., and Zhang, H.-Y. (1992) An enzymatic method for random- (site-specific) mutagenesis of Ginseng gene in vitro. *Chin. Biochem. J.*, **8** (1), 115–120.

41. Chen, K.Q. and Arnold, F.H. (1993) Tuning the activity of an enzyme for unusual environments – sequential random mutagenesis of subtilisin-E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U.S.A.*, **90** (12), 5618–5622.

42. Moore, J.C. and Arnold, F.H. (1996) Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotechnol.*, **14** (4), 458–467.

43. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes – gene-splicing by overlap extension. *Gene*, **77** (1), 61–68.

44. (a) Stemmer, W.P.C. (1994) Rapid evolution of a protein in-vitro by DNA shuffling. *Nature*, **370** (6488), 389–391; (b) Stemmer, W.P.C. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **91** (22), 10747–10751.

45. Brown, S. (1992) Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage lambda receptor. *Proc. Natl. Acad. Sci. U.S.A.*, **89** (18), 8651–8655.

46. Gilbert, W. (1978) Why genes in pieces? *Nature*, **271** (5645), 501.

47. Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, **391** (6664), 288–291.

48. (a) Posner, B.A., Li, L.Y., Bethell, R., Tsuji, T., and Benkovic, S.J. (1996) Engineering specificity for folate into dihydrofolate reductase from *Escherichia coli*. *Biochemistry*, **35** (5), 1653–1663; (b) Warren, M.S., Marolewski, A.E., and Benkovic, S.J. (1996) A rapid screen of active site mutants in glycinamide ribonucleotide transformylase. *Biochemistry*, **35** (27), 8855–8862.

49. Kamtekar, S., Schiffer, J.M., Xiong, H.Y., Babik, J.M., and Hecht, M.H. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science*, **262** (5140), 1680–1685.

50. Lutz, S. and Benkovic, S. (2002) Engineering protein evolution, in *Directed Molecular Evolution of Proteins* (eds S. Brakmann and K. Johnsson), Wiley-VCH Verlag GmbH, Weinheim.

51. (a) Smith, G. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228** (4705), 1315–1317; (b) Smith, G.P. and Petrenko, V.A. (1997) Phage display. *Chem. Rev.*, **97** (2), 391–410.

52. (a) Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D., and Winter, G. (1991) By-passing immunization. *J. Mol. Biol.*, **222** (3), 581–597; (b) Clackson, T., Hoogenboom, H.R., Griffiths, A.D., and Winter, G. (1991) Making antibody fragments using phage display libraries. *Nature*, **352** (6336), 624–628.

53. Barbas, C.F. III,, Kang, A.S., Lerner, R.A., and Benkovic, S.J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. U.S.A.*, **88** (18), 7978–7982.

54. Soumillion, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G., and Fastrez, J. (1994) Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J. Mol. Biol.*, **237** (4), 415–422.

55. (a) Janda, K.D., Lo, C.H., Li, T., Barbas, C.F. III,, Wirsching, P., and Lerner, R.A. (1994) Direct selection for a catalytic mechanism from combinatorial antibody libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **91** (7), 2532–2536; (b) Janda, K.D., Lo, L.C., Lo, C.H., Sim, M.M., Wang, R., Wong, C.H., and Lerner, R.A. (1997) Chemical selection for catalysis in combinatorial antibody libraries. *Science*, **275** (5302), 945–948.

56. Jestin, J.L., Kristensen, P., and Winter, G. (1999) A method for the selection of catalytic activity using phage display and proximity coupling. *Angew. Chem. Int. Ed.*, **38** (8), 1124–1127.

57. Pedersen, H., Holder, S., Sutherlin, D.P., Schwitter, U., King, D.S., and Schultz, P.G. (1998) A method for directed evolution and functional cloning of enzymes. *Proc. Natl. Acad. Sci. U.S.A.*, **95** (18), 10523–10528.

58. Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **94** (23), 12297–12302.

59. Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.*, **15** (6), 553–557.

60. (a) Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., and Curtiss, R. III, (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol.*, **15** (1), 29–34; (b) Daugherty, P.S., Iverson, B.L., and Georgiou, G. (2000) Flow cytometric screening of cell-based libraries. *J. Immunol. Methods*, **243** (1-2), 211–227.

61. Griffiths, A.D. and Tawfik, D.S. (2000) Man-made enzymes – from design to in vitro compartmentalisation. *Curr. Opin. Biotechnol.*, **11** (4), 338–353.

62. Acevedo-Rocha, C.G., Agudo, R., and Reetz, M.T. (2014) Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems. *J. Biotechnol.*, **191**, 3–10.

63. (a) Zhao, H. and Arnold, F.H. (1997) Combinatorial protein design: strategies for screening protein libraries. *Curr. Opin. Struct. Biol.*, **7** (4), 480–485; (b) Taylor, S.V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed.*, **40** (18), 3310–3335.

64. (a) Schultz, P.G. and Lerner, R.A. (1993) Antibody catalysis of difficult chemical transformations. *Acc. Chem. Res.*, **26** (8), 391–395; (b) Mader, M.M. and Bartlett, P.A. (1997) Binding energy and catalysis: the implications for transition-state analogs and catalytic antibodies. *Chem. Rev.*, **97** (5), 1281–1302; (c) Hilvert, D. (2000) Critical analysis of antibody catalysis. *Annu. Rev. Biochem.*, **69**, 751–793; (d) Keinan, E. (ed) (2005) *Catalytic Antibodies*, Wiley-VCH Verlag GmbH, Weinheim.

65. (a) Firth, A.E. and Patrick, W.M. (2005) Statistics of protein library construction. *Bioinformatics*, **21** (15), 3314–3315; (b) Firth, A.E. and Patrick, W.M. (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.*, **36** (Web Server issue), W281–W285.

66. Bosley, A.D. and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng*, **22** (1-3), 57–61.

67. Denault, M. and Pelletier, J.N. (2007) in *Protein Engineering Protocols* (eds K.M. Arndt and K.M. Müller), Humana Press, Totowa, NJ, pp. 127–154.

68. Reetz, M.T. and Carballeira, J.D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.*, **2** (4), 891–903.

69. (a) Reetz, M.T., Zonta, A., Schimossek, K., Liebeton, K., and Jaeger, K.E. (1997) Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. *Angew. Chem. Int. Ed. Engl.*, **36** (24), 2830–2832; (b) Reetz, M.T. (1999)

Strategies for the development of enantioselective catalysts. *Pure Appl. Chem.*, **71** (8), 1503–1509.

70. Personal account of directed evolution of stereoselective enzymes: Reetz, M.T. (2012) Laboratory evolution of stereoselective enzymes as a means to expand the toolbox of organic chemists. *Tetrahedron*, **68** (37), 7530–7548.
71. Reetz, M.T., Becker, M.H., Klein, H.W., and Stöckigt, D. (1999) A method for high-throughput screening of enantioselective catalysts. *Angew. Chem. Int. Ed.*, **38** (12), 1758–1761.
72. (a) Reetz, M.T. (2004) Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proc. Natl. Acad. Sci. U.S.A.*, **101** (16), 5716–5722; (b) Lutz, S. and Patrick, W.M. (2004) Novel methods for directed evolution of enzymes: quality, not quantity. *Curr. Opin. Biotechnol.*, **15** (4), 291–297.
73. (a) Keasling, J.D. (2010) Manufacturing molecules through metabolic engineering. *Science*, **330** (6009), 1355–1358; (b) Marcheschi, R.J., Gronenberg, L.S., and Liao, J.C. (2013) Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol. J.*, **8** (5), 545–555; (c) Bar-Even, A. and Salah Tawfik, D. (2013) Engineering specialized metabolic pathways – is there a room for enzyme improvements? *Curr. Opin. Biotechnol.*, **24** (2), 310–319; (d) Sun, X., Shen, X., Jain, R., Lin, Y., Wang, J., Sun, J., Wang, J., Yan, Y., and Yuan, Q. (2015) Synthesis of chemicals by metabolic engineering of microbes. *Chem. Soc. Rev.*, **44** (11), 3760–3785; (e) Jullesson, D., David, F., Pfleger, B., and Nielsen, J. (2015) Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.*, **33** (7), 1395–1402.
74. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., Aryee, M.J., and Joung, J.K. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523** (7561), 481–485.
75. (a) Ihssen, J., Haas, J., Kowarik, M., Wiesli, L., Wacker, M., Schwede, T., and Thony-Meyer, L. (2015) Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.*, **5** (4), 140227; (b) Ye, J., Wen, F., Xu, Y., Zhao, N., Long, L., Sun, H., Yang, J., Cooley, J., Todd Pharr, G., Webby, R., and Wan, X.F. (2015) Error-prone PCR-based mutagenesis strategy for rapidly generating high-yield influenza vaccine candidates. *Virology*, **482**, 234–243; (c) Horiya, S., MacPherson, I.S., and Krauss, I.J. (2014) Recent strategies targeting HIV glycans in vaccine design. *Nat. Chem. Biol.*, **10** (12), 990–999; (d) Kwan, D.H., Constantinescu, I., Chapanian, R., Higgins, M.A., Kötzler, M.P., Samain, E., Boraston, A.B., Kizhakkedathu, J.N., and Withers, S.G. (2015) Toward efficient enzymes for the generation of universal blood through structure-guided directed evolution. *J. Am. Chem. Soc.*, **137**, 5695–5705.
76. (a) Grimm, S.K., Battles, M.B., and Ackerman, M.E. (2015) Directed evolution of a yeast-displayed HIV-1 SOSIP gp140 spike protein toward improved expression and affinity for conformational antibodies. *PLoS One*, **10** (2), e0117227; (b) Temme, J.S., MacPherson, I.S., DeCourcey, J.F., and Krauss, I.J. (2014) High temperature SELMA: evolution of DNA-supported oligomannose clusters which are tightly recognized by HIV bnAb 2G12. *J. Am. Chem. Soc.*, **136** (5), 1726–1729; (c) Julian, M.C., Lee, C.C., Tiller, K.E., Rabia, L.A., Day, E.K., Schick, A.J. III,, and Tessier, P.M. (2015) Co-evolution of affinity and stability of grafted amyloid-motif domain antibodies. *Protein Eng. Des. Sel.*, **28** (10), 339–350.
77. (a) Zhan, T., Zhang, K., Chen, Y., Lin, Y., Wu, G., Zhang, L., Yao, P., Shao, Z., and Liu, Z. (2013) Improving glyphosate oxidation activity of glycine oxidase from *Bacillus cereus* by directed evolution. *PLoS One*, **8** (11), e79175; (b) Pollegioni, L. and Molla, G. (2011) New biotech applications from evolved D-amino acid oxidases. *Trends Biotechnol.*,

29 (6), 276–283; (c) Tian, Y.S., Xu, J., Zhao, W., Xing, X.J., Fu, X.Y., Peng, R.H., and Yao, Q.H. (2015) Identification of a phosphinothricin-resistant mutant of rice glutamine synthetase using DNA shuffling. *Sci. Rep.*, **5**, 15495; (d) Han, H., Zhu, B., Fu, X., You, S., Wang, B., Li, Z., Zhao, W., Peng, R., and Yao, Q. (2015) Overexpression of D-amino acid oxidase from *Bradyrhizobium japonicum*, enhances resistance to glyphosate in *Arabidopsis thaliana*. *Plant Cell Rep.*, **34** (12), 2043–2051; (e) Yao, P., Lin, Y., Wu, G., Lu, Y., Zhan, T., Kumar, A., Zhang, L., and Liu, Z. (2015) Improvement of glycine oxidase by DNA shuffling and site-saturation mutagenesis of F247 residue. *Int. J. Biol. Macromol.*, **79**, 965–970.

78. Steensels, J., Snoek, T., Meersman, E., Picca Nicolino, M., Voordeckers, K., and Verstrepen, K.J. (2014) Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.*, **38** (5), 947–995.

79. Cai, Z., Liu, G., Zhang, J., and Li, Y. (2014) Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of rubisco. *Protein Cell*, **5** (7), 552–562.

80. (a) Pan, J., Wu, F., Wang, J., Yu, L., Khayyat, N.H., Stark, B.C., and Kilbane, J.J. II, (2013) Enhancement of desulfurization activity by enzymes of the *Rhodococcus* dsz operon through coexpression of a high sulfur peptide and directed evolution. *Fuel*, **112**, 385–390; (b) Fosso-Kankeu, E. and Mulaba-Bafubiandi, A.F. (2014) Implication of plants and microbial metalloproteins in the bioremediation of polluted waters: a review. *Phys. Chem. Earth.*, **67-69**, 242–252; (c) Peixoto, R.S., Vermelho, A.B., and Rosado, A.S. (2011) Petroleum-degrading enzymes: bioremediation and new prospects. *Enzyme Res.*, **2011**, 475193; (d) Fukukawa, K. (2006) Oxygenases and dehalogenases: molecular approaches to efficient degradation of chlorinated environmental pollutants. *Biosci. Biotechnol., Biochem.*, **70** (10), 2335; (e) Janssen, D.B. (2007) Biocatalysis by dehalogenating enzymes. *Adv. Appl. Microbiol.*, **61**, 233–252; (f) Duprey, A., Chansavang, V., Fremion, F., Gonthier, C., Louis, Y., Lejeune, P., Springer, F., Desjardin, V., Rodrigue, A., and Dorel, C. (2014) “NiCo buster”: engineering *E. coli* for fast and efficient capture of cobalt and nickel. *J. Biol. Eng.*, **8**, 19; (g) Shen, S., Li, X.-F., Cullen, W.R., Weinfeld, M., and Le, X.C. (2013) Arsenic binding proteins. *Chem. Rev.*, **113** (10), 7769–7792.

81. (a) Weinreich, D.M., Delaney, N.F., DePristo, M.A., and Hartl, D.L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, **312** (5770), 111–114; (b) Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., and Cooper, T.F. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, **332** (6034), 1193–1196; (c) Salverda, M.L., Dellus, E., Gorter, F.A., Debets, A.J., van der Oost, J., Hoekstra, R.F., Tawfik, D.S., and deVisser, J.A. (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.*, **7**, e1001321.

82. (a) Laos, R., Shaw, R., Leal, N.A., Gaucher, E., and Benner, S. (2013) Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry*, **52** (31), 5288–5294; (b) Zhang, L., Yang, Z., Sefah, K., Bradley, K.M., Hoshika, S., Kim, M.J., Kim, H.J., Zhu, G., Jimenez, E., Cansiz, S., Teng, I.T., Champanhac, C., McLendon, C., Liu, C., Zhang, W., Gerloff, D.L., Huang, Z., Tan, W., and Benner, S.A. (2015) Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.*, **137**, 6734–6737.

83. Turner, N.J. and O’Reilly, E. (2013) Biocatalytic retrosynthesis. *Nat. Chem. Biol.*, **9** (5), 285–288.

84. (a) Muschiol, J., Peters, C., Oberleitner, N., Mihovilovic, M.D., Bornscheuer, U.T., and Rudroff, F. (2015) Cascade catalysis – strategies and challenges en route to preparative synthetic biology. *Chem. Commun.*, **51**, 5798–5811; (b) Fessner, W.-D. (2015) Systems biocatalysis: development and engineering of cell-free artificial metabolisms for

preparative multi-enzymatic synthesis. *New Biotechnol.*, **32**, 658–664; (c) Riva, S. and Fessner, W.-D. (eds) (2014) *Cascade Biocatalysis*, Wiley-VCH Verlag GmbH, Weinheim; (d) Denard, C.A., Hartwig, J.F., and Zhao, H. (2013) Multistep one-pot reactions combining biocatalysts and chemical catalysts for asymmetric synthesis. *ACS Catal.*, **3**, 2856–2864; (e) Sato, H., Hummel, W., and Gröger, H. (2015) Cooperative catalysis of noncompatible catalysts through compartmentalization: wacker oxidation and enzymatic reduction in a one-pot process in aqueous media. *Angew. Chem. Int. Ed.*, **54**, 4488–4492; (f) Tessaro, D., Pollegioni, L., Piubelli, L., D'Arrigo, P., and Servi, S. (2015) Systems biocatalysis: an artificial metabolism for interconversion of functional groups. *ACS Catal.*, **5**, 1604–1608; (g) Agudo, R. and Reetz, M.T. (2013) Designer cells for stereocomplementary de novo enzymatic cascade reactions based on laboratory evolution. *Chem. Commun.*, **49**, 10914–10916.