

Manfred T. Reetz

Directed Evolution of Selective Enzymes

Manfred T. Reetz

Directed Evolution of Selective Enzymes

Catalysts for Organic Chemistry and Biotechnology

WILEY-VCH
Verlag GmbH & Co. KGaA

Author

Manfred T. Reetz
MPI für Kohlenforschung
Kaiser-Wilhelm-Platz 1
45470 Mülheim
Germany

and

Philipps-Universität Marburg
Fachbereich Chemie
Hans-Meerwein-Straße 4
35032 Marburg
Germany

Cover

Enzyme structure -
<http://dx.doi.org/10.2210/pdb3g02/pdb>

All books published by **Wiley-VCH** are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <<http://dnb.d-nb.de>>.

© 2017 Wiley-VCH Verlag GmbH & Co.
KGaA, Boschstr. 12, 69469 Weinheim,
Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Print ISBN: 978-3-527-31660-1

ePDF ISBN: 978-3-527-65549-6

ePub ISBN: 978-3-527-65548-9

Mobi ISBN: 978-3-527-65547-2

oBook ISBN: 978-3-527-65546-5

Cover Design Schulz Grafik-Design,
Fußgönheim, Germany
Typesetting SPi Global, Chennai, India
Printing and Binding

Printed on acid-free paper

Contents

Preface *IX*

1	Introduction to Directed Evolution	1
1.1	General Definition and Purpose of Directed Evolution of Enzymes	1
1.2	Brief Account of the History of Directed Evolution	4
1.3	Applications of Directed Evolution of Enzymes	16
	References	17
2	Selection versus Screening in Directed Evolution	27
2.1	Selection Systems	27
2.2	Screening Systems	44
2.3	Conclusions and Perspectives	52
	References	53
3	Gene Mutagenesis Methods	59
3.1	Introductory Remarks	59
3.2	Error-Prone Polymerase Chain Reaction (epPCR) and Other Whole-Genome Mutagenesis Techniques	60
3.3	Saturation Mutagenesis: Away from Blind Directed Evolution	70
3.4	Recombinant Gene Mutagenesis Methods	85
3.5	Circular Permutation and Other Domain Swapping Techniques	91
3.6	Solid-Phase Combinatorial Gene Synthesis for Library Creation	92
3.7	Computational Tools	96
	References	101
4	Strategies for Applying Gene Mutagenesis Methods	115
4.1	General Guidelines	115
4.2	Rare Cases of Comparative Studies	118
4.3	Choosing the Best Strategy when Applying Saturation Mutagenesis	130
4.3.1	General Guidelines	130

4.3.2	Choosing Optimal Pathways in Iterative Saturation Mutagenesis (ISM)	135
4.3.3	Systematization of Saturation Mutagenesis	142
4.3.4	Single Code Saturation Mutagenesis (SCSM): Use of a Single Amino Acid as Building Block	149
4.3.5	Triple Code Saturation Mutagenesis (TCSM): A Viable Compromise when Choosing the Optimal Reduced Amino Acid Alphabet	151
4.4	Techno-Economical Analyses of Saturation Mutagenesis Strategies	154
4.5	Combinatorial Solid-Phase Gene Synthesis: An Alternative for the Future?	159
	References	160
5	Selected Examples of Directed Evolution of Enzymes with Emphasis on Stereo- and Regioselectivity, Substrate Scope, and/or Activity	167
5.1	Explanatory Remarks	167
5.2	Collection of Selected Examples from the Literature 2010 up to 2016	189
	References	189
6	Directed Evolution of Enzyme Robustness	205
6.1	Introduction	205
6.2	Application of epPCR and DNA Shuffling	207
6.3	B-FIT Approach	211
6.4	Iterative Saturation Mutagenesis (ISM) at Protein–Protein Interfacial Sites for Multimeric Enzymes	215
6.5	Ancestral and Consensus Approaches and their Structure-Guided Extensions	216
6.6	Computationally Guided Methods	219
6.6.1	SCHEMA Approach	219
6.6.2	FRESCO Approach	221
6.6.3	FireProt Approach	223
6.6.4	Constrained Network Analysis (CNA) Approach	224
6.6.5	Alternative Approaches	226
	References	227
7	Directed Evolution of Promiscuity: Artificial Enzymes as Catalysts in Organic Chemistry	237
7.1	Introductory Background Information	237
7.2	Tuning the Catalytic Profile of Promiscuous Enzymes by Directed Evolution	245
7.3	Conclusions and Perspectives	259
	References	260

8	Learning from Directed Evolution	267
8.1	Background Information	267
8.2	Case Studies Featuring Mechanistic, Structural, and/or Computational Analyses of the Source of Evolved Stereo- and/or Regioselectivity	269
8.2.1	Epoxide Hydrolase	269
8.2.2	Ene-Reductase of the Old Yellow Enzyme (OYE)	273
8.2.3	Esterase	279
8.2.4	Cytochrome P450 Monooxygenase	282
8.3	Additive versus Non-additive Mutational Effects in Fitness Landscapes	287
	References	296
	Index	303

Preface

Directed evolution is a term that is used in two distinctly different research areas: (i) The genetic manipulation of functional RNAs, a discipline initiated by S. Spiegelmann half a century ago and extending to the present day in the laboratories of J. W. Szostak, J. F. Joyce, and others and (ii) the genetic manipulation of genes (DNA) with the aim to engineer the catalytic profiles of enzymes as catalysts in organic chemistry and biotechnology, especially stereoselectivity. This monograph focuses on the latter field. It begins with an introductory chapter that features the basic principles of directed evolution, and is followed by a chapter on screening and selection methods. Critical analyses of recent developments constitute the heart of the monograph. Rather than being comprehensive, emphasis is placed on methodology development in the quest to maximize efficiency, reliability, and speed when performing this type of protein engineering. The primary applications concern the synthesis of chiral pharmaceuticals, fragrances, and plant protecting agents.

The directed evolution methods and strategies featured in this book can also be used when engineering metabolic pathways, developing vaccines, engineering antibodies, creating genetically modified yeasts for the food industry, engineering proteins for pollution control, developing photosynthetic CO₂ fixation, genetically modifying plants for agricultural and medicinal purposes, engineering CRISPR-Cas9 nucleases for genome editing, and modifying DNA polymerases for forensic purposes and for accepting non-natural nucleotides. A few studies of these applications are included here.

This monograph is intended not only for those who are interested in learning the basics of directed evolution of enzymes, but also for advanced researchers in academia and industry who seek guidelines for performing protein engineering efficiently.

I wish to thank Dr Zhoutong Sun for reading Chapters 3 and 4 and discussing some of the issues related to molecular biology. Thanks also goes to Dr Gheorghe-Doru Roiban and Dr Adriana Ilie for editing all the chapters and constructing some of the figures. Any errors that may remain are the responsibility of the author.

1

Introduction to Directed Evolution

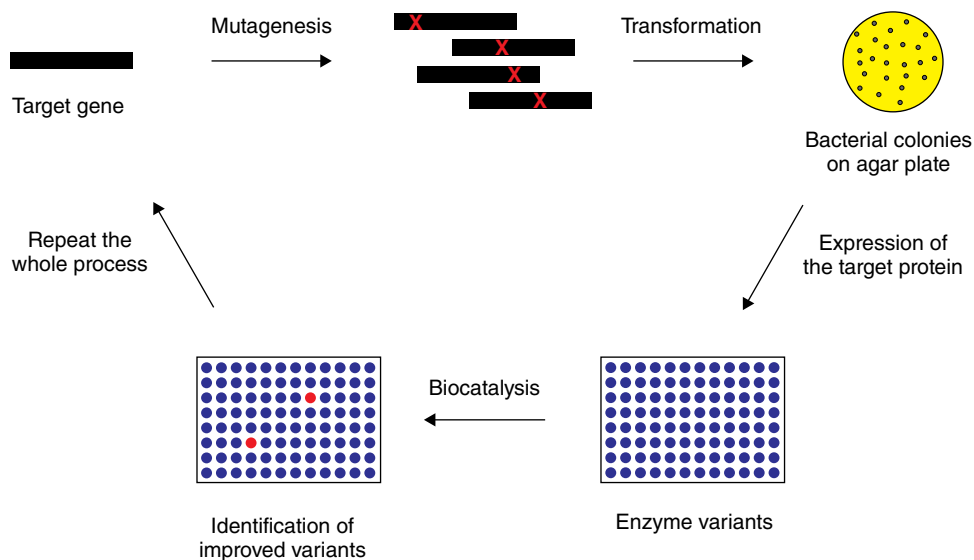
1.1

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



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 (bottle-neck). 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]:

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

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).

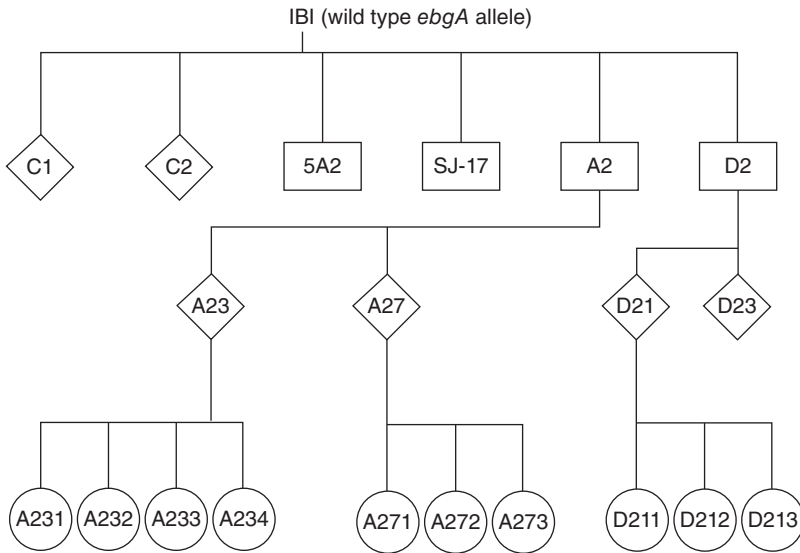
1.2

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 β -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.)

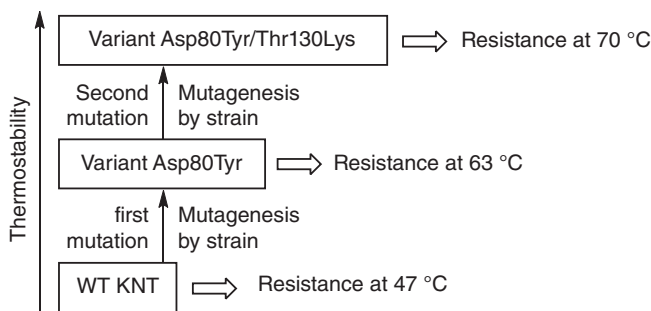
Although originally not specifically related to directed evolution, developments such as the Kunkel method of mutational specificity based on depurination [15] deserves mention because it was used two decades later in mutant library design based on error-prone rolling circle amplification (epRCA) [16]. 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 Q ϕ 3. 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



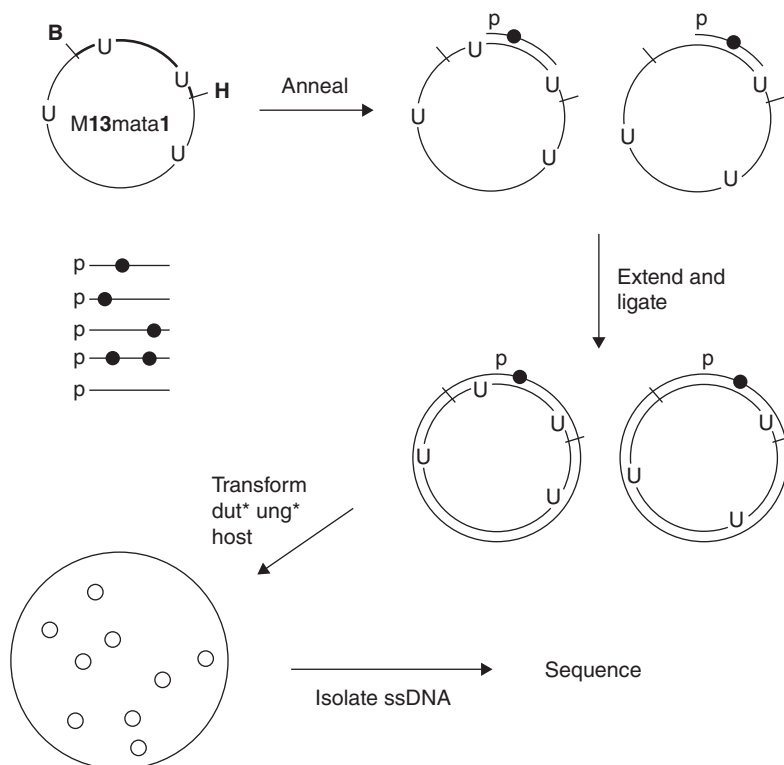
Scheme 1.4 Early example of directed evolution of thermostability with kanamycin nucleotidyltransferase (KNT) serving as the enzyme and a mutator strain as the random mutagenesis technique in an iterative manner [22].

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 H₂O₂ 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 *MATa1* 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 β -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 Phe⁶⁶XXXSer⁷⁰XXLys⁷³ contains 15 base pairs of chemically synthesized random sequences that code for 2.5 million amino acid exchanges. It should be noted that β -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 QuikChangeTM 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 MgCl₂ or MnCl₂ 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 *Hind*III/*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 β -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

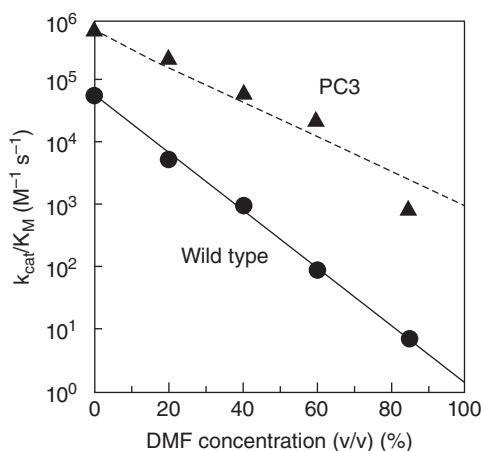
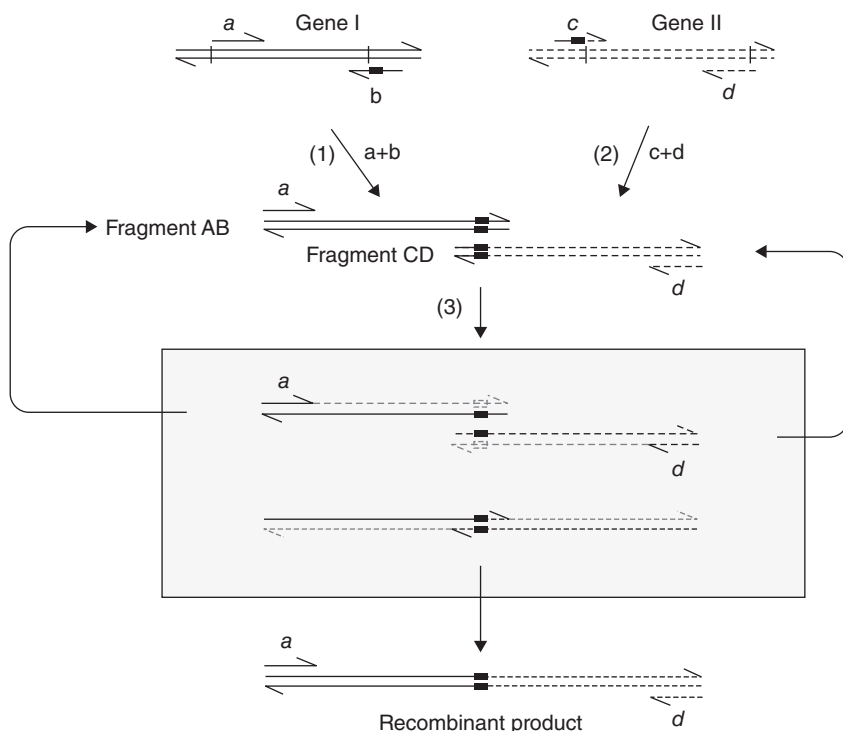


Figure 1.1 Catalytic efficiency of WT subtilisin E and variant PC3 as catalysts in the hydrolytic cleavage of *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Met-*p*-nitroanilide [41]. (Adapted from Chen and Arnold [41]. Reproduced with permission of National Academy of Sciences.)

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

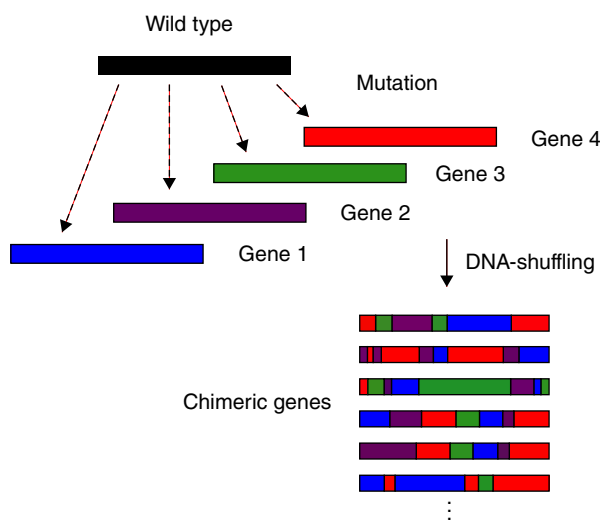


Scheme 1.6 Steps in the recombinant technique of splicing by overlap extension (SOE), illustrated here using two different genes [43]. (Adapted from Horton *et al.* [43]. Reproduced with permission of Elsevier.)

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, β -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).

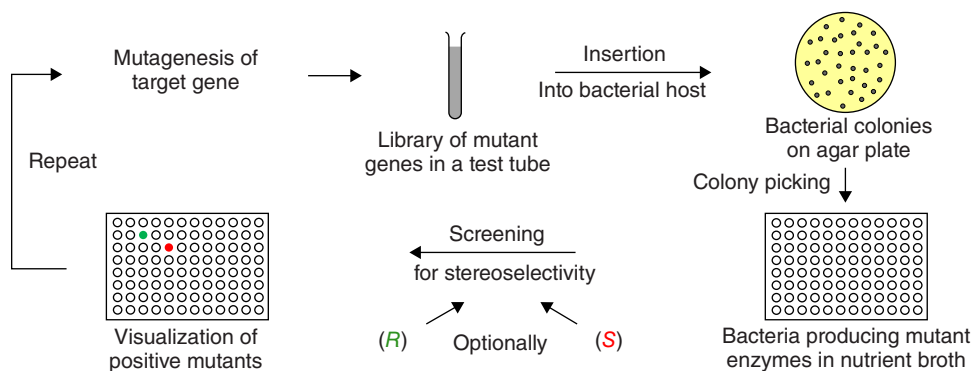
The distinction between selection and screening [63a] was recognized by Hilvert and coworkers in the 1990s, who consequently developed impressive selection systems in which the host organism experiences a growth advantage due to the generation of enzyme mutants displaying desired properties [63b]. Applying 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).

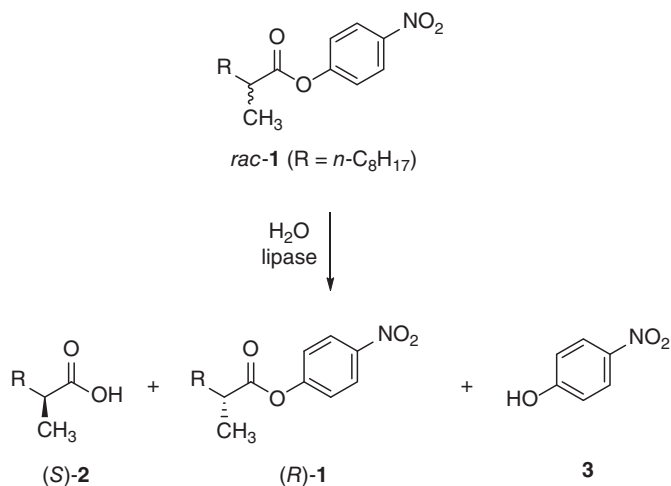
While the creation of enhanced enzyme thermostability paved the way for potential applications in biotechnology, realizing the potentially broad utility of directed evolution as a prolific source of selective catalysts in synthetic organic chemistry was still to come. In the mid-1990s the Reetz group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

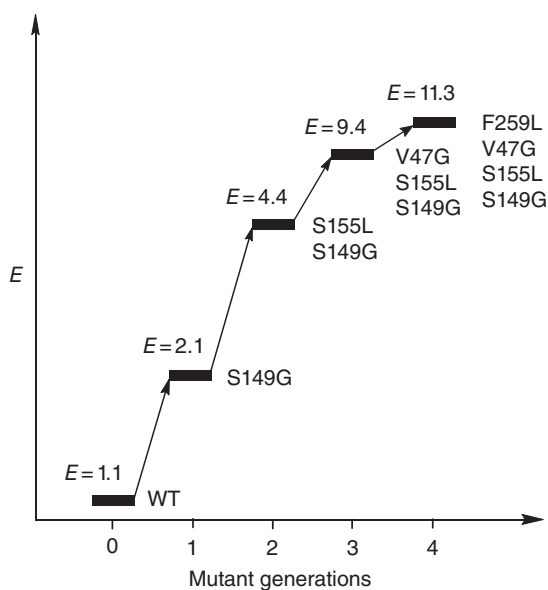


Scheme 1.8 Concept of directed evolution of stereoselective enzymes with (*R*)- or (*S*)-selective mutants being accessible on an optional basis [69]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

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 CO₂ 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 (2S,4R)-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 β -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 β -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 β -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. Cramer, 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 glycnamide 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) Bypassing 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. Soumillon, P., Jespers, L., Bouchet, M., Marchand-Brynaert, J., Winter, G., and Fastez, 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) Julleson, 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. (2015) 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., DeCoursey, 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-Bafubandi, 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.

2

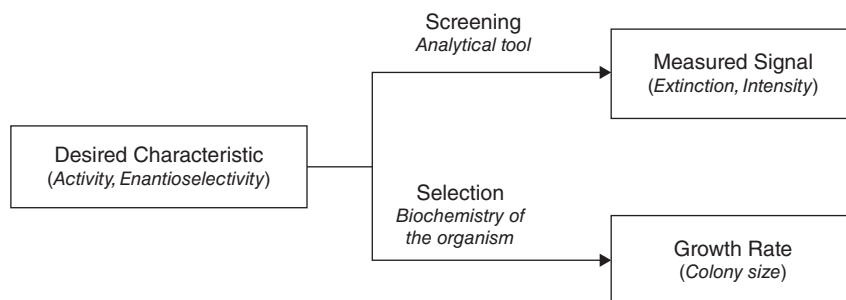
Selection versus Screening in Directed Evolution

2.1

Selection Systems

Efficient high-throughput assays for assessing activity, stereoselectivity, and thermostability of enzymes constitute essential components of directed evolution [1]. The terms “screening” and “selection” are sometimes used interchangeably, albeit in a confusing manner. In fact, they are succinctly different [2]. Screening means the (medium- or high-throughput) measurement of a given enzyme property such as activity and/or enantioselectivity by an automated analytical technique such as UV/vis spectroscopy, fluorescence, multiplex mass spectrometry (MS), robotic gas chromatography (GC), or HPLC. Genetic selection, on the other hand, involves an experimental platform in which the host organism has a growth and survival advantage because it harbors an enzyme or mutants thereof with a desired (evolved) catalytic profile (Scheme 2.1). This definition of selection is favored by many researchers [3]. It means that only those colonies that harbor the desired improved variants will appear on agar plates, which is a highly attractive feature. However, the fact that many examples of selection have been reported [1–3], does not mean that this method can be used in a general way to identify improved mutants in directed evolution. Whatever approach is chosen, screening or selection, the linkage of genotype to phenotype must be maintained.

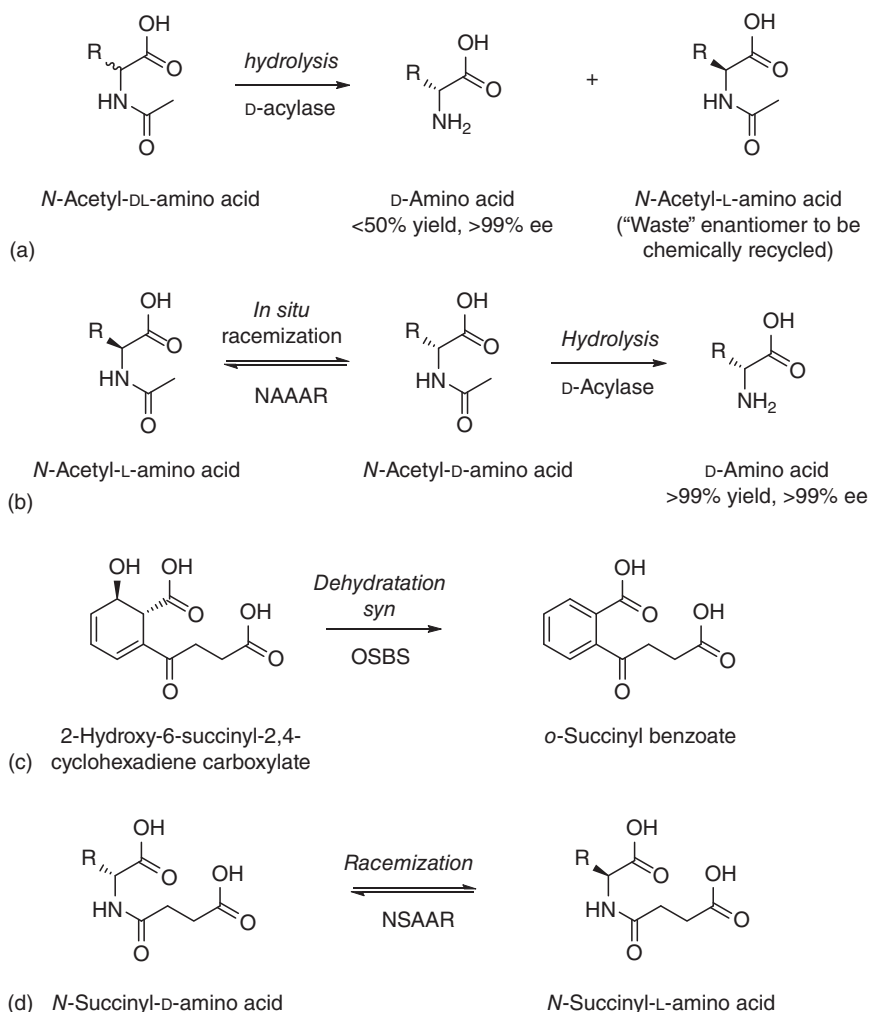
In the study reporting DNA shuffling for the first time, the activity of TEM-1 β -lactamase in *Escherichia coli* as the catalyst in the inactivating hydrolysis of the antibiotic cefotaxime was increased stepwise, selection pressure being based on antibiotic resistance [4]. Another typical example pertains to the directed evolution of an aspartate aminotransferase with extended substrate acceptance in which the selection system makes use of the auxotrophy of an *E. coli* strain deficient in the branched-chain amino acid transferase gene *ilvE* [5]. This enzyme catalyzes the last step of the biosynthesis of these amino acids, which means that the *ilvE*-deficient strain cannot grow on a minimal plate in the absence of the supplement valine, isoleucine, and leucine. The concept of employing auxotrophic strains that grow only when a given (mutant) enzyme is generated, which replaces a missing cellular protein, has been implemented in other studies as well. Unfortunately, in



Scheme 2.1 Screening versus selection in directed evolution [2]. (Acevedo-Rocha *et al.* [2]. Reproduced with permission of Elsevier.)

such cases even low enzyme activity may suffice for cell growth, which hampers the development of an efficient selection system. An elegant way to solve this problem is to tune transcription with an enzyme-degradation tag, which reduces intracellular protein concentrations from high to low levels, a concept that was applied successfully to chorismate mutase [6]. Another example of genetic complementation pertains to an *E. coli* strain that was engineered so as not to accept glucose as a carbon source with the aim of discovering genes encoding latent glucokinase activity in an overexpression library [7]. Chemical complementation (*in vivo*) avoids some of the drawbacks of conventional genetic complementation (each enzyme needs a new assay). It is based on the use of reporters such as β -galactosidase or amino acid selectable markers of yeast, which are linked to the substrate of interest [8].

A different selection method was devised for increasing the activity of an *N*-acyl amino acid racemase (NAAAR) by directed evolution, the goal being the establishment of a system enabling dynamic kinetic resolution (DKR) of amino acids on an industrial scale [9]. Efficient enantio-differentiation from a rapidly racemizing mixture of chiral *N*-acyl amino acids was ensured by a *D*-acylase known to be highly stereoselective. Selection pressure with appearance of improved NAAAR variants was implemented by linking the racemization rate to the viability of the *E. coli* host. This requirement was addressed by disabling its natural *L*-methionine biosynthetic pathway, while also eliminating a *D*-amino acid racemization pathway. The chemistry involved in this system is shown in Scheme 2.2 [9]. The combined action of improved NAAAR and stereoselective *D*-acylase ensured efficient DKR of different amino acids at an industrially practical level. Initial mutagenesis was performed using a mutator strain (XL1-Red) ($>10^7$ variants), which delivered an improved variant G291D showing higher activity. This hot spot was then subjected to saturation mutagenesis (100 variants), which led to a better variant G291D. This mutant was subsequently subjected to error-prone polymerase chain reaction (epPCR) ($>10^5$ variants), leading to the final double mutant G291D/F323Y with a sixfold increase in activity relative to wild-type (WT) NAAAR. It needs to be pointed out that this elegant system involves directed evolution of activity, not enantioselectivity.



Scheme 2.2 Chemistry involved in the directed evolution of *N*-acyl amino acid racemase (NAAAR) with the aim of increasing its activity for dynamic kinetic resolution of amino acids [9]. (Baxter *et al.* [9]. Reproduced with permission of American Chemical Society.)

A recent example of designed growth-selection pressure in directed evolution and pathway engineering utilizes a related method in order to increase the efficiency of an NADPH-dependent homophenylalanine dehydrogenase [10]. Again, this is not selection-based directed evolution of stereoselectivity. Earlier, a simple and efficient on-plate selection system for identifying active epoxide hydrolases in an *E. coli* strain had been devised [11]. Epoxides are known to be toxic to many organisms, but hydrolysis with formation of the respective diol generally causes detoxification. Thus, the more reactive an epoxide hydrolase under defined conditions is, the better the chances for cell growth and survival.

Accordingly, agar plates containing *E. coli* were first treated with various amounts of a chiral epoxide before the normal directed evolution procedure was initiated. After a certain growth period (several days), visual inspection of the plates was all that was necessary to identify positive (active) hits in a large mutant library [11]. The improved variants can then be isolated, characterized, and tested as catalysts in the hydrolytic kinetic resolution of racemic epoxides as substrates for possible enhanced enantioselectivity. In order to check the viability of this crude but useful pre-selection system, an agar plate harboring *E. coli* and an epoxide was charged with 92 inactive and 4 active epoxide hydrolase mutants at defined positions. Following incubation, visual inspection correctly identified the active variants (Figure 2.1) [11]. The system can be automated for high-throughput identification of active epoxide hydrolases from large collections of mutants. It requires the simple transfer of fresh transformants manually or automatically by a robot (e.g., QPix or Genetix) to the epoxide-containing agar plates harboring *E. coli* in 96-well format. It is possible to produce hundreds of such plates per day. Extension to the selection of stereoselectivity still needs to be developed.

Most of the above and numerous other examples of selection-based directed evolution constitute impressive achievements, yet genetic selection is not as general as one would like it to be [2]. For many enzyme types, it is difficult to construct such experimental platforms. Even more challenging is the development of selection systems for laboratory evolution directed toward the control of stereoselectivity, and indeed very few examples have been reported [2]. The difficulty is related to the following question: Why should an organism have a

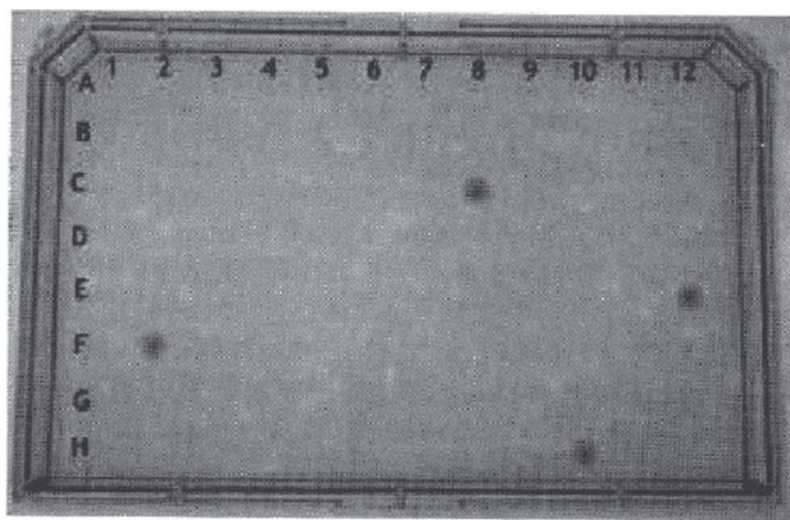
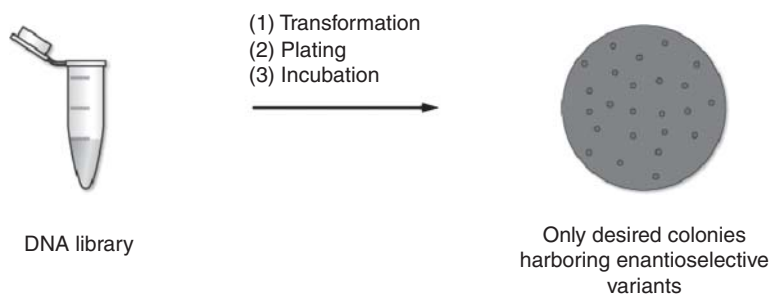


Figure 2.1 Agar plate harboring 96 *E. coli* colonies in the presence of 8 mM of an epoxide after 8 days of incubation [11]. The four spots (colonies) indicate the presence of

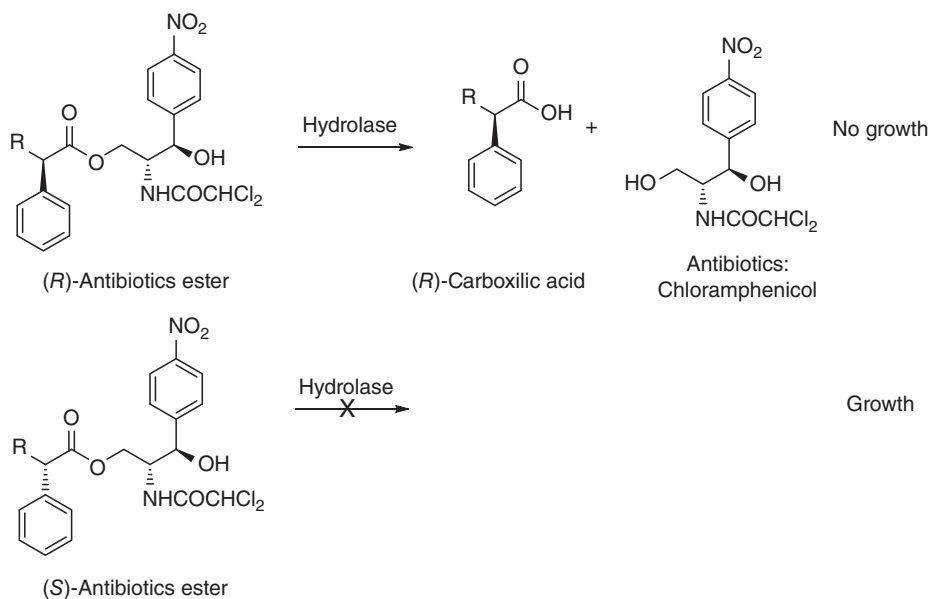
active epoxide hydrolase mutants. (Reetz and Wang [11]. Reproduced with permission of Bentham Science Publishers.)

growth advantage just because it harbors an enantioselective enzyme mutant? If this could be accomplished, then only those colonies harboring variants with enhanced enantioselectivity for a defined asymmetric reaction would appear on agar plates, which would be of tremendous advantage (Scheme 2.3).



Scheme 2.3 Genetic selection in the directed evolution of enantioselective enzymes [2]. (Acevedo-Rocha *et al.* [2]. Reproduced with permission of Elsevier.)

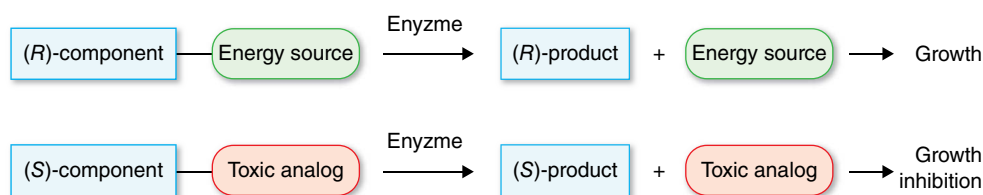
In order to identify enantioselective lipases, a system utilizing pro-antibiotic substrates has been devised [12]. Scheme 2.4 shows the general strategy, which involves monitoring the growth of *E. coli* or *Exiguobacterium acetylicum* cells during hydrolysis of (*R*)- and (*S*)-esters. The viability of this interesting concept was



Scheme 2.4 Concept of growth-based selection method employing pro-antibiotic substrates [12]. (Hwang *et al.* [12]. Reproduced with permission of Springer.)

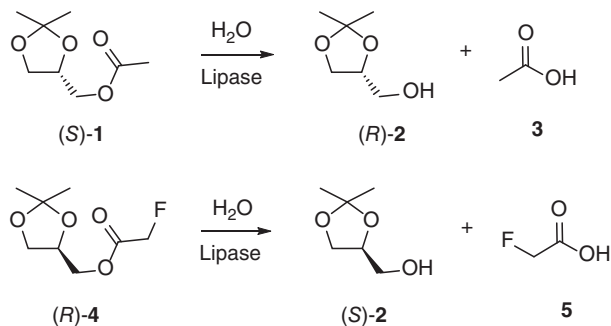
tested by studying several lipases, but it has not been employed in directed evolution of lipase mutants needed in stereoselective hydrolytic kinetic resolution. One problem that may be encountered in such a venture is the fact that surrogates are required as substrates, which would not be used in directed evolution aimed at real (industrial) applications (see discussion concerning surrogates as substrates in Section 2.2).

A different and conceptionally promising approach makes use of pseudo-enantiomers in which one of the pseudo-enantiomeric pair has an isosteric yet toxic component as shown in Scheme 2.5 [13]. It is based on the bond-breaking reaction of an appropriately designed substrate which in one enantiomeric form generates an energy source for the host organism (promotion of cell growth), while the mirror-image substrate constitutes an isosteric pseudo-enantiomer, which upon bond breakage releases a poison (inhibition of cell growth or cell death) (Scheme 2.5).



Scheme 2.5 A genetic selection system for directed evolution of enantioselective enzymes in kinetic resolution [13]. (Reetz *et al.* [13a]. Reproduced with permission of Royal Society of Chemistry.)

Since the number of synthetically interesting stereoselective bond-breaking reaction types is limited, the concept is far from general. Nevertheless, it was implemented experimentally in a proof-of-principle study in which the lipase-catalyzed hydrolytic kinetic resolution of an acetate derived from a chiral alcohol (acetic acid ester) was used as the model reaction [13]. Hydrolysis of both (*R*)- and (*S*)-substrate generates acetate, which in both cases serves as an energy source for the organism. Therefore, the host organism has no reason to prefer the hydrolysis of either one of the enantiomers. In order to construct selection pressure favoring one of the enantiomers, a pseudo-racemate was designed following the principle outlined in Scheme 2.5. The acetate and the sterically similar fluoro-acetate of isopropylidene glycerol (IPG; **2**) were first prepared separately, specifically (*S*)-**1** and (*R*)-**4** (Scheme 2.6). Lipase-catalyzed hydrolysis of (*S*)-**1** would provide acetic acid (**3**) as an energy source, while hydrolysis of the isosteric pseudo-enantiomer (*R*)-**4** would be expected to generate fluoro acetic acid (**5**) as a poison. The latter inhibits the acotinase step of the essential citric acid cycle. The lipase from *Candida antarctica* B (CALB) was chosen as the enzyme and *Pichia pastoris* as the host organism [13].



Scheme 2.6 Genetic selection system utilizing a pseudo-racemate (S)-1/(R)-4 in the CALB-catalyzed hydrolytic kinetic resolution [13]. (Note that the designation of absolute

configuration upon going from (S)-1 to (R)-2 or from (R)-4 to (S)-2 switches according to the priority rules of the CIP convention).

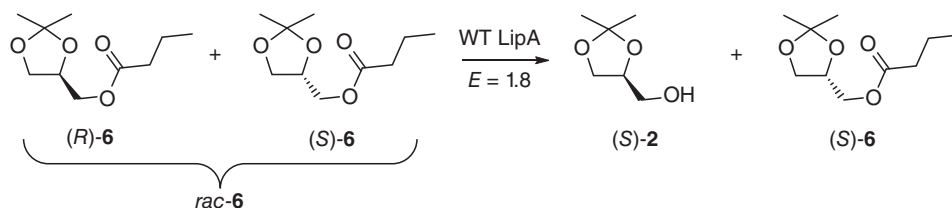
It was known that the conventional hydrolytic kinetic resolution using the traditional acetates of *rac*-1 slightly favors the formation of (S)-2 ($E = 1.9$). The experiment was designed to induce reversal of enantioselectivity as opposed to enhancing (S)-selectivity. Control experiments showed that neither of the two starting compounds themselves is toxic to the host organism. Upon optimizing this selection system, it was found that the pseudo-racemate need not consist exactly of a 1 : 1 mixture of the two pseudo-enantiomers. The use of too much of the fluoroacetate containing substrate (R)-4 leads to excessive formation of fluoroacetic acid (5), which causes undesired immediate cell death. Therefore, the amount of (R)-4 had to be decreased to an optimal level [13].

In exploratory experiments, the expected inhibition of growth by fluoroacetate (5) in the presence of acetate as the carbon source was first demonstrated, which indicated the viability of the concept [13]. Since problems associated with catabolite repression and background growth in the case of the methanol-inducible pPICZ α system could occur, the constitutive pGAPZ α was employed. Following a series of optimization experiments in liquid cultures and on solid plates, selection plates with 0.3% (17 mM) of the acetate (S)-1 and 0.003% of the fluoroacetate (R)-4 were found to be optimal. A minimum of directed evolution experiments were performed in this study, consisting of saturation mutagenesis at a two-residue site next to the CALB binding pocket at Leu278/Ala281 using NNK codon degeneracy encoding all 20 canonical amino acids. The fairly small library was spread out on an agarose plate followed by incubation. Approximately 70–80 colonies were observed. The 10 largest ones were harvested and the respective mutants characterized by sequence determination, which were then tested as catalysts in the hydrolytic kinetic resolution of the real racemic acetates (1 : 1 mixture of (S)-1 and (R)-1). Eight of the 10 variants led to the expected preferential reaction of (S)-1, which means reversal of enantioselectivity. Only one mutant showed a slight preference for the opposite enantiomer, while one proved to be essentially inactive. The percentage of false positives (20%) is low, which speaks for the viability of the selection system. The measured selectivity

of the (*S*)-selective variants ranged between $E = 3$ and $E = 8$, the double mutant Leu278Asp/Ala281Leu leading to the highest (*S*)-selectivity [13].

This study constitutes proof of principle, but it suffers from several drawbacks. Firstly, *P. pastoris* is probably not the optimal host organism. Secondly, the library was much too small; simultaneous saturation mutagenesis at, for example, a 10-residue site would ensure much greater structural diversity and most likely lead to variants showing considerably higher (*S*)-selectivity. Finally, it would be interesting to see how the analogous system performs in which the acetate and fluoro acetate are interchanged. Such a switch should provide colonies housing mainly CALB variants of opposite enantioselectivity.

A different genetic selection system for enhancing or inverting enantioselectivity of a lipase was published around the same time [14]. It is based on an alternative concept. In this case, *Bacillus subtilis* lipase A (LipA) was chosen as the catalyst in the hydrolytic kinetic resolution of *rac*-**6**, the butyrate of IPG (Scheme 2.7). As in the case of the acetate *rac*-**1** in the above study, this is a “difficult” substrate, WT LipA being only slightly (*R*)-selective ($E = 1.8$). The researchers also aimed for reversal of enantioselectivity with evolution of an (*S*)-selective variant [14].

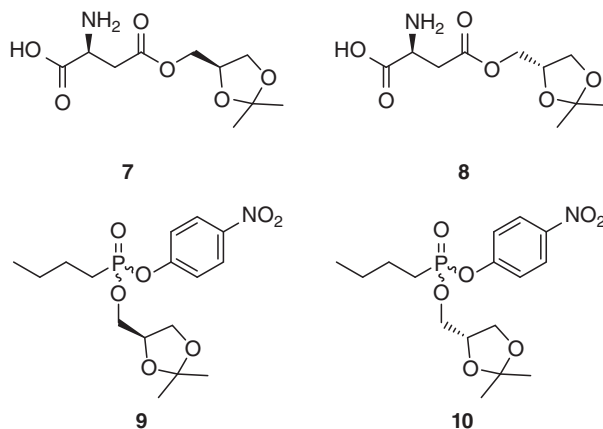


Scheme 2.7 LipA-catalyzed hydrolytic kinetic resolution of *rac*-**6** [14]. (Notice the switch in the designation of absolute configuration due to a change in priority

of the substituents in accord with the Cahn–Ingold–Prelog (CIP) nomenclature.) (Boersma *et al.* [14]. Reproduced with permission of John Wiley & Sons.)

In order to evolve (*S*)-selective mutants as catalysts in the model reaction of *rac*-**6**, the researchers developed a kind of a dual selection system, requiring the synthesis of aspartate esters of (*S*)- and (*R*)-**6** (compounds **7** and **8**, respectively) as well as LipA inhibitors **9** and **10** derived from (*S*)- and (*R*)-**2**, respectively (Scheme 2.8) [14].

A mutant library was first generated by saturation mutagenesis at a site comprising residues 132–136, which was then transformed into *E. coli* K-12 PA340/T6, this being a strain in which both pathways leading to the synthesis of aspartate have been blocked by conventional knock-out tools. Plating this strain onto selective minimal medium plates containing aspartate ester **7** was an essential part of the plan. Only those LipA mutants expressed in the periplasm capable of hydrolyzing this ester would be expected to liberate aspartate necessary for bacterial growth. The addition of phosphonate inhibitor **10** derived from the undesired IPG was expected to minimize the growth of bacteria that express less enantioselective variants [14].



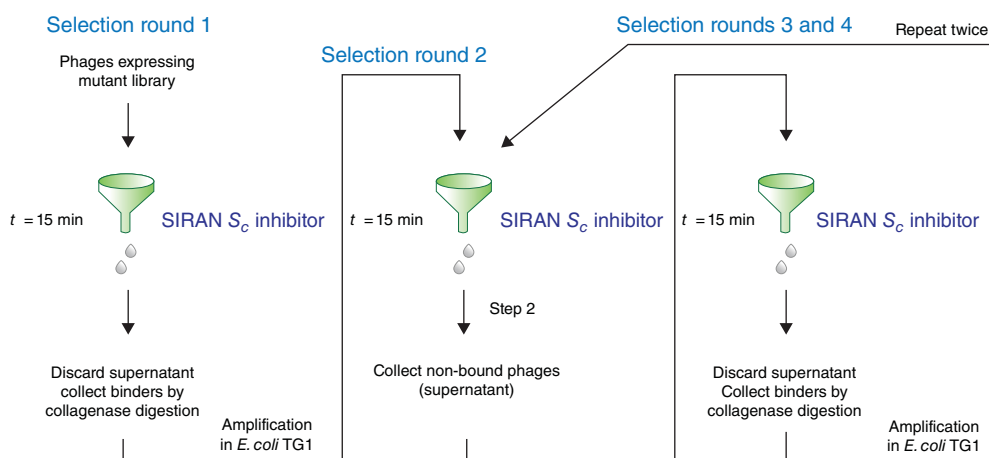
Scheme 2.8 Compounds used in devising a selection system for the evolution of *S*-selective LipA variants as catalysts in the hydrolytic kinetic resolution of *rac*-9 [14]. (Boersma *et al.* [14]. Reproduced with permission of John Wiley & Sons.)

About 1000 bacterial colonies appeared within two days in the first cycle. In the second cycle, an increased amount of phosphonate inhibitor was used, which led to 750 colonies. In each case, the 50 largest colonies were picked. Following two rounds, the most enantioselective variant was identified as Asp133Gly/Met134Leu/Ile135Asn, showing in a standard hydrolytic kinetic resolution using *rac*-6 as substrate a selectivity factor of $E = 12$ in favor of the desired (*S*)-2 with reversed absolute configuration [14]. Moreover, seven other variants favoring (*S*)-2 were also identified, although with lower E -values. As in the previous study [13], more research is needed to see how far enantioselectivity can be boosted. The fact that surrogates were used as substrates that would not be employed in practical applications is a disadvantage.

As already alluded to, the term “*selection*” has been used in different ways [2]. In addition to the strict sense of the word when considering systems in which cell growth and cell survival form the basis of selection (a terminology which the author prefers), the term is also used by many scientists when working with various systems in which proteins are displayed on the surface of an organism, phage display in combination with fluorescence-activated cell sorting (FACS) being a commonly used detection technique [1a, 15]. The display systems are particularly well suited for detecting proteins with specific binding properties (e.g., antibodies), but such systems are generally less reliable when considering catalysis that involves binding and turnover. Nevertheless, the advantage of display systems is evident, one reason being the enormous size of libraries ($\approx 10^{10}$) that can be rapidly screened. Several studies describe cell surface display of active enzymes [1a, 15b, 16].

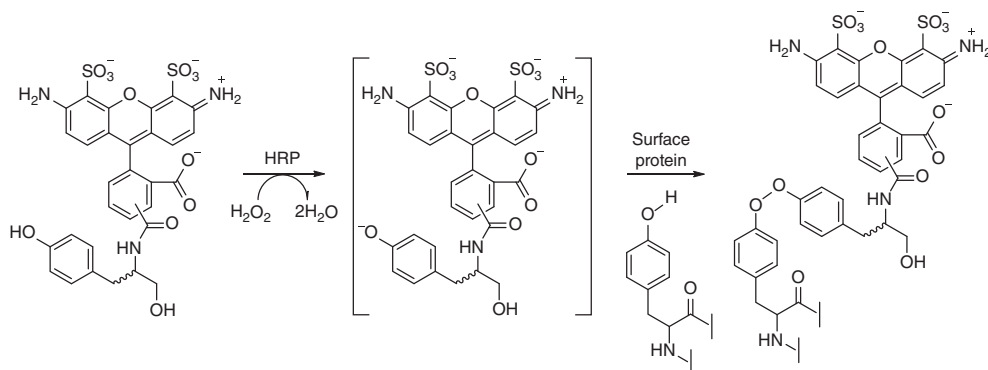
Few display systems have been employed in directed evolution for the purpose of enhancing or inverting enantioselectivity [15b]. The advantage is the possibility of generating and selecting (or is screening the better term?) extremely large

libraries which would be impossible to screen by conventional means. In an early example, phage display was employed in the attempt to evolve an enantioselective mutant of *B. subtilis* lipase (LipA) as the catalyst in the hydrolytic kinetic resolution of *rac*-6 [17]. It was an attempt to use the same concept that was employed in the dual selection approach featured above (Scheme 2.7). In the phage-display system, the same type of inhibitors were utilized, but these were attached covalently to a highly porous silicate carrier (SIRAN) (Scheme 2.9). As before, saturation mutagenesis was performed at large randomization sites. First, about 8.3×10^{10} phages were incubated with the phosphonate inhibitor immobilized on the solid carrier SIRAN and allowed to react for 15 min. Panning and digestion followed as illustrated in Scheme 2.9, leading to the isolation of a LipA mutant showing inverted enantioselectivity (33% *ee* at 32% conversion) [17b]. Inversion of enantioselectivity corresponds to the design of the experimental setup, but the degree of stereoselectivity remains meager. Several reasons for this somewhat disappointing result were postulated, the most likely being the fact that phage display is based on (selective) complexation, but not on catalysis [17b].



Scheme 2.9 Dual selection system based on the use of phosphonate inhibitors immobilized on the solid carrier SIRAN [17b]. (Dröge *et al.* [17b]. Reproduced with permission of John Wiley & Sons.)

In a different and more successful approach, yeast surface display employing *Saccharomyces cerevisiae* was used in order to evolve horseradish peroxidase (HRP) as a catalyst in the enantioselective oxidative reaction of L- and D-tyrosinol [18]. The known concept of covalent attachment of products via enzyme display for the purpose of assessing activity was utilized. Labeling the absolute configuration of the substrate by a fluorescent dye (Alexa 488) was performed using radical chemistry (Scheme 2.10). Two 2×10^6 -sized libraries were displayed and subjected to FACS analysis. The first library was generated by extensive



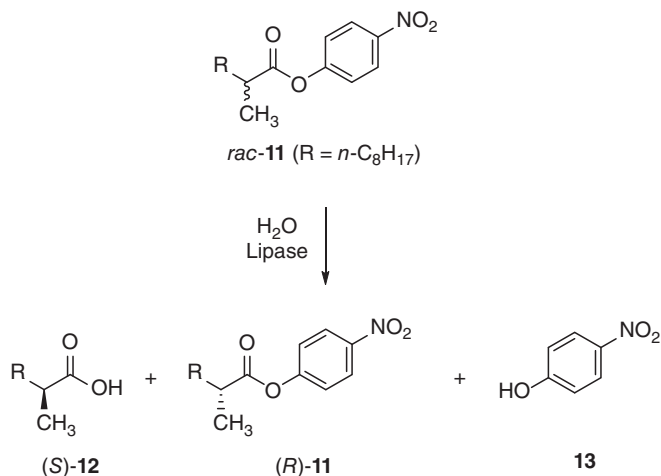
Scheme 2.10 HRP-catalyzed radical polymerization of L- or D-tyrosine and Alexa-488 derivatives [18]. (Lipovsek *et al.* [18]. Reproduced with permission of John Wiley & Sons.)

epPCR, the other by simultaneous saturation mutagenesis at a randomization site composed of five amino acid positions near the active site. In the case of the saturation mutagenesis library at sites lining the binding pocket (CASTing), the combination of positive and negative selection led to the identification of mutants showing a four- to eightfold enhanced enantioselectivity for the D- and the L-substrate, respectively. The analogous strategy using the epPCR library failed because no stereoselective mutants could be identified. The authors ascribe this failure to the fact that in epPCR mutations at multiple residues all near the active site are statistically unlikely, among other factors [18].

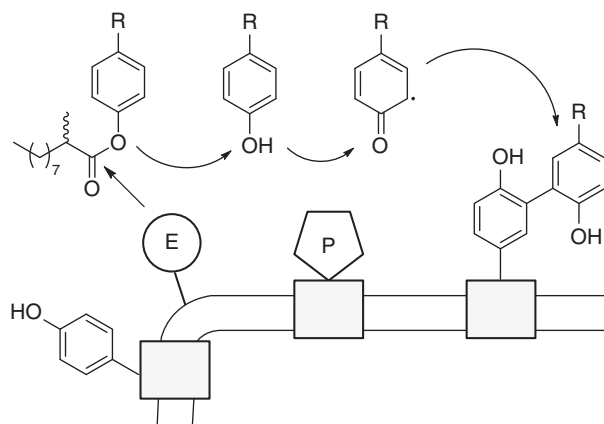
This experimental platform was designed in a logical way, which led to intriguing results. It represents the first case of directed evolution using a eukaryotic organism for displaying an enzyme under this kind of “selection pressure” for stereoselectivity. It was suggested that the scope of enzyme yeast surface display could be extended by employing “any nontoxic substrate that can be conjugated to a standard linker” [18]. In order to generalize this interesting concept, more research is necessary.

In a proof-of-concepts study, single-cell FACS-based high-throughput identification of enantioselective enzymes was demonstrated [19]. The esterase from *Pseudomonas aeruginosa* (EstA) served as the catalyst in the hydrolytic kinetic resolution of the chiral ester derived from 2-methyldecanoic acid, (*R*)- and (*S*)-2-MDA (*R*)- and (*S*)-**11** (Scheme 2.11). This is the same transformation that was used in the first example of directed evolution of an enantioselective enzyme, at that time employing a lipase (Chapter 1) [20].

Control experiments showed that WT EstA is a poor catalyst in the hydrolytic kinetic resolution of *rac*-**11** ($E = 1.2$ in favor of (*S*)-**12**). The two enantiomeric substrates were each labeled with a different fluorescent dye (green and red, respectively). This allowed 10^8 cells to be assayed individually within a few hours. Several events occur during the analytical process: the tyramide ester substrates undergo hydrolysis and peroxidase-mediated formation of short-lived radicals ensures the



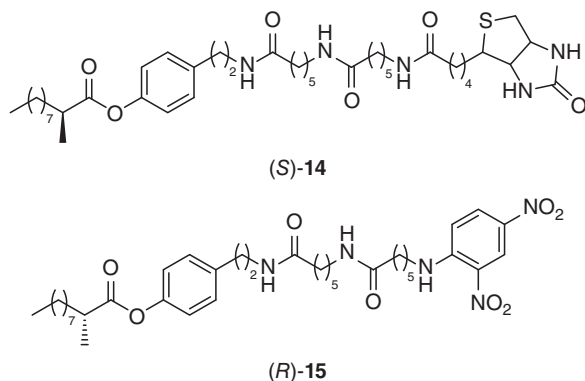
Scheme 2.11 Model hydrolytic kinetic resolution catalyzed by the esterase EstA and used in FACS-based assessment of enantioselectivity [19]. (Becker *et al.* [19]. Reproduced with permission of John Wiley & Sons.)



Scheme 2.12 Schematic representation of coupling reactions ensuring covalent attachment of tyramide species on the surface of *E. coli* cells [19]; E, esterase; P, peroxidase. (Becker *et al.* [19]. Reproduced with permission of John Wiley & Sons.)

immediate covalent attachment of reaction products to the surface of the esterase-proficient bacterial cell (Scheme 2.12) [19]. Since the attached products are fluorescence labeled, cells harboring reactive enzymes can be identified by FACS.

In order to address enantioselectivity, two different pseudo-enantiomeric substrates (*S*)-14 and (*R*)-15 were used, labeled with indicator groups 2,4-dinitrophenyl (DNP) and biotin, respectively (Scheme 2.13). The experimental setup was designed so that hydrolysis triggers green or red fluorescence signaling (*R*)- and (*S*)-selectivity, respectively.



Scheme 2.13 Differentially labeled enantiomers of 2-MDA tyramide ester substrates (S)-14 and (R)-15 [19]. (Becker *et al.* [19]. Reproduced with permission of John Wiley & Sons.)

The obligatory cell labeling was performed individually by conjugation of HRP (P) to the cell surface. In the case of DNP tyramide labeling, green fluorescent detection was ensured with an Alexa Fluor 488-labeled antibody, whereas biotin tyramide deposition was detected by streptavidin (*R*)-phycoerythrin conjugate (red fluorescence). Flow cytometry distinguished between individually labeled cells. Typical examples of FACS histograms are displayed in Figure 2.2 [19].

More than 10^7 clones generated by epPCR were rapidly analyzed by FACS. After applying the peroxidase, the search for (*R*)-selective mutants in the large library was initiated by incubation with a 1:1 mixture of (S)-14 and (R)-15. Out of 6.8×10^7 FACS-analyzed cells, only 86 active clones were obtained. The respective *E*-values were determined for 35 mutants by kinetics under standard conditions using (S)- and (R)-11 (Scheme 2.11), respectively. Most of the variants showed moderate to poor enantioselectivity, the best one being a triple mutant W185R/G224D/G263S with *E* = 16.3 in favor of the (R)-12 as determined in separate experiments [19]. This means reversal of enantioselectivity. A few variants were identified showing very slight enhancement of (*S*)-selectivity [19].

It can be speculated that the somewhat moderate results in terms of stereoselectivity may be due to the sub-optimal choice of the mutagenesis method (epPCR). Indeed, in the original directed evolution study using the same substrate (*rac*-11) where a lipase was used instead of an esterase, multiple cycles of epPCR also delivered a mutant showing only moderate enantioselectivity (*E* = 11) [20a]. Subsequent experiments using saturation mutagenesis at sites surrounding the binding pocket proved to be much more efficient (*E* = 594 in favor of (S)-12) [20b]. Simultaneous saturation mutagenesis at sites next to the binding pocket comprising 10 amino acid positions using NNK codon degeneracy encoding all 20 canonical amino acids could be even better if combined with FACS-based flow cytometry. This would cover an almost astronomically large protein sequence space ($\approx 10^{15}$ variants), which cannot be assessed by any other screening method.

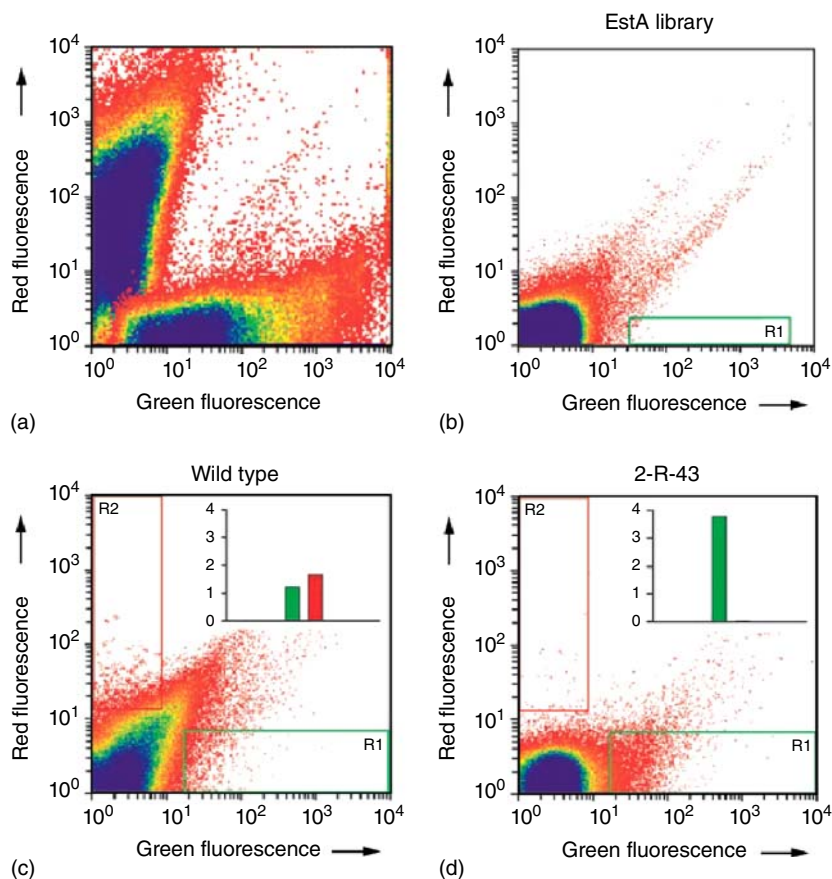
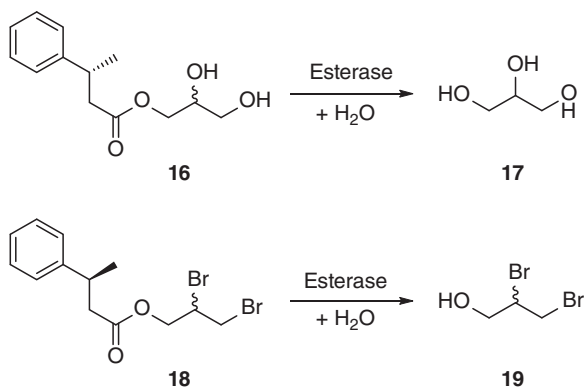


Figure 2.2 FACS-based high-throughput analysis of enantioselectivity [19]. (a) Overlay of flow-cytometry analyses of esterase-displaying cells that were incubated for 60 min with either (S)- or (R)-enantiomer of tyramide ester. (b) EstA library sort. The green window indicates the sorting gate. (c, d) FACS histogram of WT EstA (c) and

clone 2-R-43 (d) after 5 min incubation with a 1:1 mixture of both enantiomeric substrates and fluorescence staining. The inset shows the percentage of cells within the respective green and red gate. (Becker *et al.* [19]. Reproduced with permission of John Wiley & Sons.)

In fact, such experiments would be limited by the amount of available DNA. Both (R)- and (S)-selective mutants would then be expected to occur in such a large library.

Another study [21] describes the combination of *in vivo* selection along the lines of the earlier system in which one enantiomer generates a cell-killing product [13], FACS-based screening being applied as in a different investigation [19]. The model system involves esterase-catalyzed hydrolytic kinetic resolution of the pseudo-racemate comprising ester **16** with (R)-configuration in the acid part and ester **18** with the respective opposite (S)-configuration (Scheme 2.14) [21]. Hydrolysis



Scheme 2.14 Hydrolytic kinetic resolution utilizing the pseudo-racemate (*R*)-**16**/*S*)-**18** catalyzed by esterase PFE in which an energy source (**17**) and a cell poison (dibromide **19**) are generated [21]. (Fernandez-Alvaro *et al.* [21]. Reproduced with permission of John Wiley & Sons.)

of **16** releases glycerol (**17**), which is an energy source for the host bacterium *E. coli*, whereas hydrolysis of the pseudo-enantiomer **18** generates the dibromide **19**, which is a poison that causes cell death. Preliminary control experiments of the growth behavior of *E. coli* in a medium supplemented with mixtures of the two substrates in the presence of esterases BS2, PestE, or CL1 confirmed the working hypothesis. In the final experiments for demonstrating the concept, a different esterase was used (*Pseudomonas fluorescens* esterase (PFE)) [21].

PFE was known to catalyze the hydrolytic kinetic resolution of the ethyl ester analog of **16/18** with low enantioselectivity in slight favor of the (*R*)-substrate ($E = 3.5$). In an earlier study the authors had used this substrate to improve enantioselectivity up to $E = 80$ by employing saturation mutagenesis at sites W28, V121, F198, and V225 next to the binding pocket (CASTing), the choice of the amino acids in a reduced alphabet being guided by bioinformatics data [21]. Conserved amino acid substitutions were identified by standard sequence alignment [22] followed by choosing the respective highly reduced amino acid alphabets [21]: For W2 (A, V, L, F, W, S, G, C); for V121 (A, V, I, S, T, G); for F198 (V, L, F, W, G, C); for V225 (A, V, I, L, S, T). For a discussion of the use of reduced amino acid alphabets [23], see Chapter 3. It was clear that the use of these reduced amino acid alphabets requires the screening of only $\approx 10\,000$ transformants for 95% library coverage, whereas NNK codon degeneracy encoding all 20 canonical amino acids would involve much more screening for 95% library coverage ($\approx 3 \cdot 10^6$ transformants).

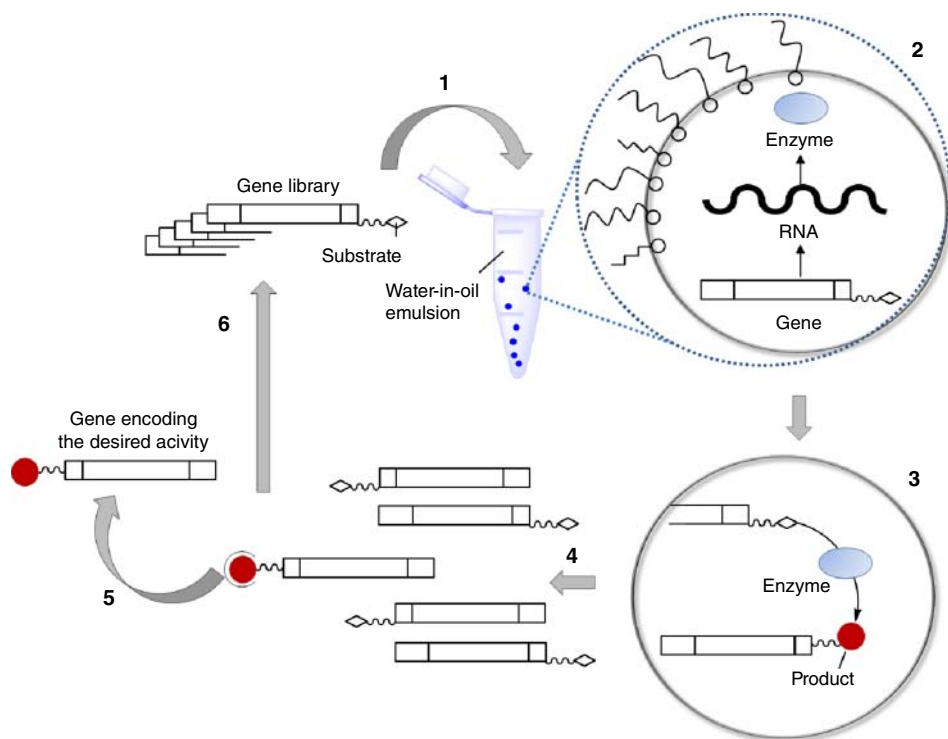
When employing FACS-based directed evolution, such numbers do not impose any restrictions. Nevertheless, the researchers employed their previously designed relatively small library and performed FACS-based cell sorting on it [21]. Those cells that were expected to harbor (*R*)-selective variants were spread on pretreated agar plates. It was discovered that 28 clones were able to grow on the (*R*)-**16**/*S*)-**18** pair, while only a few proved to be false positives. Variants showing

enhanced enantioselectivity in the hydrolytic kinetic resolution of the surrogate substrate *rac*-**16** were identified with selectivity factors in the moderate range of $E = 4$ –25. Upon testing these mutants in the hydrolytic kinetic resolution of the ethyl ester of 3-phenylbutyric acid, several mutants were found that had also been identified in the library generated previously using bioinformatics-guided saturation mutagenesis, including variant V121I/F198C ($E = 80$). An even better variant V121S/F198G/V225A led to a selectivity factor of $E \approx 100$ [21].

This investigation demonstrates that the combination of selection and FACS-based cell sorting reproduces many of the results observed previously by standard saturation mutagenesis at the binding pocket using highly reduced amino acid alphabets [21]. An even better variant was evolved. Currently, it is not clear how this experimental platform would fare if a much larger protein sequence space were to be studied with considerably higher genetic diversity, rather than restricting the search to a rather small previously designed library. It remains to be seen if the selection part of the process can be used alone without FACS-based screening, which would correspond to the selection principle outlined in Scheme 2.1.

Various *in vitro* methods have also been developed [1, 2], including lysate assays and ribosome display [24], and especially systems based on *in vitro* compartmentalization (IVC) [3c, 16d]. In the latter system, originally reported in 1998 [25], the addition of an *in vitro* transcription/translation reaction mixture to stirred mineral oil harboring appropriate surfactants such as Span80 and Tween80 generates an emulsion with particles having a mean droplet diameter of about 2.6 μm comparable in size to *E. coli* cells (Scheme 2.15). The translation of several proteins proved to be successful in such droplets, including dihydrofolate reductase and the DNA methyltransferase *Hae*III.

Since this initial report of IVC [25], the concept has been extended to include directed evolution of proteins and RNAs for catalytic and binding properties as well as regulatory activities [3c, 16d]. Due to the polydispersity of the emulsions, which makes quantitative assessments difficult and because the addition of new reagents to pre-formed droplets is not trivial, the general use of the technology was somewhat limited. However, with the extension to droplet-based microfluidic devices, these problems were largely solved [26]. Prominent recent examples include a high-throughput cellulase screening system for activity [27a] and a highly parallel microfluidic droplet device that enables single-molecule counting for digital enzyme detection [27b], among other developments [27c–g]. An ultrahigh-throughput technique for mapping enzyme sequence–function relationships has been devised by combining droplet microfluidic screening with next-generation DNA sequencing [27g]. This approach was used to identify glycosidase variants displaying higher activity and stability. All members of a library resulting from deep mutational scanning [28] were assayed in aqueous microdroplets, and sorted and unsorted variant pools were subsequently characterized by high-throughput DNA sequencing (Figure 2.3) [27g]. An in-house constructed microfluidic droplet sorter was used which requires specialized instrumentation.



Scheme 2.15 Illustration of gene selection in compartmentalized oil-in-droplet emulsions [25]. (Tawfik and Griffiths [25]. Reproduced with permission of Nature Publishing Group.)

Other major applications of these microfluidic devices involve physiologically relevant cell-based assays (e.g., proliferation of human cells) [29], genetic analysis, and retroviral display using droplet-based microfluidics [30a] It will be interesting to see how the droplet-based technologies perform when applied to the directed evolution of stereo- and regioselective enzymes as catalysts in organic chemistry and biotechnology. It should be pointed out that microfluidic devices can also be used in screening as opposed to selection [30b].

In conclusion, a number of seminal studies have appeared describing selection systems (including display systems) for directed evolution of enantioselective enzymes, while in other cases the respective techniques have not yet been tested for this purpose. While the different approaches and concepts are clear, the degree of success, in most cases, remains limited when applied to the directed evolution of enantio-, diastereo-, and regioselective enzymes. Screening systems (see subsequent text) have been considerably more successful in these important areas of application. Nevertheless, further efforts and ideas in the intriguing research area of selection-based or microfluidics-based laboratory evolution may lead to new assessments.

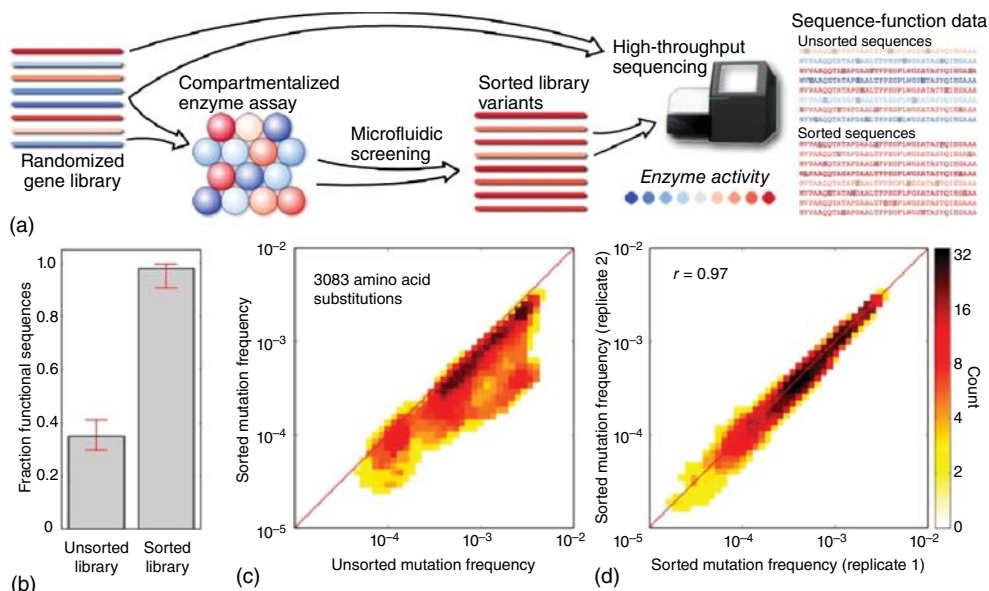


Figure 2.3 Ultrahigh-throughput sequence-function mapping based on deep mutational scanning [27g]. (a) General overview of mapping protocol; (b) droplet-based microfluidic screening leading to the recovery of functional sequences from the initial random mutagenesis library; (c) frequency of 3083 amino acid exchanges in the

unsorted and sorted glycosidase libraries; and (d) reproducibility of the sequence-function mapping protocol with two independent replicates showing good agreement in amino acid frequencies. (Romero *et al.* [27g]. Reproduced with permission of National Academy of Sciences.)

2.2

Screening Systems

A monograph published in 2006 provides a comprehensive survey of screening systems used in the directed evolution of activity, substrate scope, and stereoselectivity [1a], which was later augmented by review articles [1b–e, 2]. Since then a number of innovations have emerged, which have been summarized in a 2015 review [1f]. Only the basic principles are reiterated here, illustrated by selected examples. The advent of directed evolution of stereoselective enzymes included the first medium-throughput ee-assay [20a] and sparked research directed toward developing further and more efficient medium- and high-throughput ee-screening systems [1, 2, 31]. In the original study and subsequent reports [20] involving a lipase-catalyzed kinetic resolution of a racemic *p*-nitrophenyl ester (*rac*-**11**; Scheme 2.11), a UV/vis plate reader was used to monitor the formation of *p*-nitrophenolate **13** (yellow color) as a function of time (typically during the first 8 min). It is highlighted here because the underlying principle forms the basis of many other assays needed for different enzyme types. Measuring the time-dependent formation of **13** released from the racemic ester provides

information only of the approximate overall rate. Consequently, the (*R*)- and (*S*)-substrates were prepared separately in enantiomerically pure form and then studied pairwise on 96-well microtiter plates, allowing 48 mutants to be assayed for enantioselectivity on one plate. Whenever the slopes of the absorption/time curves differ considerably, a hit is indicated, which is subsequently used as the catalyst in the process of measuring the kinetics of the real hydrolytic kinetic resolution of the racemic *p*-nitrophenyl ester (*rac*-**11**). This allows the exact determination of the selectivity factor *E* (relative rate of reaction of one enantiomer relative to that of its mirror image). Figure 2.4 shows the respective curves of the WT lipase and a typical variant displaying enhanced enantioselectivity [20a].

The measurement of about 500–800 plots of this kind is possible per day. However, this does not have to be performed for all of the transformants, because an on-plate pre-test for activity can be carried out (Figure 2.5). Only the active clones are subsequently studied by the UV/vis assay or directly by kinetics. The agar plate contains tributyrin, which on hydrolysis forms a halo that can be easily identified by the human eye, a classical test used to assess lipase or esterase activity, which was simply adapted for high-throughput analysis [20, 31]. The so-called casein-test for identifying active proteases is also based on the appearance of halos on casein-spiked agar plates [32].

Measuring the initial rates of reaction of (*R*)- and (*S*)-enantiomers separately in the wells of microtiter plates forms the basis of many other ee-assays

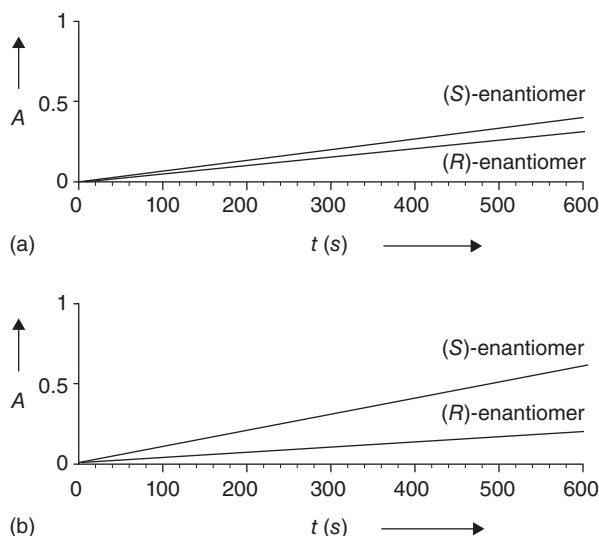


Figure 2.4 (a) Time course of the lipase-catalyzed hydrolysis of two enantiomeric *p*-nitrophenyl esters (*S*)- and (*R*)-**11** (Scheme 2.11) separately using WT enzyme with poor stereoselectivity. (b) Time course of the lipase-catalyzed hydrolysis of the

two enantiomeric *p*-nitrophenyl esters **11** using an enzyme variant with enhanced (*S*)-enantioselectivity [20a]. (Reetz *et al.* [20a]. Reproduced with permission of John Wiley & Sons.)

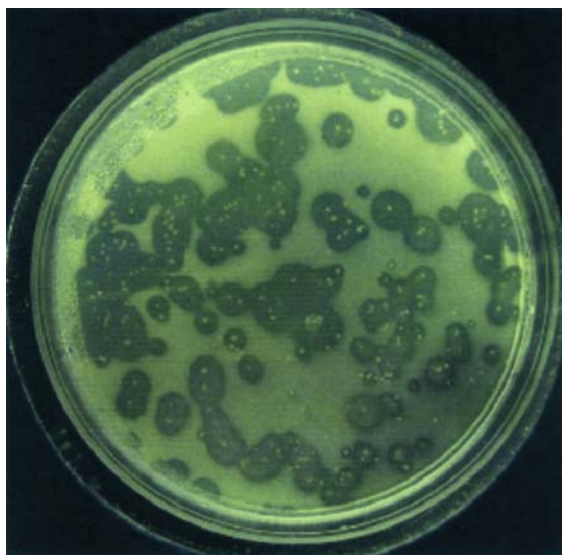
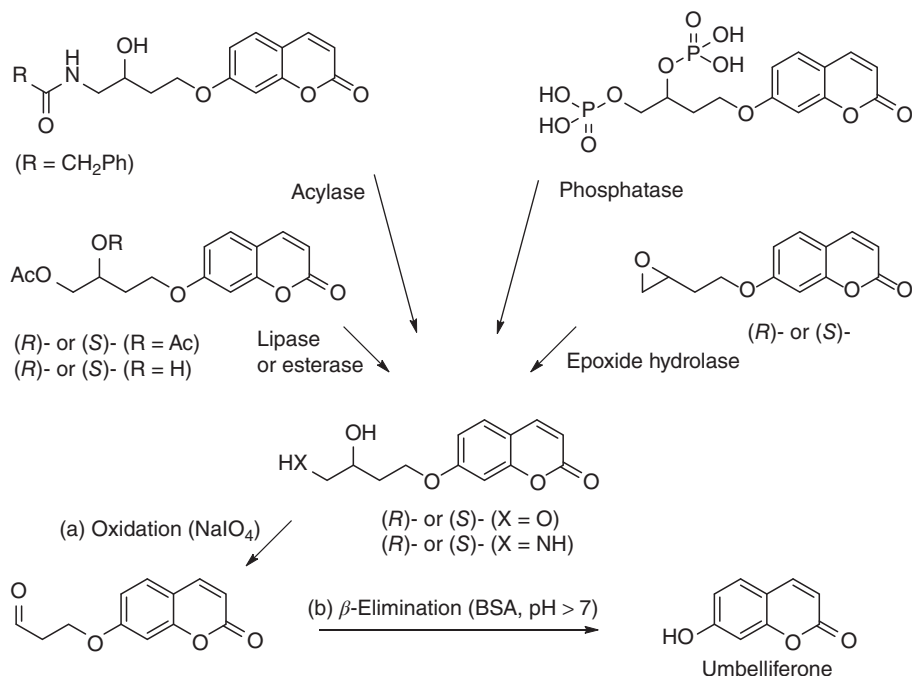


Figure 2.5 On-plate pre-test for lipase activity based on halos that form upon hydrolysis of tributyrin [31]. White dots represent bacterial colonies harboring active lipases; those

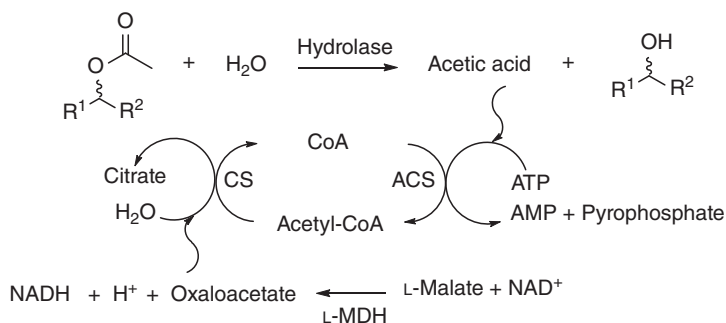
having no (clear) black background contain inactive mutants. (Reetz [31]. Reproduced with permission of John Wiley & Sons.)

[1, 20, 31]. The so-called umbelliferone-based fluorescence assay is an example, which can also be used for activity determination [1a]. In one embodiment, a periodate-coupled fluorogenic assay was used specifically for hydrolases such as lipases, esterases, and epoxide hydrolases (Scheme 2.16) [33].

If enantioselectivity is not involved, that is, whenever mutant libraries need to be screened for activity only, these and similar UV/vis- or fluorescence-based screens can of course be applied. However, all of them suffer from the disadvantage that surrogates are involved that are spectroscopically active when reacting. Unfortunately, the mutants evolved for these substrates may well show different selectivity and activity when the corresponding “real” substrates are employed for industrial applications. For example, a chiral ethyl ester is more likely to be utilized industrially than the respective *p*-nitrophenyl ester. For this reason, alternative assays are often preferred. For illustrative purposes, an enzyme-coupled UV/vis-based assay for assessing the activity of lipases or esterases is featured here, a measure of enantioselectivity also being possible by using “real” (*R*)- and (*S*)-substrates separately (Scheme 2.17) [34a]. Lipase- or esterase-catalyzed hydrolysis forms acetic acid which is converted by acetyl-coenzyme A synthetase (ACS) to acetyl-CoA in the presence of ATP and coenzyme A (CoA). Citrate synthase (CS) is used to catalyze the reaction between acetyl-CoA and oxaloacetate leading citrate. The required oxaloacetate is generated from L-malate and NAD⁺ in the presence of L-malate dehydrogenase (L-MDH). In this way, the initial rates of acetic acid formation are



Scheme 2.16 Periodate-coupled fluorogenic assay designed for assessing the enantioselectivity of hydrolases [33]. (Badalassi *et al.* [33]. Reproduced with permission of John Wiley & Sons.)



Scheme 2.17 Enzyme-coupled assay for assessing the activity of lipases or esterases, measurement of the apparent enantioselectivity also being possible when using

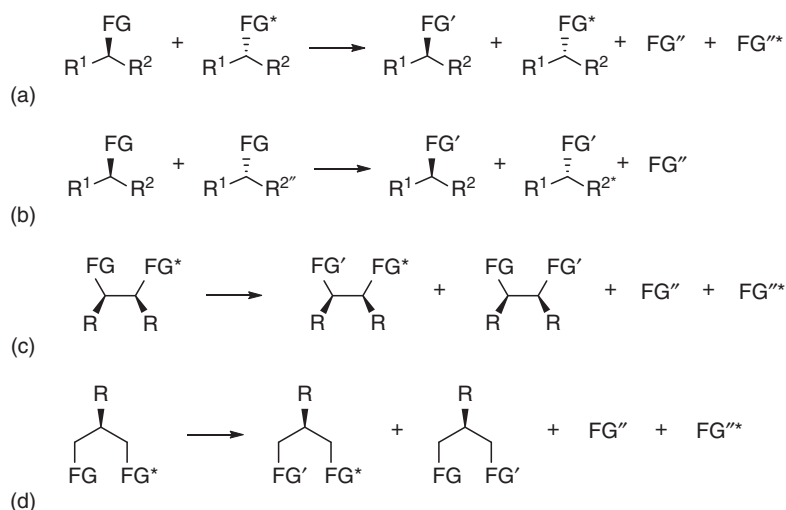
(R)- and (S)-substrates separately [34a]. (Baumann *et al.* [34a]. Reproduced with permission of John Wiley & Sons.)

accessible by monitoring the increase in absorption at 340 nm due to the increase in NADH concentration.

A different high-throughput ee-screen likewise utilizing enzyme-coupled reactions for the assay can be applied to any enzyme-catalyzed stereoselective

formation of chiral alcohols (e.g., ADH reduction of prochiral ketones, P450-catalyzed oxidative hydroxylation of achiral compounds, and lipase- or esterase catalyzed hydrolysis of esters) [34b]. Further screens that also avoid the use of surrogates have been developed [1], for example, those that measure the pH change in a lipase- or esterase-catalyzed reaction by means of appropriate dye indicators [35].

Alternative approaches utilize isotopically labeled pseudo-enantiomers in hydrolytic or oxidative kinetic resolution or (pseudo) meso substrates in hydrolytic or oxidative desymmetrization (Scheme 2.18) [36–38]. For labeling, deuterium, ^{13}C or ^{15}N isotopes have been used, quantitative high-throughput detection of enantioselectivity being possible by applying multiplexing MS [36a–d], IR-spectroscopy which is very effective while requiring a relatively cheap instrument [37], and even fast NMR-spectroscopy [38]. For example, in the industrial directed evolution of a nitrilase as a catalyst in the desymmetrization of 1,3-dicyano-3-hydroxypropane needed in the chemoenzymatic synthesis of the cholesterol-lowering therapeutic drug Lipitor[®], one of the cyano-groups was ^{15}C -labeled, enantioselectivity being measured by multiplexing MS, which allows thousands of samples to be analyzed in one day [36d]. As a note of caution, not



Scheme 2.18 Utilization of isotopically labeled pseudo-enantiomers in high-throughput screening of mutant libraries generated by directed evolution [36]. (a) Asymmetric transformation of a mixture of pseudo-enantiomers involving cleavage of the functional groups FG and labeled functional groups FG*. (b) Asymmetric transformation of a mixture of pseudo-enantiomers involving either cleavage or bond formation

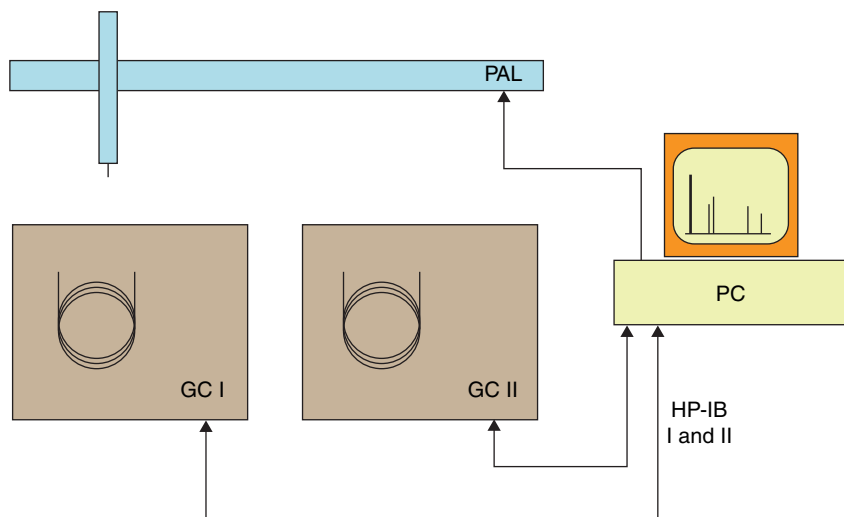
at the functional group FG; isotopic labeling at R2 is indicated by the asterisk. (c) Asymmetric transformation of a pseudo-meso substrate involving cleavage of the functional groups FG and labeled functional groups FG*. (d) Asymmetric transformation of a pseudo-prochiral substrate involving cleavage of the functional group FG and labeled functional group FG*.

all asymmetric processes can be detected by these otherwise efficient screens, for example, enantioselective reduction of prochiral ketones. In contrast to the IR-based system, the multiplexing MS instrument is currently too expensive for most academic laboratories.

High-throughput screening methods have also been developed for oxidoreductases, the majority of which require cofactors NAD(P)⁺ and NAD(P)H [39]. However, the absorbance of NAD(P)H at 340 nm cannot generally be used to monitor its formation or depletion as a function of time in a fully reliable manner due to background signals originating from cell lysates. Therefore, the redox reactions need to be coupled with a dye-forming reaction, of which various types have been developed [39]. Focusing on the oxidized form NAD(P)⁺ is an alternative in some cases, absorption at 580 nm being monitored by a UV/vis-plate reader. In this case, an agar plate pre-test based on the identification of white spots on a purple background is useful for spotting active clones. In special applications, other tests are called for, as in the case of P450-catalyzed terminal oxidative hydroxylation of linear alkanes in which *p*-nitrophenyl ethers are used as surrogate substrates that generate the yellow *p*-nitrophenolate for straightforward monitoring by a plate reader [39b]. Among other recent developments, several convenient colorimetric screening assays have been developed for application in the directed evolution of fungal laccases as catalysts in the conversion of plant biomass [39c]. Other examples include a high-throughput color-based screening system for assaying the activity of terpene-synthetases in cyclization reactions [39d], and a microplate assay for real-time screening of aldolases [39e].

For decades chiral GC and HPLC were considered to be slow processes. However, especially with the advent of directed evolution, the need for high- or at least medium-throughput arose. The first attempt to implement medium-throughput ee-determination in the directed evolution of an enantioselective enzyme (lipase from *Pseudomonas aeruginosa*, PAL) utilized two GC instruments in one unit (Scheme 2.19), rather than two columns in one unit [40]. This saves space, but in many laboratories this may not be an issue. Today, automated GC instruments for handling samples taken directly from microtiter plates, including washing steps, are commercially available at costs less than \$50 000. One of recent numerous examples pertains to the directed evolution of the haloalkane dehalogenase DhaH-31 [41]. Depending upon the nature of the product/starting material, about 300–700 samples can be handled per day [40, 41]. Automated HPLC has been used in a similar medium-throughput manner [42]. The respective throughputs do not seem to be very high, but in view of the fact that methodology development in directed evolution has enabled the generation of small libraries of increasingly higher quality, these screening systems suffice (Section 2.3).

A two-step screening strategy can be recommended for future directed evolution studies of stereo- and/or regioselective enzymes: Identifying mutants using an appropriate on-plate pretest for activity, and then subjecting the hits to automated GC or HPLC analysis. In addition to the tributyrin test as illustrated in Figure 2.5, a number of other on-plate assays for various enzyme types have been

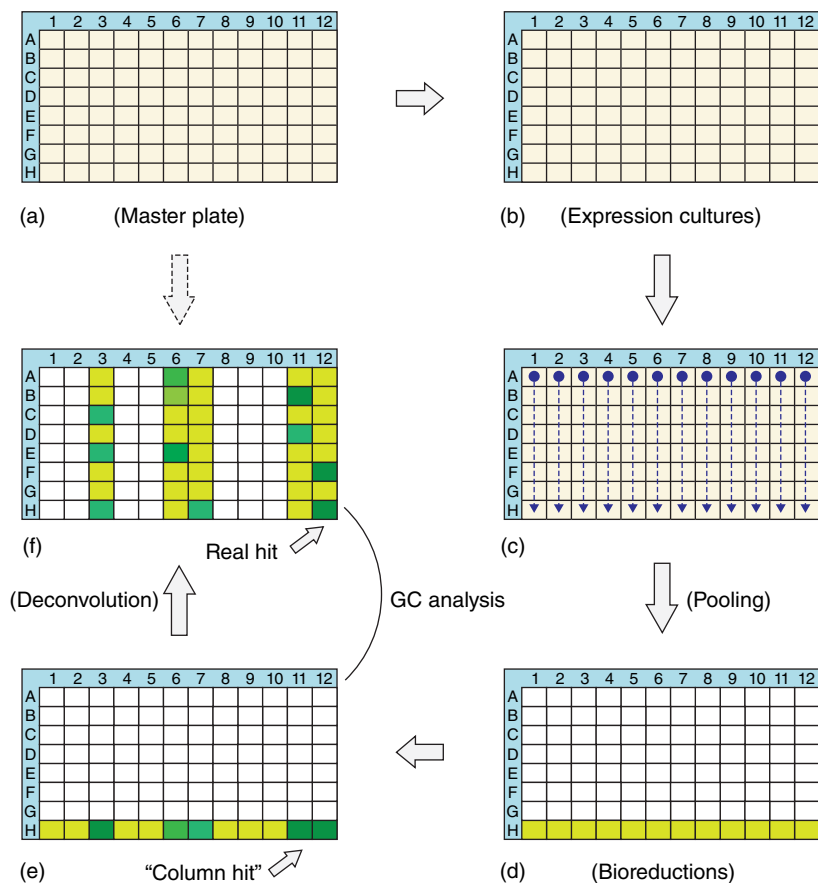


Scheme 2.19 Medium-throughput unit containing two GC instruments and one PC used in screening PAL-mutants generated by directed evolution for enhanced enantioselectivity [40]. (Reetz *et al.* [40]. Reproduced with permission of Elsevier.)

developed [1, 2, 31]. However, if very high-quality library design is ensured by the proper choice of mutagenesis techniques, then taking samples directly from the wells of microtiter plates and analyzing them by automated GC or HPLC constitutes the better strategy because it avoids the “uncertainty” accompanying on-plate tests (different expression rate of mutants, etc.). The same applies to laboratory evolution directed toward expanding enzyme substrate scope or thermostability.

In many cases, the concept of pooling mutant libraries offers another way to reduce the overall screening effort to a notable extent [43]. An early example pertains to the directed evolution of the enoate reductase YqjM as a catalyst in the enantioselective conjugate reduction of three substituted cyclohexenone derivatives. Careful analysis of the pooling/screening results demonstrated that the screening effort can be reduced by a factor of at least two (Scheme 2.20) [43b].

In a novel approach, fluorescence was used for designing an aptamer-based biosensor for high-throughput ee-determination, but application in directed evolution needs to be demonstrated [44]. A 2015 review of high-throughput screening in protein engineering covers the most recent developments including a list of FACS-based screening platforms and microfluidic assays [1f]. Microfluidic screening as part of lab-on-a-chip is indeed rapidly emerging for different areas of potential industrial application, including directed evolution [26–30]. For example, integrated microfluidic chips have been used in enzyme-catalyzed asymmetric reactions and electrophoresis-based chiral analysis at a nanoliter scale in dimensions of time and space by employing only a few hundred cells [30b]. Parallelization still needs to be implemented.



Scheme 2.20 General protocol for screening by pooling defined-cell cultures overexpressing enzyme variants [43b]. (a) Pick and inoculate individual colonies. (b) Induce expression of variants. (c) Recover by centrifuging individual cell pellets and combine all eight cell pellets belonging to the same column. (d) Lyse cells and incubate 12 biotransformations per plate by adding the appropriate reagents. (e) Extract product with organic

solvent and analyze organic layer by GC; hits will be identified in this step by setting an appropriate threshold. (f) Using as a reference the plate from step (e) return to master plate in step (a) and re-examine the columns of interest, if there is any, by adding reagents to each well separately. (Bougioukou *et al.* [43b]. Reproduced with permission of John Wiley & Sons.)

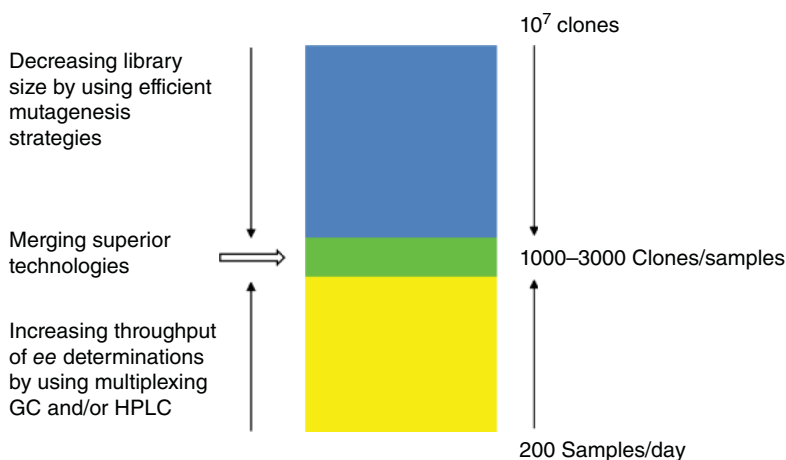
Along a different line, rapid exploration of the substrate scope of newly discovered enzymes or mutants generated by directed evolution is also a crucial goal. For this purpose cluster-screening has been developed, as in the case of new cytochrome P450 monooxygenases (CYPs) [45]. A library of 51 compounds to be tested with two different enzymes of this type (CYP 154E1) and (CYP 154A8) was organized into nine groups according to their structural properties. Following

CYP-catalyzed oxidation automated GC/MS was applied, leading to useful and practical data.

2.3

Conclusions and Perspectives

The optimal choice of a selection or screening system depends upon the particular goal of a directed evolution project. On-plate pre-tests for activity are always recommended. When aiming for improved or reversed stereo- and/or regioselectivity, screening assays as opposed to selection platforms are currently the only practical option. Since in these cases screening is the bottleneck of the whole process, much effort has gone into developing molecular biological methods for generating higher-quality mutant libraries (Chapters 3 and 4). Methodology development over the past decade has indeed led to smaller and smarter libraries that require less screening. Utilizing these advanced genetic methods and strategies, the analytical effort has been reduced generally to the screening of several thousand transformants or less. Therefore, state-of-the-art automated GC and HPLC constitute the option of choice, which means that in many, if not most cases, screening is no longer the slow step in the overall process of directed evolution of stereo- and/or regioselective enzymes. The Trapp-method of multiplexing GC or HPLC [46] offers the potential of further reduction of time expenditure when performing such high-throughput assays. Merging advanced mutagenesis methods and strategies (Chapter 4) with improved analytical procedures such as automated GC and HPLC, optionally in the form of multiplexing chromatography (Scheme 2.21) [47], constitutes the best way to perform directed evolution.



Scheme 2.21 Merging efficient mutagenesis strategies for generating smaller but higher quality mutant libraries with increased ee-assay capacity on the basis of multiplexing GC and/or HPLC [47]. (Reetz [47]. Reproduced with permission of John Wiley & Sons.)

References

- Reviews of screening and selection systems [3]: (a) Reymond, J.-L. (ed) (2006) *Enzyme Assays—High-throughput Screening, Genetic Selection and Fingerprinting*, Wiley-VCH Verlag GmbH, Weinheim; (b) Reymond, J.-L., Fluxa, V.S., and Maillard, N. (2008) Enzyme assays. *Chem. Commun.*, (1), 34–46; (c) McLachlan, M.J., Sullivan, R.P., and Zhao, H. (2009) Directed enzyme evolution and high-throughput screening, in *Biocatalysis for the Pharmaceutical Industry: Discovery, Development and Manufacturing* (eds J. Tao, G.-Q. Lin, and A. Liese), John Wiley & Sons, Inc., Hoboken, NJ; (d) Boersma, Y.L., Dröge, M.J., and Quax, W.J. (2007) Selection strategies for improved biocatalysts. *FEBS J.*, **274** (9), 2181–2195; (e) Martinez, R. and Schwaneberg, U. (2013) A roadmap to directed enzyme evolution and screening systems for biotechnological applications. *Biol. Res.*, **46** (4), 395–405; (f) Wojcik, M., Telzerow, A., Quax, W.J., and Boersma, Y.L. (2015) High-throughput screening in protein engineering: recent advances and future perspectives. *Int. J. Mol. Sci.*, **16** (10), 24918–24945.
- 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.
- (a) Taylor, S.V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed.*, **40** (18), 3310–3335; (b) Lin, H. and Cornish, V.W. (2002) Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed.*, **41** (23), 4402–4425; (c) Aharoni, A., Griffiths, A.D., and Tawfik, D.S. (2005) High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.*, **9** (2), 210–216; (d) van Sint Fiet, S., van Beilen, J.B., and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **103** (6), 1693–1698.
- Stemmer, W.P.C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370** (6488), 389–391.
- Yano, T., Oue, S., and Kagamiyama, H. (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U.S.A.*, **95** (10), 5511–5515.
- Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007) A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.*, **25** (10), 1145–1147.
- Miller, B.G. and Raines, R.T. (2004) Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. *Biochemistry*, **43** (21), 6387–6392.
- Lin, H., Tao, H., and Cornish, V.W. (2004) Directed evolution of a glycosynthase via chemical complementation. *J. Am. Chem. Soc.*, **126** (46), 15051–15059.
- Baxter, S., Royer, S., Grogan, G., Brown, F., Holt-Tiffin, K.E., Taylor, I.N., Fotheringham, I.G., and Campopiano, D.J. (2012) An improved race-mase/acylase biotransformation for the preparation of enantiomerically pure amino acids. *J. Am. Chem. Soc.*, **134** (47), 19310–19313.
- Li, H. and Liao, J.C. (2014) Development of an NADPH-dependent homophenylalanine dehydrogenase by protein engineering. *ACS Synth. Biol.*, **3** (1), 13–20.
- Reetz, M.T. and Wang, L.-W. (2006) High-throughput selection system for assessing the activity of epoxide hydrolases. *Comb. Chem. High Throughput Screening*, **9**, 295–299.
- 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.
- (a) Reetz, M.T., Höbenreich, H., Soni, P., and Fernandez, L. (2008) A genetic selection system for evolving enantioselectivity of enzymes. *Chem. Commun.*, **43**, 5502–5504; (b) Reetz, M.T. and Rüggeberg, C.J. (2002) A screening system for enantioselective enzymes

- based on differential cell growth. *Chem. Commun.*, 1428–1429.
14. Boersma, Y.L., Dröge, M.J., van der Sloot, A.M., Pijning, T., Cool, R.H., Dijkstra, B.W., and Quax, W.J. (2008) A novel genetic selection system for improved enantioselectivity of *Bacillus subtilis* lipase a. *ChemBioChem*, **9** (7), 1110–1115.
 15. (a) Sidhu, S.S. (2005) *Phage Display in Biotechnology and Drug Discovery*, CRC Press, Boca Raton, FL; (b) Yang, G. and Withers, S.G. (2009) Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *Chem-BioChem*, **10** (17), 2704–2715.
 16. (a) Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., and Iverson, B.L. (2000) Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol.*, **18** (10), 1071–1074; (b) Santoro, S.W. and Schultz, P.G. (2002) Directed evolution of the site specificity of Cre recombinase. *Proc. Natl. Acad. Sci. U.S.A.*, **99** (7), 4185–4190; (c) Griswold, K.E., Kawarasaki, Y., Ghoneim, N., Benkovic, S.J., Iverson, B.L., and Georgiou, G. (2005) Evolution of highly active enzymes by homology-independent recombination. *Proc. Natl. Acad. Sci. U.S.A.*, **102** (29), 10082–10087; (d) Mastrobattista, E., Taly, V., Chanudet, E., Treacy, P., Kelly, B.T., and Griffiths, A.D. (2005) High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.*, **12** (12), 1291–1300.
 17. (a) Reetz, M.T., Rüggeberg, C.J., Dröge, M.J., and Quax, W.J. (2002) Immobilization of chiral enzyme inhibitors on solid supports by amide-forming coupling and olefin metathesis. *Tetrahedron*, **58**, 8465–8473; (b) Droge, M.J., Boersma, Y.L., van Pouderooyen, G., Vrenken, T.E., Rüggeberg, C.J., Reetz, M.T., Dijkstra, B.W., and Quax, W.J. (2006) Directed evolution of *Bacillus subtilis* lipase a by use of enantiomeric phosphonate inhibitors: crystal structures and phage display selection. *ChemBioChem*, **7** (1), 149–157.
 18. Lipovsek, D., Antipov, E., Armstrong, K.A., Olsen, M.J., Klibanov, A.M., Tidor, B., and Wittrup, K.D. (2007) Selection of horseradish peroxidase variants with enhanced enantioselectivity by yeast surface display. *Chem. Biol.*, **14** (10), 1176–1185.
 19. Becker, S., Höbenreich, H., Vogel, A., Knorr, J., Wilhelm, S., Rosenau, F., Jaeger, K.E., Reetz, M.T., and Kolmar, H. (2008) Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew. Chem. Int. Ed.*, **47** (27), 5085–5088.
 20. (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**, 2830–2832; (b) Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y., and Bocola, M. (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.*, **132** (26), 9144–9152.
 21. Fernandez-Alvaro, E., Snajdrova, R., Jochens, H., Davids, T., Bottcher, D., and Bornscheuer, U.T. (2011) A combination of in vivo selection and cell sorting for the identification of enantioselective biocatalysts. *Angew. Chem. Int. Ed.*, **50** (37), 8584–8587.
 22. Reetz, M.T. and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chem. Commun.*, **43**, 5499–5501.
 23. Reetz, M.T., Kahakeaw, D., and Sanchis, J. (2009) Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. *Mol. Biosyst.*, **5** (2), 115–122.
 24. Hanes, J. and Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.*, **94** (10), 4937–4942.
 25. Tawfik, D.S. and Griffiths, A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, **16** (7), 652–656.

26. Fallah-Araghi, A., Baret, J.C., Ryckelynck, M., and Griffiths, A.D. (2012) A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip*, **12** (5), 882–891.
27. (a) Ostafe, R., Prodanovic, R., Lloyd Ung, W., Weitz, D.A., and Fischer, R. (2014) A high-throughput cellulase screening system based on droplet microfluidics. *Biomicrofluidics*, **8** (4), 041102; (b) Guan, Z., Zou, Y., Zhang, M., Lv, J., Shen, H., Yang, P., Zhang, H., Zhu, Z., and James Yang, C. (2014) A highly parallel microfluidic droplet method enabling single-molecule counting for digital enzyme detection. *Biomicrofluidics*, **8** (1), 014110; (c) Baret, J.C., Miller, O.J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M.L., Hutchison, J.B., Agresti, J.J., Link, D.R., Weitz, D.A., and Griffiths, A.D. (2009) Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip*, **9** (13), 1850–1858; (d) Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.C., Marquez, M., Klibanov, A.M., Griffiths, A.D., and Weitz, D.A. (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107** (9), 4004–4009; (e) Kintses, B., Hein, C., Mohamed, M.F., Fischlechner, M., Courtois, F., Laine, C., and Hollfelder, F. (2012) Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.*, **19** (8), 1001–1009; (f) Abate, A.R., Hung, T., Mary, P., Agresti, J.J., and Weitz, D.A. (2010) High-throughput injection with microfluidics using picoinjectors. *Proc. Natl. Acad. Sci. U.S.A.*, **107** (45), 19163–19166; (g) Romero, P.A., Tran, T.M., and Abate, A.R. (2015) Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.*, **112** (23), 7159–7164.
28. (a) Hietpas, R.T., Jensen, J.D., and Bolon, D.N. (2011) Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.*, **108** (19), 7896–7901; (b) Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., Bercot, B., Petit, E., Poulain, J., Barnaud, G., Gros, P.A., and Tenaillon, O. (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (32), 13067–13072.
29. Clausell-Tormos, J., Lieber, D., Baret, J.C., El-Harrak, A., Miller, O.J., Frenz, L., Blouwolff, J., Humphry, K.J., Koster, S., Duan, H., Holtze, C., Weitz, D.A., Griffiths, A.D., and Merten, C.A. (2008) Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. *Chem. Biol.*, **15** (5), 427–437.
30. (a) Granieri, L., Baret, J.C., Griffiths, A.D., and Merten, C.A. (2010) High-throughput screening of enzymes by retroviral display using droplet-based microfluidics. *Chem. Biol.*, **17** (3), 229–235; (b) Krone, K.M., Warias, R., Ritter, C., Li, A., Acevedo-Rocha, C.G., Reetz, M.T., and Belder, D. (2016) Analysis of enantioselective biotransformations using a few hundred cells on an integrated microfluidic chip. *J. Am. Chem. Soc.*, **138**, 2102–2105.
31. Reetz, M.T. (2006) in *Enzyme Assays – High-throughput Screening, Genetic Selection and Fingerprinting* (ed. J.-L. Reymond), Wiley-VCH Verlag GmbH, Weinheim, pp. 41–76.
32. 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.
33. Badalassi, F., Wahler, D., Klein, G., Crotti, P., and Reymond, J.-L. (2000) A versatile periodate-coupled fluorogenic assay for hydrolytic enzymes. *Angew. Chem. Int. Ed.*, **39** (22), 4067–4070.
34. (a) Baumann, M., Stürmer, R., and Bornscheuer, U.T. (2001) A high-throughput-screening method for the identification of active and enantioselective hydrolases. *Angew. Chem. Int. Ed.*, **40** (22), 4201–4204; (b) Li, Z., Bütikofer, L., and Witholt, B. (2004) High-throughput measurement of the enantiomeric excess of chiral alcohols by

- using two enzymes. *Angew. Chem. Int. Ed.*, **43** (13), 1698–1702.
35. (a) Janes, L.E., Löwendahl, A.C., and Kazlauskas, R.J. (1998) Quantitative screening of hydrolase libraries using pH indicators: identifying active and enantioselective hydrolases. *Chem. Eur. J.*, **4** (11), 2324–2331; (b) Liu, A.M.F., Somers, N.A., Kazlauskas, R.J., Brush, T.S., Zocher, F., Enzelberger, M.M., Bornscheuer, U.T., Horsman, G.P., Mezzetti, A., Schmidt-Dannert, C., and Schmid, R.D. (2001) Mapping the substrate selectivity of new hydrolases using colorimetric screening: lipases from *Bacillus thermocatenulatus* and *Ophiostoma piliferum*, esterases from *Pseudomonas fluorescens* and *Streptomyces diastatochromogenes*. *Tetrahedron: Asymmetry*, **12** (4), 545–556.
 36. (a) 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; (b) Schrader, W., Eipper, A., Pugh, D.J., and Reetz, M.T. (2002) Second-generation MS-based high-throughput screening system for enantioselective catalysts and biocatalysts. *Can. J. Chem.*, **80**, 626–632; (c) Cedrone, F., Niel, S., Roca, S., Bhatnagar, T., Ait-abdelkader, N., Torre, C., Krumm, H., Maichele, A., Reetz, M.T., and Baratti, J.C. (2003) Directed evolution of the epoxide hydrolase from *Aspergillus niger*. *Biocatal. Biotransform.*, **21** (6), 357–364; (d) DeSantis, G., Wong, K., Farwell, B., Chatman, K., Zhu, Z., Tomlinson, G., Huang, H., Tan, X., Bibbs, L., Chen, P., Kretz, K., and Burk, M.J. (2003) Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM). *J. Am. Chem. Soc.*, **125** (38), 11476–11477.
 37. Tielmann, P., Boese, M., Luft, M., and Reetz, M.T. (2003) A practical high-throughput screening system for enantioselectivity by using FTIR spectroscopy. *Chem. Eur. J.*, **9** (16), 3882–3887.
 38. Reetz, M.T., Eipper, A., Tielmann, P., and Mynott, R. (2002) A practical NMR-based high-throughput assay for screening enantioselective catalysts and biocatalysts. *Adv. Synth. Catal.*, **344**, 1008–1016.
 39. (a) Johannes, T.W., Woodyer, R.D., and Zhao, H. (2006) High-throughput screening methods developed for oxidoreductases, in *Enzyme Assays: High-throughput Screening, Genetic Selection and Fingerprinting* (ed. J.-L. Reymond), Wiley-VCH Verlag GmbH, Weinheim; (b) Schwaneberg, U., Schmidt-Dannert, C., Schmitt, J., and Schmid, R.D. (1999) A continuous spectrophotometric assay for P450 BM-3, a fatty acid hydroxylating enzyme, and its mutant F87A. *Anal. Biochem.*, **269**, 359–366; (c) Pardo, I., Chanaga, X., Vicente, A.I., Alcalde, A., and Camarero, S. (2013) New colorimetric screening assays for the directed evolution of fungal laccases to improve the conversion of plant biomass. *BMC Biotech.*, **13**, 90; (d) Lauchli, R., Rabe, K.S., Kalbarczyk, K.Z., Tata, A., Heel, T., Kitto, R.Z., and Arnold, F.H. (2013) High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase. *Angew. Chem. Int. Ed.*, **52**, 5571–5574; (e) Ma, H., Enugala, T.R., and Widersten, M. (2015) A microplate format assay for real-time screening for new aldolases that accept aryl-substituted acceptor substrates. *ChemBioChem*, **16**, 2595–2598.
 40. Reetz, M.T., Kühling, K.M., Wilensek, S., Husmann, H., Häusig, U.W., and Hermes, M. (2001) A GC-based method for high-throughput screening of enantioselective catalysts. *Catal. Today*, **67** (4), 389–396.
 41. van Leeuwen, J.G.E., Wijma, H.J., Floor, R.J., van der Laan, J.-M., and Janssen, D.B. (2012) Directed evolution strategies for enantiocomplementary haloalkane dehalogenases: from chemical waste to enantiopure building blocks. *ChemBioChem*, **13** (1), 137–148.
 42. (a) Reetz, M.T., Daligault, F., Brunner, B., Hinrichs, H., and Dege, A. (2004) Directed evolution of cyclohexanone monooxygenases: enantioselective biocatalysts for the oxidation of prochiral thioethers. *Angew. Chem. Int. Ed.*, **43** (31), 4078–4081; (b) Kille, S., Zilly, F.E.,

- Acevedo, J.P., and Reetz, M.T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.*, **3** (9), 738–743.
43. (a) Polizzi, K.M., Parikh, M., Spencer, C.U., Matsumura, I., Lee, J.H., Realf, M.J., and Bommaris, A.S. (2006) Pooling for improved screening of combinatorial libraries for directed evolution. *Biotechnol. Progr.*, **22** (4), 961–967; (b) Bougioukou, D.J., Kille, S., Taglieber, A., and Reetz, M.T. (2009) Directed evolution of an enantioselective enoate-reductase: testing the utility of iterative saturation mutagenesis. *Adv. Synth. Catal.*, **351** (18), 3287–3305.
44. Feagin, T.A., Olsen, D.P., Headman, Z.C., and Heemstra, J.M. (2015) High-throughput enantiopurity analysis using enantiomeric DNA-based sensors. *J. Am. Chem. Soc.*, **137** (12), 4198–4206.
45. von Buhler, C., Le-Huu, P., and Urlacher, V.B. (2013) Cluster screening: an effective approach for probing the substrate space of uncharacterized cytochrome P450s. *ChemBioChem*, **14** (16), 2189–2198.
46. Trapp, O. (2007) Boosting the throughput of separation techniques by “multiplexing”. *Angew. Chem. Int. Ed.*, **46** (29), 5609–5613.
47. 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.

3

Gene Mutagenesis Methods

3.1

Introductory Remarks

Traditionally, gene mutagenesis was performed for different reasons using chemicals, radiation, or mutator strains. Bacterial mutator strains are based on artificially enhanced mutation during replication and are easy to apply, requiring little or no knowledge of recombinant DNA techniques [1], as in the case of the *Escherichia coli* strain XL1-Red, which was engineered to have a 5000-fold higher mutation rate than wild-type (WT). Natural mutation rate is low, amounting to about 0.0025/1000 bp in 30 generations due to defects in repair mechanisms, in contrast to the commercially available strain XL1-Red, which causes 0.5/1000 bp mutations. Pitfalls when using this strain are outlined in the Agilent instruction manual [1e]. In directed evolution, mutator strains have been rarely used [1c]; molecular biological techniques have been the preferred choice. As delineated in the introductory Chapter 1, three polymerase chain reaction (PCR)-based gene mutagenesis methods were developed at an early stage in advancing protein engineering:

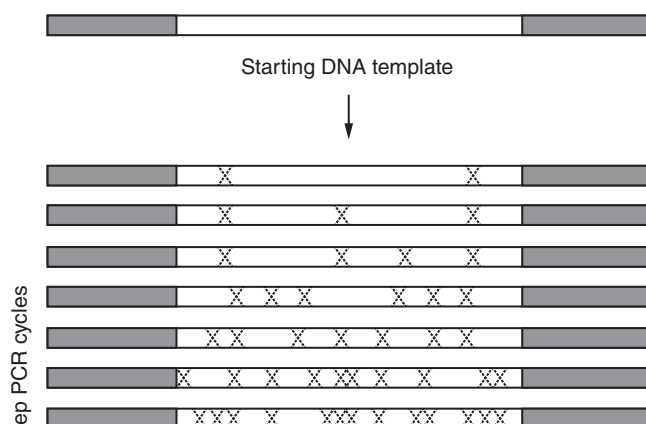
- Error-prone polymerase chain reaction (epPCR)
- Saturation mutagenesis
- DNA shuffling.

These three methods continue to be the most commonly used techniques as reflected in several recent reviews [2]. Any of these or other gene mutagenesis methods can be expected to result in the creation of improved enzyme variants, but the degree of improvement and the invested amount of lab work may be very different, which is sometimes ignored. When applying directed evolution for mechanistic purposes, efficiency plays no role. However, when the purpose is to evolve active, stereoselective, and robust enzymes as catalysts in organic chemistry and/or biotechnology, efficient, reliable, and fast techniques are required. Since the bottleneck of directed evolution continues to be the screening step (Chapter 2), advanced directed evolution techniques are featured here which aim for “smart” mutant libraries exhibiting a high density of notably improved variants. Where appropriate, statistical analyses regarding the degree of oversampling are included, which are essential for optimal library design.

3.2

Error-Prone Polymerase Chain Reaction (epPCR) and Other Whole-Genes Mutagenesis Techniques

For a long time, epPCR (Section 1.1) was the most popular gene mutagenesis method [3], but in recent years a shift to the use of structure-based saturation mutagenesis with creation of focused mutant libraries has occurred (see Section 3.3). Whenever structural information of an enzyme is lacking, epPCR is certainly the preferred approach, although the method is often used even when X-ray structural information is available [4]. It introduces mutations more or less randomly over the whole gene/protein, as opposed to saturation mutagenesis in which random mutations are restricted to defined sites in the protein (Section 3.3). The full-gene epPCR mutagenesis method is illustrated in Scheme 3.1 [5].



Scheme 3.1 Illustration of epPCR [5]. (McCullum *et al.* [5]. Reproduced with permission of Springer.)

In the first step of epPCR, the DNA of a given gene is denatured into single strands by heat treatment, followed by the second step in which annealing of a primer to the DNA single strands occurs, the primers consisting of two oligonucleotides having lengths of 15–30 bp complementary to the ends of the amplified region. In the third step, primer extension is ensured by an appropriate DNA polymerase which lacks proof-reading ability. Based on the original sequence, nucleotides complementary to the single-strand template are added until normal DNA double strands are restored. The *Taq* polymerase from the thermophilic bacterium *Thermus aquaticus* is used most often [3]. It incorporates “wrong” nucleotides at a fairly low frequency of $0.1 - 2 \times 10^{-4}$ [6]. In epPCR experiments, the error rate can be continuously increased to $1 - 20 \times 10^{-3}$. This is achieved by performing the PCR “sloppily” so that mistakes in DNA base pairings are introduced, which encode point mutations in the protein. The mutation rate can be controlled empirically by varying such parameters as the MgCl_2 or MnCl_2

concentrations, employing unbalanced amounts of nucleotides and utilizing higher concentrations of *Taq* polymerase. The error-rate can also be influenced by the incorporation of synthetic mutagenic dNTPs such as 8-oxo-dGTP, which is subsequently eliminated in PCR employing natural dNTPs [7]. On average, one, two, three, or more amino acid exchange events on the protein level can be induced by epPCR. Following amplification, the DNA products (new sequences) are cloned into an appropriate vector, which can be the “rate-limiting” step in terms of lab work [8].

An informative review summarizing the virtues of different techniques for cloning into expression vectors, thereby enabling protein synthesis in a host, has appeared [9]. A number of epPCR kits are commercially available, but they are not identical in terms of performance. Assessing the relative efficiency by performing systematic comparative studies of the kits using different enzymes has not been done to date, although many companies offer helpful information. For example, the GeneMorph II Random Mutagenesis Kits (Stratagene) have been developed to a point where a more uniform mutational epPCR spectrum results as a consequence of reduced bias. It utilizes Mutazyme II DNA polymerase, an epPCR enzyme blend leading to equivalent mutation rates at A's and T's versus G's and C's.

The size of epPCR libraries varies according to the amount of lab work the researcher is willing to invest, 10^4 – 10^6 clones being typical if screening (not selection) is involved. When a selection or fluorescence-activated cell sorting (FACS)-based screening system is available, epPCR libraries comprising 10^6 – 10^9 clones can be handled. A detailed protocol that considers some of the limitations of epPCR was published in 2003 [10], and further developments continue to this day (see further text).

Less experienced users of epPCR should be aware of all the drawbacks of this “shotgun” technique, which is often considered to cover more or less the whole gene/protein. Ideally, the full-gene mutagenesis method would ensure the equal occurrence of all four transitions T_s [$AT \rightarrow GC$ and $GC \rightarrow AT$] and eight transversions T_v [$AT \rightarrow TA$, $AT \rightarrow CG$, $GC \rightarrow CG$, and $GC \rightarrow TA$], which means a probability of 16.67% for each nucleotide substitution pair as well as a $T_s:T_v$ ratio of 0.5 [9]. Also, in an ideal technique, deletions and insertions should not occur. Bias occurs in epPCR due to the redundancy of the generic code, among other factors. For example, $AT \rightarrow GC$ transitions and $AT \rightarrow TA$ transversions are favored. Due to statistical reasons, only single bases are replaced within the triplet codon, which restricts diversity considerably. The event of two or even three base-pair exchanges per codon is extremely unlikely. Therefore, at best, one nucleotide of a given codon is exchanged, leading to just nine (instead of 64 possible) different codons encoding four to seven (instead of 20) different amino acids. The number of “designed” amino acid substitution events in reality depends on the type of original codon. For example, silent mutations are more likely for some types of codons, for example, CGA coding for arginine, than for other types, for example, AAC coding for asparagine. These and other sources of bias have been discussed in an experimental study flanked by a statistical analysis [11]. By analyzing every single

Table 3.1 Theoretical number of variants in a library obtained for a protein consisting of 181 amino acids (lipase A from *Bacillus subtilis*) with one to five amino acid exchanges per enzyme molecule [11].

Number of amino acid exchanges (M)	Number of variants (N)
1	3 439
2	5 880 690
3	6 666 742 230
4	5 636 730 555 465
5	3 791 264 971 605 760

Source: Eggert *et al.* [11]. Reproduced with permission of Taylor & Francis.

codon of the lipase from *Bacillus subtilis* composed of 181 residues, a model calculation was performed, which reveals the real number of enzyme variants obtained by epPCR with one mutation per gene. Employing the conventional algorithm in Eq. (3.1),

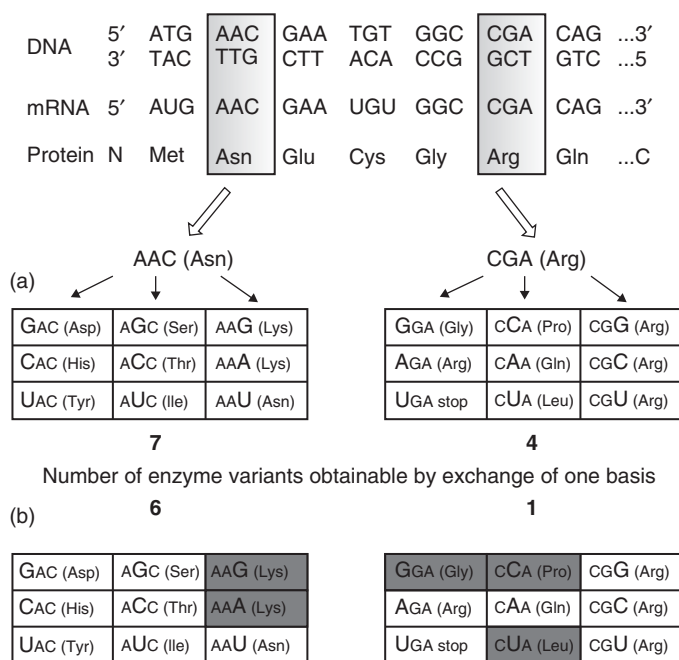
$$N = 19^M X! / [(X - M)! M!] \quad (3.1)$$

where N denotes the number of variants at maximal diversity, M the total number of amino acid substitutions, and X the number of amino acids, the theoretical number of variants assuming the absence of bias or other problems can be calculated (Table 3.1) [11]. The mutational bias for this particular model system was then estimated (Scheme 3.2).

Table 3.2 summarizes the calculated maximum number of variants versus the real number of variants in this system. It can be seen that serious amino acid bias occurs. In further analyses, it was concluded that the calculated library sizes represent only about 20% of the theoretical diversity in this system [11]. Potential users of epPCR should be aware of this phenomenon. Consequently, here and in other systems, Eq. (3.1) cannot be used to estimate the real diversity when using epPCR. Moreover, due to statistical reasons, mutations remote from the active center are more likely to occur.

Bias also results from the exponential nature of PCR amplification [12]. Several other studies that model epPCR by statistical means and in part by experimental data have appeared in the literature. Although different approaches were taken, these analyses likewise illuminate the scope and limitations of this mutagenesis method [13]. Variations and/or improvements have appeared, for example, by alcohol-mediated epPCR [14] or epPCR influenced by heavy water (D_2O) [15], both showing a clear shift in bias. Other DNA polymerases have been used showing different bias [16], and still others have been re-engineered for the purpose of using them in random mutagenesis [17]. In order to reduce bias, two DNA polymerases have been used in one experimental setup, *Taq* and Mutazyme combined with a recombinant mutagenesis method [16].

A mutagenesis method called “*hypermutagenic PCR*”, which involves all four transitions and a large proportion of transversions, can be considered to be an



Scheme 3.2 Mutational bias of ep-PCR in the case of the lipase from *Bacillus subtilis* [11]. The substitution of one nucleotide per codon results in nine new triplets which may encode four to seven different amino acids depending on the type of codon. (a) The example shows that the mutation of the codon AAC coding for asparagine can yield a maximum of seven different amino acids, whereas the mutation of the codon CGA coding for arginine can yield a maximum of four different amino acids. (b) Low frequencies of transversions G → T, C → A, G → C, and C → G result in a further decrease of

diversity: for codon AAC, six different amino acid exchanges may occur, and for the GC-rich codon CGA just a single new amino acid exchange is expected. Background color coding: white shows codons that encode new amino acids, gray indicates silent mutations, or the formation of stop codons, and black shows codons that would require the formation of an unfavored basepair exchange (G → T, C → A, G → C, or C → G). The bold letters indicate nucleotides exchanged by ep-PCR. (Eggert *et al.* [11]. Reproduced with permission of Taylor & Francis.)

extension of epPCR [18], but thus far it has not been used very often in directed evolution. In order to enhance the proficiency of epPCR by alternative means, “megaprimer PCR of whole plasmid” (MEGAWHOP) has been introduced, which is a successful PCR amplification procedure used in many other cloning applications as well [19]. It has also been applied in directed evolution not based on epPCR [20], for example, when opting for saturation mutagenesis [21].

Guidelines and carefully worked out experimental protocols for applying epPCR are available [5, 10, 22]. Unfortunately, in some published directed evolution studies based on epPCR, details are missing, which means that it is not clear which version was actually used.

Table 3.2 Codon usage (left panel), and theoretical and actual numbers of enzyme variants to be obtained upon ep-PCR mutagenesis of *B. subtilis* lipase LipA (right panel) [11].

Number	Codon usage			<i>B. subtilis</i> lipase LipA		
	Codon	Amino acid	Amino acid exchanges ^{a)}	Number of codons	Maximum number of variants ^{b)}	Real number of variants ^{c)}
1	gca	Ala	6	3	57	18
2	gcc	Ala	6	1	19	6
3	gcg	Ala	6	5	95	3
4	gcu	Ala	6	3	57	18
5	aga	Arg	5	2	38	10
6	agg	Arg	6	0	0	0
7	cga	Arg	4	1	19	4
8	cgc	Arg	6	0	0	0
9	cgg	Arg	5	1	19	5
10	cgu	Arg	6	1	19	6
11	aac	Asn	7	7	133	49
12	aau	Asn	7	10	190	70
13	gac	Asp	7	3	57	21
14	gau	Asp	7	6	114	42
15	ugc	Cys	6	0	0	0
16	ugu	Cys	6	0	0	0
17	caa	Gln	6	4	76	24
18	cag	Gln	6	2	38	12
19	gaa	Glu	6	3	57	18
20	gag	Glu	6	0	0	0
21	gga	Gly	4	6	114	24
22	ggc	Gly	6	11	209	66
23	ggg	Gly	5	4	76	20
24	ggu	Gly	6	3	57	18
25	cac	His	7	4	76	28
26	cau	His	7	1	19	7
27	aua	Ile	6	1	19	6
28	auc	Ile	7	2	38	14
29	auu	Ile	7	7	133	49
30	cua	Leu	5	0	0	0
31	cuc	Leu	6	1	19	6
32	cug	Leu	5	5	95	25
33	cuu	Leu	6	4	76	24
34	uua	Leu	4	5	95	20
35	uug	Leu	5	1	19	5
36	aaa	Lys	6	5	95	30
37	aag	Lys	6	6	114	36
38	aug	Met	6	4	76	24
39	uuc	Phe	6	1	19	6
40	uuu	Phe	6	3	57	18
41	cca	Pro	6	2	38	12
42	ccc	Pro	6	0	0	0

Table 3.2 (continued)

Codon usage				<i>B. subtilis</i> lipase LipA		
Number	Codon	Amino acid	Amino acid exchanges ^{a)}	Number of codons	Maximum number of variants ^{b)}	Real number of variants ^{c)}
43	ccg	Pro	6	2	38	12
44	ccu	Pro	6	0	0	0
45	agc	Ser	6	6	114	36
46	agu	Ser	6	1	19	6
47	uca	Ser	4	3	57	12
48	ucc	Ser	6	1	19	6
49	ucg	Ser	5	1	19	5
50	ucu	Ser	6	1	19	6
51	aca	Thr	6	6	114	36
52	acc	Thr	5	0	0	0
53	acg	Thr	6	4	76	24
54	acu	Thr	5	0	0	0
55	ugg	Trp	6	2	38	12
56	uac	Tyr	6	6	114	36
57	uau	Tyr	6	3	57	18
58	gua	Val	5	2	38	10
59	guc	Val	6	5	95	30
60	gug	Val	5	3	57	15
61	guu	Val	6	7	133	42
62	uaa	Stop				
63	uag	Stop	Total number of variants		3439	1077
64	uga	Stop			100%	31.3%

a) Maximum number of amino acid exchanges calculated for each of the naturally occurring 64 codons.

b) Maximum theoretical number of enzyme variants if each amino acid is replaced by the 19 remaining ones.

c) Actual number of amino acid exchanges to be obtained for each amino acid.

Source: Eggert *et al.* [11]. Reproduced with permission of Taylor & Francis.

It should be noted that the above mentioned disadvantages of epPCR does not at all mean that this method is disqualified. The user should simply know what is being generated in terms of diversity and what is not formed. Indeed, this message applies to all gene mutagenesis methods, because none is “perfect.” How much screening is performed depends upon the amount of effort the researcher is willing to invest. Algorithms for estimating diversity and library (in)completeness in epPCR and other gene mutagenesis methods have been developed [22b, 23].

Once the optimal choice of an epPCR version has been made, the question of mutation rate arises. Should the epPCR conditions be chosen so that on average a single amino acid substitution occurs per enzyme, or should they be adjusted leading to two, three, or even more exchange events? In the latter case, protein sequence space increases astronomically. A general answer cannot be given. If an assay system based on selection with or without FACS-analysis (Chapter 2) is

available, then a high mutational rate is the preferred option. If screening is the only possibility as in most directed evolution studies aimed at improving activity, stereo- and or regioselectivity, the answer remained unclear for some time. In the original publication reporting the directed evolution of an enantioselective enzyme (lipase from *Pseudomonas aeruginosa*) as the catalyst in a hydrolytic kinetic resolution (Scheme 1.10) [24], a low error rate averaging one amino acid substitution was applied. This was the general guideline in several previous papers on thermostability, because it was thought at the time that as the number of mutations rises, so does the probability of deleterious unfolding [25]. In the case of the stereoselectivity study, low error-rate epPCR in all four cycles resulted in moderate catalyst improvement as shown in Scheme 1.10 [24]. In contrast, upon repeating the epPCR experiments at higher mutational rate averaging about three amino acid substitutions per enzyme molecule, considerably better results featuring higher degrees of enantioselectivity in the same hydrolytic kinetic resolution were observed [26]. However, no improvements in further epPCR cycles were observed, leading to the conclusion that recursive epPCR is not the most efficient mutagenesis method for enhancing [26] or reversing enantioselectivity [27].

These results originate from a single study and cannot be generalized as such. Nevertheless, the trend suggests that creating theoretically higher diversity while not increasing the screening effort in the expanded protein sequence space is more rewarding than opting for low diversity at the same screening effort [26]. This assessment contradicts the previous study [25] as well as an earlier conclusion stating that no “optimal mutational load for protein engineering” is possible [7]. Just a few years later several studies (not focusing on stereoselectivity) suggested that high mutational rates are in fact preferred [28]. Indeed, original recommendations concerning low mutational rate [25a] were later reversed by coming to the conclusion that “high error-rate random mutagenesis libraries are enriched in functional and improved proteins” [29].

In addition to the above improvements of epPCR leading to reduced bias, several alternative whole-gene random mutagenesis techniques have been proposed, although thus far they are rarely applied. One of several approaches combines the generality of epPCR with Kunkel-mutagenesis (Section 1.2), in which it was shown that competent cells engineered for expressing restriction endonuclease *in vivo* by epPCR are efficient in the elimination of partial and parental clones in digestion and subsequent re-transformation [30]. Error-prone rolling circle amplification (epRCA) [31] and sequence saturation mutagenesis (SeSaM) [32] are further examples. epRCA is a “sloppy” form of traditional rolling circle amplification (RCA) used for diagnostic and biosensing assays [33]. Mutations are introduced by varying the MnCl_2 concentration. The method has been described as the “simplest random mutagenesis method” [31]. A single RCA step is necessary followed by direct transformation of the host strain, leading to mutants characterized by 3–4 mutations per kilobase. Advantages include the fact that no restriction enzymes, ligases, specific primers, or special equipment such as a thermocycler are needed. However, some bias remains. The use of ϕ 29DNA polymerase or mutant thereof

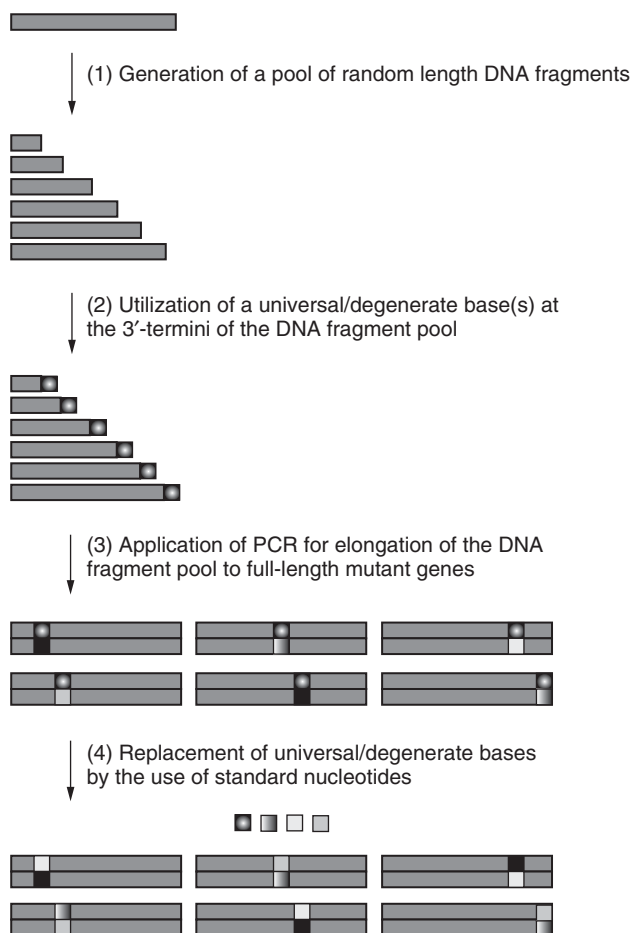
favors C → T and G → A mutations (66%), which differs considerably from *Taq* polymerase. RCA has also been used to construct large phage display antibody libraries [34], and it has been exploited in primer extension mutagenesis [35]. In another approach, the QuikChange™ protocol for site-directed mutagenesis or saturation mutagenesis (see Section 3.3) was performed with added MnCl₂, which resulted in the construction of randomly mutagenized libraries with the claimed advantage that a ligation step becomes superfluous [36]. It remains to be seen whether these methods are in fact simple and efficient enough to be used widely in directed evolution.

The same question pertains to the four-step SeSaM technique, which likewise reduces bias [32]: (i) generation of a pool of random length DNA fragments; (ii) utilization of a universal/degenerate base(s) at the 3'-termini of the DNA fragment pool; (iii) application of PCR for elongation of the DNA fragment pool to full-length mutant genes and (iv) replacement of universal/degenerate bases by the use of standard nucleotides (Scheme 3.3).

SeSaM has the advantage that it is independent of the mutational bias of DNA polymerases and that the fragment distribution of a DNA library can be controlled by the use of different concentrations of the individual Sp-dNTP α S or a combination thereof. By varying the concentrations of NaCl and/or NaOH in the DNA melting step, the mutation rate can be controlled empirically. The second generation SeSaM-Tv protocol is an improvement because it bypasses the requirement of single-stranded DNA (ssDNA) template and employs Vent (exo-) and Deep Vent (exo-) polymerases, thereby increasing the PCR yields and allowing for transversion-enriched sequences [32b]. The disadvantage of this method is the labor involved in the multi-step procedure.

Whole-gene mutagenesis techniques enabling random insertion and deletion (RID) constitute quite different and, to date, little explored approaches to directed evolution. In one of the early reports, a technique for random insertion/substitution was described [37], to be followed by RID, which allows the deletion of an arbitrary number of consecutive bases (up to 16) at random positions and enables the insertion of a specific sequence or random sequences of an arbitrary number at the defined position [38]. The eight-step protocol is summarized in Scheme 3.4.

RID was first tested successfully in the directed evolution of the green fluorescent protein GFP_{UV} by replacing three randomly chosen consecutive bases by a mixture of 20 codons [38]. In *E. coli* as host, six variants were selected, several of which showed quite different fluorescence properties. A yellow fluorescent protein and an enhanced green fluorescent mutant were identified, neither of which were accessible by conventional epPCR. It should be mentioned that the earliest directed evolution studies of GFPs utilized other mutagenesis methods such as epPCR [39] or saturation mutagenesis at a 20 residue region [40]. Unnatural amino acids using expanded genetic codes [41] can also be introduced by this technique [39a, 40, 42], but undesired secondary mutations in the PCR process may occur (as in any PCR-based method). Studies reporting other techniques for deletion mutagenesis have appeared, claiming simpler protocols [43], some based

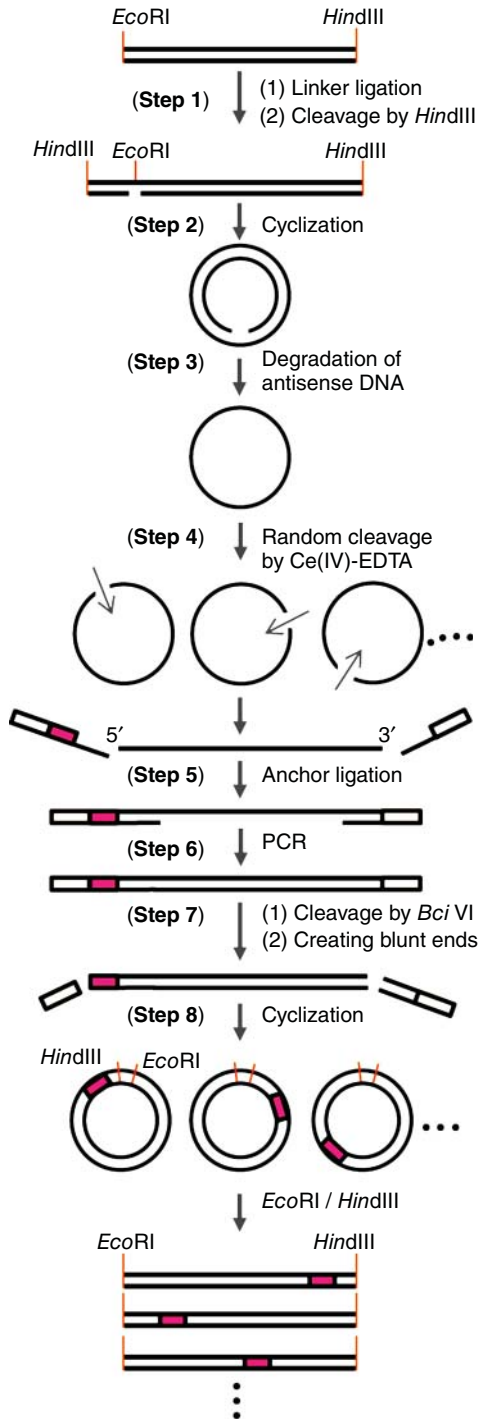


Scheme 3.3 The four basic stages of SeSaM [32a]. Step 1: Generation of a pool of DNA fragments characterized by a random size distribution; step 2: Enzymatic elongation of DNA fragments using the universal base deoxyinosine; step 3: PCR-based full-length gene synthesis using a single-stranded

template and a reverse primer which amplifies the new strand; and step 4: Replacement of deoxyinosine by one of the four standard nucleotides by PCR. (Wong *et al.* [32a]. Reproduced with permission of Oxford University Press.)

purely on rational design as opposed to randomness [44]. A technique dubbed INSULT avoids subcloning and obviates the necessity of special “ultra-competent cells” [45].

In summary, epPCR is the most commonly practiced whole-gene random mutagenesis method, which is especially useful when structural data is lacking. Relative to the original versions [3a], various improvements and alternatives have



Scheme 3.4 Illustration of random insertion/deletion (RID) mutagenesis for the construction of a library of mutant genes [38]. Step 1: (1) The fragment obtained by digesting the original gene with *EcoRI* and *HindIII* is ligated to a linker and (2) the product is then digested with *HindIII* to make a linear dsDNA with a nick in the antisense chain. Step 2: The gene fragment is cyclized with T4 DNA ligase to make a circular dsDNA with a nick in the antisense chain. Step 3: The circular dsDNA is treated with T4 DNA polymerase to produce a circular ssDNA. Step 4: The circular ssDNA is randomly cleaved at single positions by treating with Ce(IV)–EDTA complex. Step 5: The linear ssDNAs, which

have unknown sequences at both ends, are ligated to the 5'-anchor and the 3'-anchor, respectively. Step 6: The DNAs that are linked to the two anchors at both ends are amplified by PCR. Step 7: The PCR products are treated with *BciVI*, leaving several bases from the 5'-anchor, at the 5'-end. The *BciVI* treatment also deletes a specific number of bases at the 3'-end. Step 8: The digested products are treated with Klenow fragment to make blunt ends and cyclized again with T4 DNA ligase. The products are treated with *EcoRI* and *HindIII*, and the fragments are cloned into an *EcoRI*–*HindIII* site of modified pUC18 (pUM). (Murakami *et al.* [38]. Reproduced with permission of Nature Publishing Group.)

been reported (see discussions above). Several standard protocols are available [5, 10, 22]. To date, none of the alternative whole-gene mutagenesis techniques have replaced epPCR on a broad basis.

3.3

Saturation Mutagenesis: Away from Blind Directed Evolution

As outlined in the introductory Chapter 1, saturation mutagenesis involves random amino acid exchange events at a defined single amino acid position or site composed of more than one position in a protein of interest with creation of focused mutant libraries. Sometimes the term “site saturation mutagenesis” (SSM) is used, although the word “site” in this expression is redundant, since by definition saturation mutagenesis means randomization at a defined site. Maximum diversity is introduced when all 20 canonical amino acids are used as building blocks (NNK or NNS codon degeneracy), but reduced amino acid alphabets can be employed in clever ways (see further text). Restricting randomization to a single amino acid position or a site composed of several such positions generally requires X-ray structural data, homology models, and/or consensus sequence information. The decision where randomization should be focused is crucial, which depends upon the purpose of a mutagenesis project [2f]:

- At hot spots identified by epPCR for manipulating theoretically any catalytic property of interest such as thermostability [46] or stereoselectivity [47].
- At positions selected by rational design for enhancing oxidative stability [48].
- At positions predicted by a consensus approach based on multiple sequence alignment (MSA) and guided by structural information for enhancing stability [49].
- At sites lining the binding pocket for manipulating activity, stereo- and regioselectivity, and/or substrate scope [2f,49].

- At remote sites that can be expected to induce allosteric effects for manipulating activity, stereo- and regioselectivity, and/or substrate scope [50].
- At sites displaying high B-factors for manipulating thermostability and/or robustness in the presence of hostile organic solvents [51].

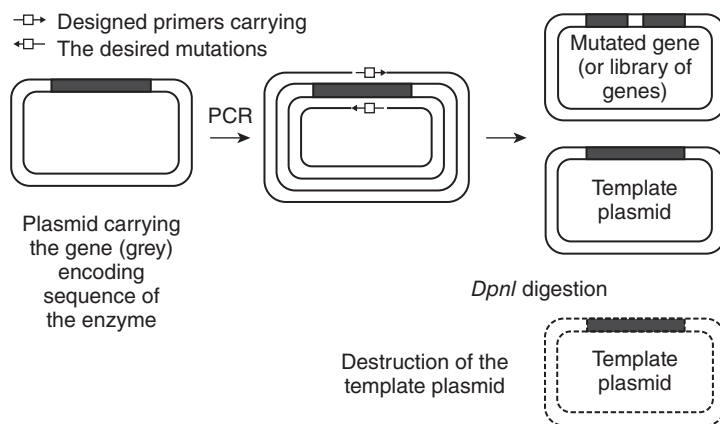
Experience has shown that some choices are more productive than others. If a completely wrong decision is made, then the probability of generating variants displaying improved catalytic profiles is low, for example, performing saturation mutagenesis at sites lining the binding pocket in order to enhance thermostability would not be a reasonable choice, although in exceptional cases such a mutation may in fact lead to enhanced stability. In the discussion that follows, different molecular biological techniques for saturation mutagenesis are described before addressing strategic questions on how to apply them optimally.

In the introductory Chapter 1, a number of older methods of saturation mutagenesis are featured, some of which are no longer in use today due to the emergence of superior protocols. As a most general description of this focused mutagenesis method, appropriate primers that carry the genetic information encoding the desired mutational changes are designed, prepared (commercially), and applied in different procedures. A bewildering number of studies have appeared and continue to be published which feature variations and improvements of saturation mutagenesis, some only reporting improved cloning procedures in specific cases (which are essential for acceptable performance). Systematic comparative studies are rare, and since they cover only a limited number of approaches [2, 21b, 52–56], general conclusions are not possible. Fortunately, trends are emerging. The following three options are chosen most often:

- QuikChange™ protocol [52]
- Megaprimer approach [19, 53]
- Overlap extension polymerase chain reaction (OE-PCR) [54].

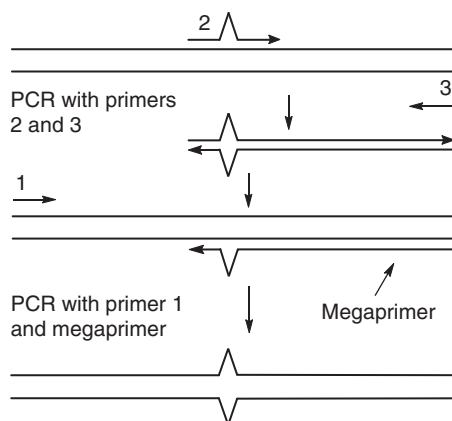
The most commonly used procedure for saturation mutagenesis is the so-called QuikChange™ protocol of Stratagene/Agilent [52], which is based on previous studies (Chapter 1). Originally, it was developed for efficient site-specific mutagenesis, but it can be used to include saturation mutagenesis at a single specified residue or at a site composed of up to five amino acid positions. It consists of several steps: (i) for plasmid amplification the utilization of a pair of complementary primers on both the sense and anti-sense strands that carry the designed mutations; (ii) *DpnI* digestion for selectively removing the methylated parental strands; and (iii) transformation into an appropriate host (e.g., *E. coli* strain) (Scheme 3.5). In the first step, a DNA polymerase such as Pfu or KOD-DNA having no nick-translation activity needs to be used. As already pointed out, algorithms for estimating diversity and degree of library completeness when using saturation mutagenesis have been developed [23].

A reasonable recommendation for potential users is to start with QuikChange™, and in the case of poor or negative results to continue empirically by testing the megaprimer approach or overlap extension based OE-PCR (or combinations



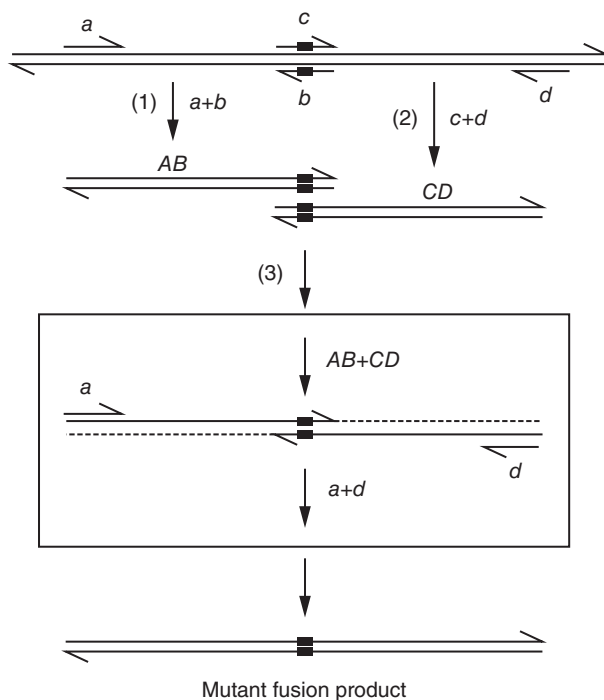
Scheme 3.5 Illustration of saturation mutagenesis based on the QuikChange™ (Stratagene/Agilent) protocol [52]. (Adapted and reproduced from Hogrefe *et al.* [52]. Reproduced with permission of Informa Healthcare USA, Inc.)

thereof). QuikChange™ may fail due to problems associated with primer length and design, especially when more than five amino acid positions are targeted simultaneously. The formal representation of the megaprimer approach is shown in Scheme 3.6 [19, 53, 55]. Two cycles of PCR are required utilizing two flanking primers and one internal mutagenic primer carrying the designed base substitution(s). The first PCR round is carried out using the internal mutagenic primer and the first flanking primer. This PCR product, the “megaprimer”, is purified and then used together with the second flanking primer for the second PCR. This ensures that the final PCR product harbors the desired mutation(s) in a specific DNA sequence.



Scheme 3.6 General illustration of megaprimer PCR [55]. (Adapted from Reikofski and Tao [55]. Reproduced with permission of Elsevier.)

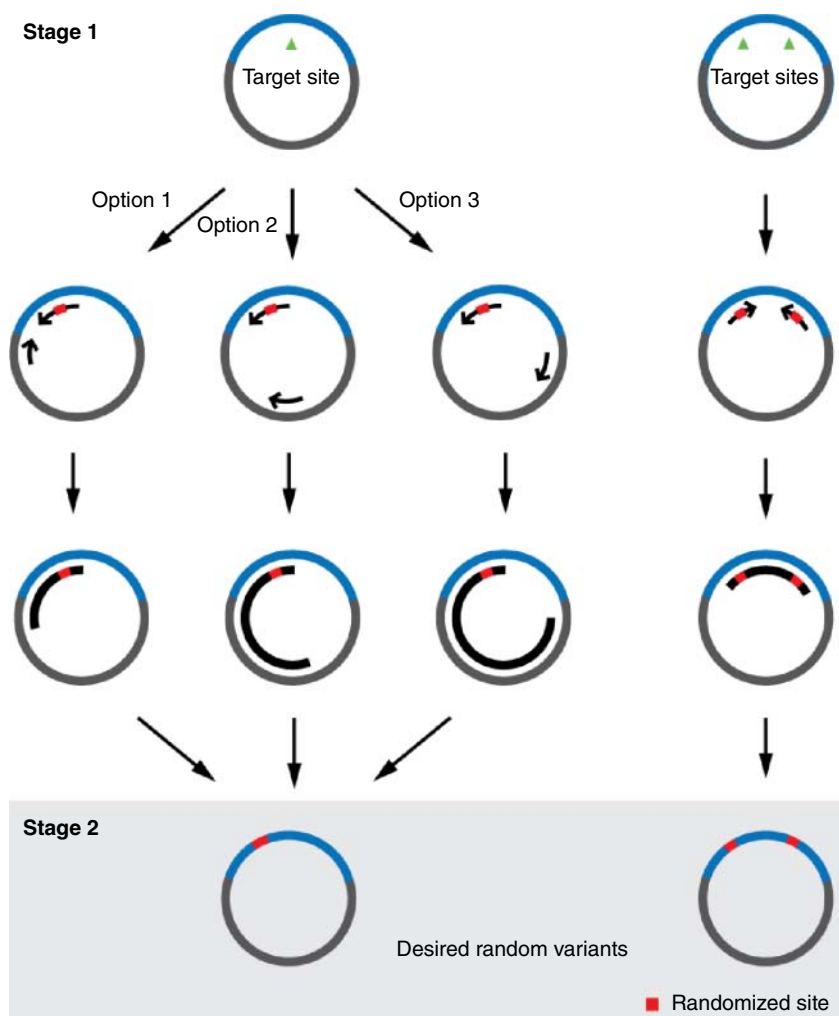
As already pointed out in Chapter 1, OE-PCR is a particularly important advancement when generating focused libraries. It involves several steps as outlined in Scheme 3.7 [54].



Scheme 3.7 Steps in site-directed mutagenesis by overlap extension PCR which can also be used for randomization at single residues or sites composed of more than one amino acid position [54]. Lines with arrows represent the dsDNA and synthetic oligonucleotides with the arrows indicating the 5' to 3' orientation. Small black rectangles denote the site of mutagenesis. Lower-case letters refer to oligos while the PCR products are

indicated by pairs of upper-case letters corresponding to the oligo primers which are employed to generate the product. The box represents the intermediate steps at which the denatured fragments anneal at the overlap and are extended by the DNA polymerase (dotted line). Further PCR amplification occurs by additional primers "a + d." (Ho *et al.* [54a]. Reproduced with permission of Elsevier.)

Despite these improvements, problems may still arise, especially in the case of recalcitrant targets such as large plasmids. An example is P450-BM3 from *Bacillus megaterium*, a case of difficult-to-amplify templates [56]. Based on the concept of using non-overlapping oligonucleotides [57], a notably improved two-stage technique for creating saturation mutagenesis libraries was developed (Scheme 3.8). In the initial stage, the mutagenic primer and the anti-primer (not complementary) anneal to the template. In the second stage, the amplified sequence is utilized as a megaprimer. Importantly, sites comprising more than one residue can be



Scheme 3.8 Efficient method for saturation mutagenesis useful for cases of difficult-to-amplify templates [56], the scheme showing variation of the antiprimer position. The gene is represented in blue, the vector backbone in gray, and the formed megaprimer in black. In the first stage of the PCR, both the mutagenic primer (positions randomized represented by a red square) and the antiprimer (or another mutagenic primer, shown to the right) anneal to the template

and the amplified sequence is used as a megaprimer in the second stage. Finally, the template plasmids are digested using *DpnI*, and the resulting library is transformed in bacteria. The scheme to the left illustrates the three possible options in the choice of the megaprimer size for a single site randomization experiment. The scheme to the right represents an experiment with two sites being simultaneously randomized.

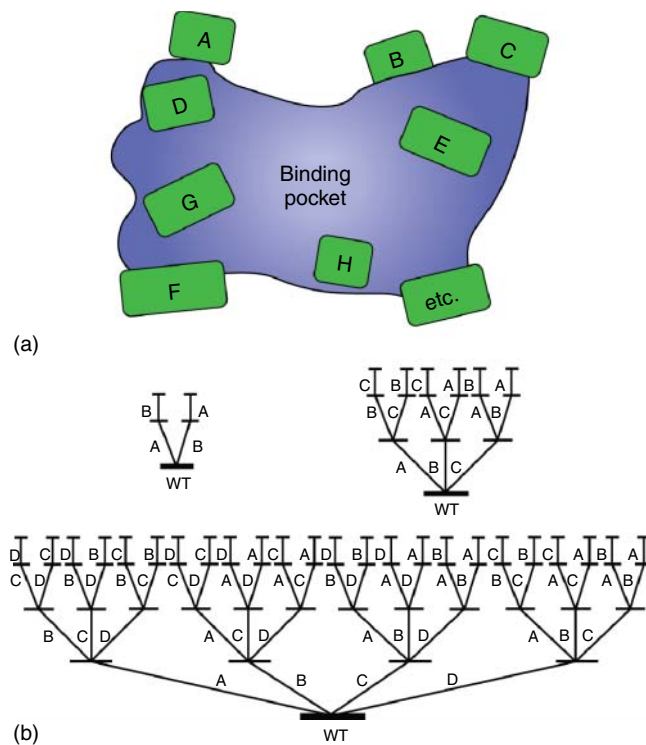
randomized efficiently in a single PCR, irrespective of their location in the gene sequence. In a systematic comparative study, the virtue of this method was carefully compared with the performance of traditional QuikChange™ and related techniques using four different enzymes, P450-BM3 from *B. megaterium*, lipase A from *P. aeruginosa*, lipase B from *Candida antarctica*, and epoxide hydrolase from *Aspergillus niger* (ANEH) [56]. In all cases it proved to be superior in terms of library quality.

Several other strategies for higher-quality saturation mutagenesis libraries, for example, leading to less amino acid bias, have been reported: (i) use of special mono-, di-, or trinucleotide phosphoramidite solutions in the synthesis of redundancy-free mutagenic primers [58] and (ii) so-called MAX strategy based on the synthesis of a template and 20 selection oligonucleotides and their hybridization [59], and its extension to contiguous codons using the ProxiMAX method [60].

A different approach for enhancing the quality of saturation mutagenesis libraries when randomizing several amino acid positions simultaneously is OmniChange [61]. Simultaneous saturation of up to five codons is possible by a four-step procedure: (i) vector and insert formation by traditional PCR using phosphorothiolated oligonucleotides containing the degenerate codons; (ii) use of I_2 as a chemical DNA-cleavage reaction for generating complementary 5'-overhangs in double-stranded DNA by cleavage of the phosphorothiodiester bonds; (iii) application of complementary DNA hybridization for assembling the mutated plasmids; and (iv) transformation in *E. coli* and nick repair. Yet another approach, dubbed OSCARR, utilizes in five steps designed mutagenic cassettes and an optimized one-pot megaprimer PCR [62] with formation of focused libraries [63].

While many of these “tricks” promise higher-quality libraries, choosing the correct randomization site(s) is just as important. When activity, substrate scope, stereoselectivity, and/or regioselectivity are the parameters of interest, rational design exploiting site-specific mutagenesis at the active site has long been used to improve such catalytic features [64, 65]. However, purely rational design is not as general as directed evolution based on focused combinatorial randomization at sites lining the binding pocket [2f]. This directed evolution approach was first employed in a successful attempt to enhance the enantioselectivity of lipase A from *P. aeruginosa* as the catalyst in the hydrolytic kinetic resolution of a chiral ester (Scheme 1.10), saturation mutagenesis being focused on a four-residue site lining the binding pocket [26] (see also Chapter 4 for an analysis of this early experiment).

In view of Emil Fischer’s lock-and-key hypothesis and Linus Pauling’s concept of transition state stabilization by the protein environment, any attempt to reshape an enzyme’s binding pocket by saturation mutagenesis at such sites for manipulating stereo- and/or regioselectivity or substrate acceptance appears logical. It is all the more surprising that this strategy [26] was not systematized until later with the emergence of the combinatorial active-site saturation test (CAST) [66], a convenient acronym for this embodiment of saturation mutagenesis. Residues



Scheme 3.9 Illustration of (a) CAST sites comprising randomization sites A, B, C, and so on [66]. (Reetz *et al.* [66]. Reproduced with permission of John Wiley & Sons.). (b) ISM scheme for 2-, 3-, and 4-site systems involving 2, 6, and 24 upward pathways, respectively [67, 68].

surrounding the binding pocket are first identified on the basis of X-ray data or homology models, designated as A, B, C, and so on, (Scheme 3.9a). Each site may comprise one, two, or more amino acid positions. The problem of how to group single-residue sites into multiple-residue sites (if at all), constitutes a strategic question to be treated in Section 3.5. The libraries A, B, C, and so on, may harbor improved hits, but the degree of improvement may not be acceptable. This problem can be solved by applying iterative saturation mutagenesis (ISM) [67a], which was first reported in 2005 using an epoxide hydrolase as the catalyst in the hydrolytic kinetic resolution of a racemic epoxide [67b]. Accordingly, the gene of a hit in a given library is used as a template for saturation mutagenesis at another site, and the process is continued until all sites have been visited once in an upward pathway. The schemes for 2-site, 3-site, and 4-site ISM systems are illustrated in Scheme 3.9b, involving 2, 6, and 24 upward pathways, respectively [2f, 67, 68]. Questions of how to identify the optimal upward pathway and how to escape from local minima are treated in Chapter 4. In an initial attempt, four cycles of ISM led to a selectivity factor of $E=49$ in the hydrolytic kinetic resolution [67b], which was boosted to $E=115$ upon exploring further pathways [67a].

When applying saturation mutagenesis, the issue of oversampling in the screening process needs to be considered [2f, 68, 69], which is essential when choosing the optimal evolutionary strategy (Chapter 4). Two types of statistical analyses have been proposed for estimating the required degree of screening (necessary number of transformants) assuming the absence of bias: The Patrick/Firth algorithm [23a,b] or similar metrics [23c,d], which focus on %-coverage of a library, and the recently introduced Nov metric that refers to the first, second, third, or *n*th best mutant in a given library as a function of the number of screened transformants [23e, 70]. The calculated minimum number of screened transformants, for example, 95% library coverage, or for finding the best mutant, serves as a useful guide when deciding on how much effort to invest in the screening step. In practice, a number of studies have shown that full library coverage, which is equivalent to ensuring that the best mutant has been identified, is not necessary for obtaining practical results [2f, 51b, 71]. The Patrick/Firth approach as featured in the computational tool GLUE-IT [23b] has been incorporated in the user-friendly CASTER computer aid for designing saturation mutagenesis libraries [51b], available free of charge on the author's homepage (<http://www.kofo.mpg.de/en/research/biocatalysis>). These and other computational tools are summarized in Section 3.7.

When designing saturation mutagenesis libraries, all 20 canonical amino acids are traditionally used as building blocks (NNK or NNS codon degeneracy), but reduced amino acid alphabets can also be considered for certain reasons [2f, 68, 69]. Constructing whole proteins with less than the normal 20 canonical amino acids had been reported earlier for different reasons that are related to the following questions:

- Did nature use less than 20 canonical amino acids for constructing primordial proteins before evolving the optimal full number [72a–d]?
- Do proteins constructed with notably reduced amino acid alphabets enable proper folds [72b,c]?
- Do proteins constructed with notably reduced amino acid alphabets enable proper folds as well as activity [72d]?

All three questions have been answered affirmatively. For example, a chorismate mutase was constructed using a set of nine amino acids, leading to a proper fold as well as enzyme activity [72d]. Therefore, the generally observed success when choosing an appropriate reduced amino acid alphabet in saturation mutagenesis at appropriate sites is not so surprising, but it took some years before it was applied for the first time with the aim of engineering stereoselective enzymes [69b].

As the number of amino acid positions in a randomization site increases, so does the screening effort at all degrees of library coverage. This effect is illustrated in Table 3.3, which features the case of 95% library coverage when NNK codon degeneracy (N: adenine/cytosine/guanine/thymine; K: guanine/thymine) encoding all 20 canonical amino acids is used versus NDT (D: adenine/guanine/thymine; T: thymine) encoding 12 amino acids that comprise a fairly balanced mixture of polar/non-polar, charged/non-charged, and hydrophobic/hydrophilic building

Table 3.3 Oversampling necessary for 95% library coverage as a function of NNK versus NDT codon degeneracy and the number of amino acid positions in a randomization site [68].

Number of amino acid positions at one site	NNK		NDT	
	Codons	Transformants needed	Codons	Transformants needed
1	32	94	12	34
2	1 028	3 066	144	430
3	32 768	98 163	1 728	5 175
4	$>1.0 \times 10^6$	$>3.1 \times 10^6$	$>2.0 \times 10^4$	$>6.2 \times 10^4$
5	$>3.3 \times 10^7$	$>1.0 \times 10^8$	$>2.5 \times 10^5$	$>5.5 \times 10^5$
6	$>1.0 \times 10^9$	$>3.2 \times 10^9$	$>2.9 \times 10^6$	$>8.9 \times 10^6$
7	$>3.4 \times 10^{10}$	$>1.0 \times 10^{11}$	$>3.5 \times 10^7$	$>1.1 \times 10^8$
8	$>1.0 \times 10^{12}$	$>3.3 \times 10^{12}$	$>4.2 \times 10^8$	$>1.3 \times 10^9$
9	$>3.5 \times 10^{13}$	$>1.0 \times 10^{14}$	$>5.1 \times 10^9$	$>1.5 \times 10^{10}$
10	$>1.1 \times 10^{15}$	$>3.4 \times 10^{15}$	$>6.1 \times 10^{10}$	$>1.9 \times 10^{11}$

Source: Acevedo-Rocha *et al.* [68]. Reproduced with permission of Springer.

blocks (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly) [2f, 68, 69]. It can be concluded that saturation mutagenesis at a single residue is best performed with NNK codon degeneracy because screening a single 96-well plate suffices for 95% library coverage. In fact, in some cases it is just as easy to perform 19 separate site-specific mutagenesis experiments, which makes screening superfluous. Practical differences between NNK and NDT arise when randomizing larger sites. If, in a given project, an efficient selection or FACS-based system is available (Chapter 2), then there is little reason to choose NDT codon degeneracy because it generates lower diversity. On the other hand, if such systems are unavailable and screening is the only option, as in the case of manipulating stereoselectivity, rate, or substrate scope, then NDT or even smaller amino acid alphabets should be considered.

A fundamental question arises from such calculations. If a researcher has chosen a randomization site, for example, a three-residue site, and for practical reasons has defined an upper limit regarding the number of transformants to be screened, for example, 5000, then two options are possible:

- Application of NNK codon degeneracy, which covers only a very small portion of the respective library, for example, 15%.
- Use of NDT codon degeneracy, which ensures 95% library coverage.

The question as to which option should be chosen has been addressed empirically [69a]. Using an epoxide hydrolase as the catalyst in the hydrolytic kinetic resolution of a racemic substrate, an NNK- and an NDT-library were constructed, and in each case 5000 transformants were screened. This means 15% versus 95% library coverage, respectively. The two equally sized libraries were checked for quality, meaning the frequency of improved hits and the degree of

improved enantioselectivity. It turned out that the NDT-library had a distinctly higher quality [69a]. Although this is presently the only study addressing this basic question, it can be concluded that striving for higher library coverage may be more important than ensuring maximum structural diversity by using all 20 canonical amino acids as building blocks, but covering only a small segment of the respective protein sequence space. Clearly, more comparative studies are needed to support this recommendation.

Irrespective of the need for more research, a wide variety of reduced amino acid alphabets have been utilized with great success, ranging from 12 amino acids (NDT or DNT codon degeneracy) to only one amino acid as building block (Chapter 4). In summary, the use of reduced amino acid alphabets when applying saturation mutagenesis and ISM constitutes a major step forward in the directed evolution of enzymes with enhanced or reversed stereo- and regioselectivity, activity, substrate scope [2f, 21, 68, 69, 71], and even thermostability [51b] (for engineering protein robustness, see Chapter 6).

As already pointed out, ensuring 95% library coverage (or even more) is not necessary in most practical applications, but knowledge of such numbers is useful when designing saturation mutagenesis experiments. For this reason, the numbers computed for the whole range of library coverage from 0% to 95% serve as an imminently useful guide. Figures 3.1 and 3.2 convey the situation for NNK versus NDT codon degeneracy, respectively, in both cases sites comprising one, two, three, four, and five amino acid positions being considered [69a]. Graphs of this kind are easily generated for any codon degeneracy by consulting the CASTER computer aid (<http://www.kofo.mpg.de/en/research/biocatalysis>) [51b].

Based on the Patrick/Firth algorithm [23a,b], another index of interest to the experimenter was derived, namely the oversampling factor O_f starting from

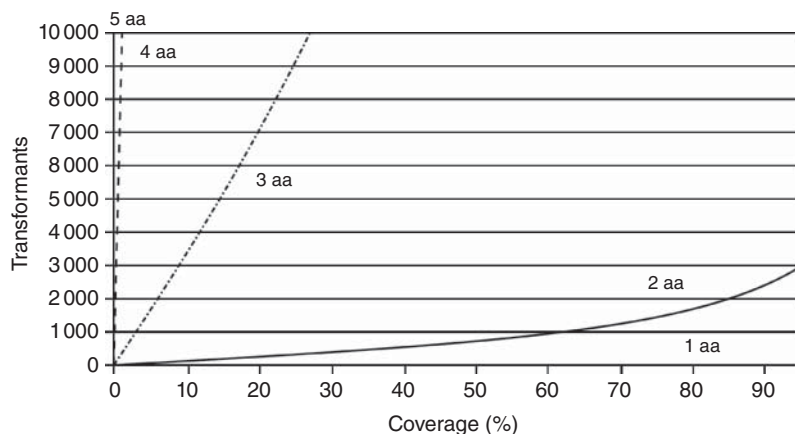


Figure 3.1 Library coverage calculated for NNK codon degeneracy at sites consisting of one, two, three, four, and five amino acid positions (aas, amino acids) [69a]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

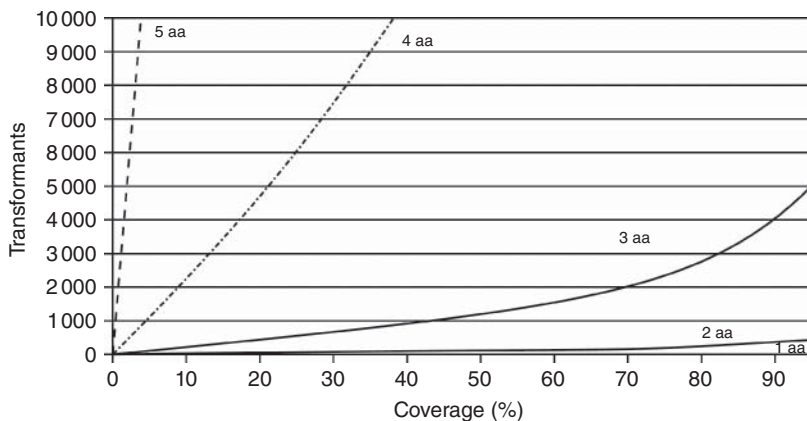


Figure 3.2 Library coverage calculated for NDT degeneracy at sites consisting of one, two, three, four, or five amino acid positions [69a]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

Eq. (3.2), where T is the number of transformants actually screened, P denotes the probability that a particular sequence occurs in the library, and F_i is the frequency [2f, 69a]:

$$T = \frac{-\ln(1 - P_i)}{F_i} \quad (3.2)$$

Upon substituting for F_i , the relationship reduces to Eq. (3.3), where V is the number of mutants on gene level in a given library:

$$T = -V \ln(1 - P_i) \quad (3.3)$$

thereby defining the correlation between the number of mutants V of a given library and the number of transformants T that need to be screened for a defined degree of completeness.

This leads to the oversampling factor O_f which defines the degree of oversampling necessary for achieving a certain completeness under the assumption of no amino acid bias (Eq. (3.4)) [69a]:

$$O_f = T/V = -\ln(1 - P_i) \quad (3.4)$$

When computing the oversampling factor O_f as a function of %-coverage, the curve featured in Figure 3.3 results, spanning the whole range. For example, when aiming for 95% library coverage, the oversampling factor O_f amounts to about 3, which means that a threefold excess of transformants relative to the theoretical number of possible mutants on DNA level needs to be screened.

Researchers preferring the Nov approach, which also assumes the absence of bias, can use the respective computer aid TopLib [70b]. The results of a Nov analysis for sites comprising one, two, three, or four amino acid positions are shown in Figure 3.4 [23e].

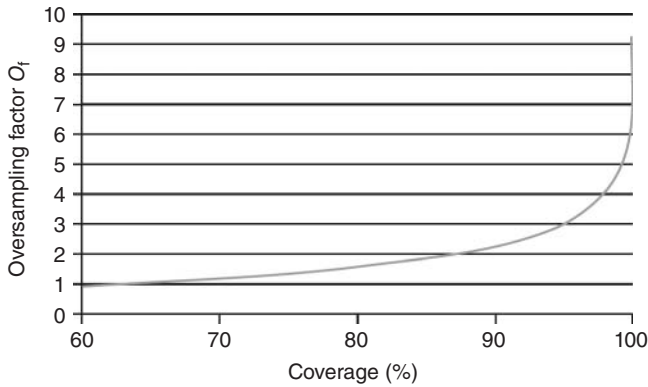


Figure 3.3 Correlation between oversampling factor O_f and percent library coverage [69a]. (Reetz *et al.* [69a]. Reproduced with permission of John Wiley & Sons.)

Obviously, there is a correlation between the Patrick/Firth and the Nov metrics, that is, between a given %-coverage and the n th best variant, respectively. In order to illustrate this for potential users of these metrics who may prefer one or the other statistical approach, the graph in Figure 3.5 has been computed [73].

Whichever saturation mutagenesis method is used, it is advisable to routinely apply the quick quality control (QQC) for checking the expected diversity of a mutant library on DNA level [74a], because screening too many plates for something that does not exist makes no sense. QQC can be performed quickly by pooling a limited number of transformants that are then sequenced. Since the cost of sequences has gone down drastically in recent years, this small investment may well reduce wasted time and resources. Not as fast as QQC, but having the advantage of quantitative quality assessment is the Stewart method, which generates so-called Q -values [74b].

In addition to the above advancements, several other molecular biological modifications have been reported for increasing the quality of saturation mutagenesis libraries, the goal again being the reduction of amino acid bias and elimination of stop codons. When randomization with introduction of all 20 canonical amino acids as building blocks is aimed for, two recent studies deserve particular attention. Traditionally, NNS or NNK codon degeneracy has been used, but as already seen, library quality may be poor. The Tang procedure dedicated to the construction of “small-intelligent” focused libraries [75a] and the related “22c-trick” [75b] are well suited to solve these fundamental problems. In the Tang procedure, a primer mixture is designed so that exactly one codon occurs per amino acid, assisted by a software tool called *DC-analyzer*. Stop codons and eight rare *E. coli* are also eliminated. The Tang mixture is composed of:

- NDT codons encoding 12 amino acids (N, S, I, H, R, L, Y, C, F, D, G, and V)
- VMA codons encoding six amino acids (E, A, Q, P, K, and T)
- ATG codon encoding one amino acid (M)
- TGG codon encoding one amino acid (W).

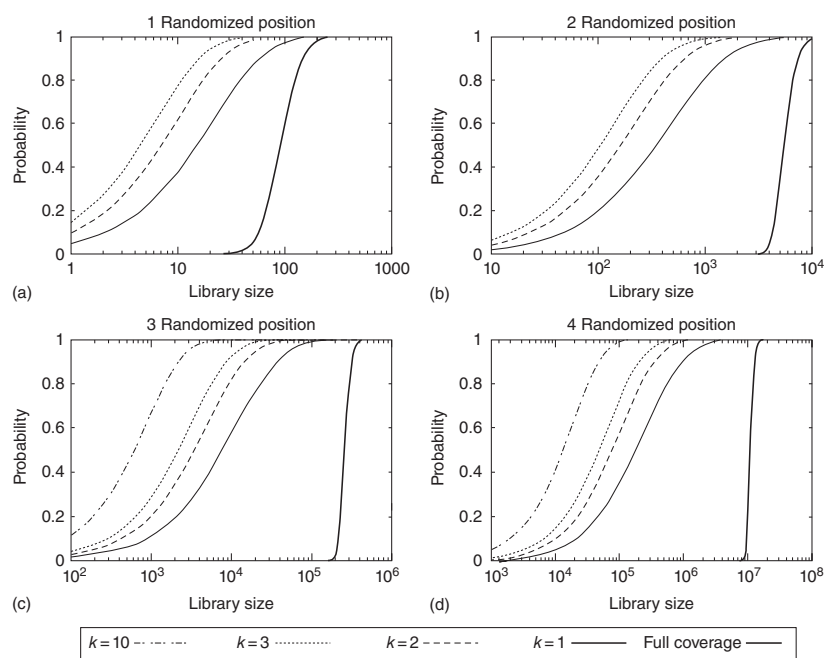


Figure 3.4 Probabilities of “full coverage” and of discovering at least one of the top k protein variants in variant space as a function of the library size when randomizing sites comprising one, two, three, and four amino acid positions in the case of NNK codon degeneracy [23e]. (Nov [23e]. Reproduced with permission of American Society for Microbiology.)

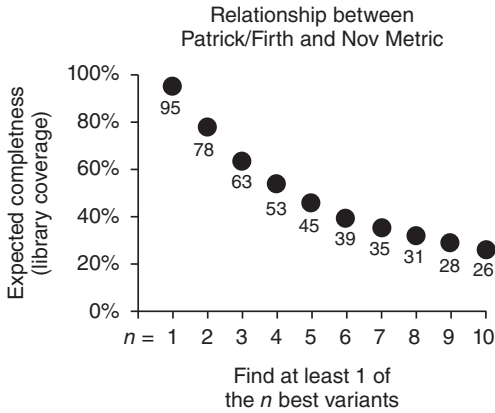


Figure 3.5 Patrick-Firth versus Nov statistical metrics [73]. Relationship between the expected completeness (i.e., library coverage) algorithm by Patrick and Firth as computed by GLUE-IT, and the concept of finding at least one of the n th-best variants (with a 95% probability) by Nov as

computed by TopLib. This mathematical relationship is independent of the number of positions randomized and of the randomization scheme. (Höbenreich *et al.* [73]. Reproduced with permission of American Society for Microbiology.)

When applying the “22c-trick,” 22 unique codons encoding the 20 canonical amino acids are used:

- NDT codons encoding 12 amino acids (N, S, I, H, R, L, Y, C, F, D, G, and V)
- VHG codons encoding nine amino acids
- TGG codon encoding one amino acid (W).

It can be seen that in the Tang approach, four primers with two degeneracies (NDT, VMA) and two coding sequences (ATG, TGG) are involved, together targeting all 20 amino acids while theoretically eliminating all bias. A codon to amino acid ratio of 20:20 is ideal. The 22c trick requires two more codons. Nevertheless, the difference between the two techniques is not great, and both approaches have advantages and disadvantages [75b]. From a cost perspective, which has been analyzed using statistical methods [76], the total number of primers should be minimized, because this increases with n number of residues. In the case of the Tang approach, one site requires four equimolar mixed sense or antisense primers, but this can double to eight depending on the technique. Along such lines, a notably higher number of primers is required when considering two or three residues, namely 16 or 64 in the case of sense or antisense and 32 or 128 sense and antisense separately synthesized primers, respectively [75b].

In the study describing the “22c-trick,” the screening efforts for NNN, NNK(S), and 22c were compared when using two- or three-residue sites (Figure 3.6). The benefits of the “22c-trick” are evident.

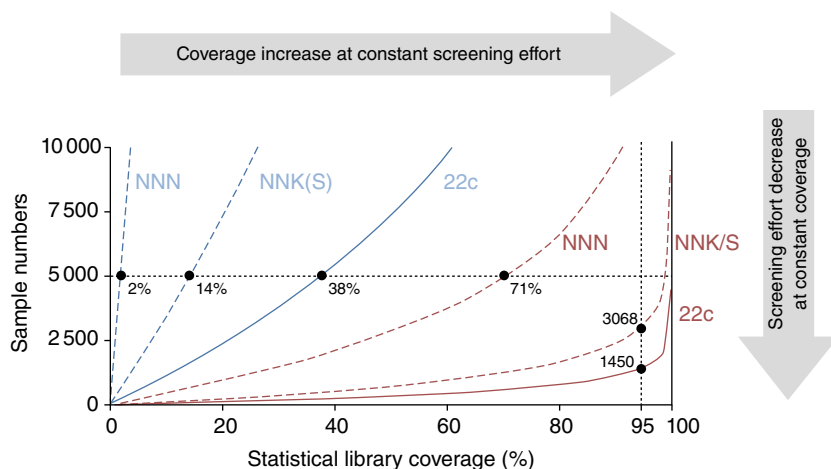


Figure 3.6 Screening effort required for different randomization schemes regarding sites composed of two or three amino acid residues [75b]. The choice of codon degeneracy dictates the sampling size for a desired statistical coverage of the library. For a 95% library coverage targeting two amino acid residues (red lines), 3068 samples have to be screened in the case of NNK/S, whereas only 1450 are necessary when applying the 22c-trick (53% lower screening effort). However, if

the assumed capacity of medium-throughput systems is limited to 5000 samples, the library coverage drops to 71% when using NNN degeneracy. Similarly, when targeting three amino acid residues (blue lines) and limiting the sample size to 5000 colonies or transformants, the library coverage changes drastically to 38, 14, and 2% in the case of the 22c-trick, NNK/S, and NNN, respectively. (Kille *et al.* [75b]. Reproduced with permission of American Chemical Society.)

Using a Baeyer–Villiger monooxygenase CHMO as the model enzyme, the QCC [74a] was applied to an NNK library and the respective 22c-library (Figure 3.7). Here again the superior library quality of “22c trick” is obvious [75b].

Rather than utilizing reduced amino acid alphabets, expanded genetic codes encoding noncanonical amino acids can also be applied in directed evolution [77]. In doing so, different strategies are possible:

- Utilize more than 20 amino acids as building blocks in a randomization procedure.
- Introduce one or more non-natural amino acid at a strategic site and then use such a mutant as a template for directed evolution employing a designed reduced amino acid alphabet for saturation mutagenesis.

Expanded genetic codes [41] have already been applied in the generation of artificial metalloenzymes [77c] and in manipulating the spectral properties of GFPs [77d]. From a practical point of view, notable advantages still need to be demonstrated, but this may well be a matter of time. Site-specifically introducing an amino acid having a synthetic side-chain characterized by a diphosphane or dipyrroline moiety which strongly binds transition metals, en route to artificial metalloenzymes, offers exciting perspectives.

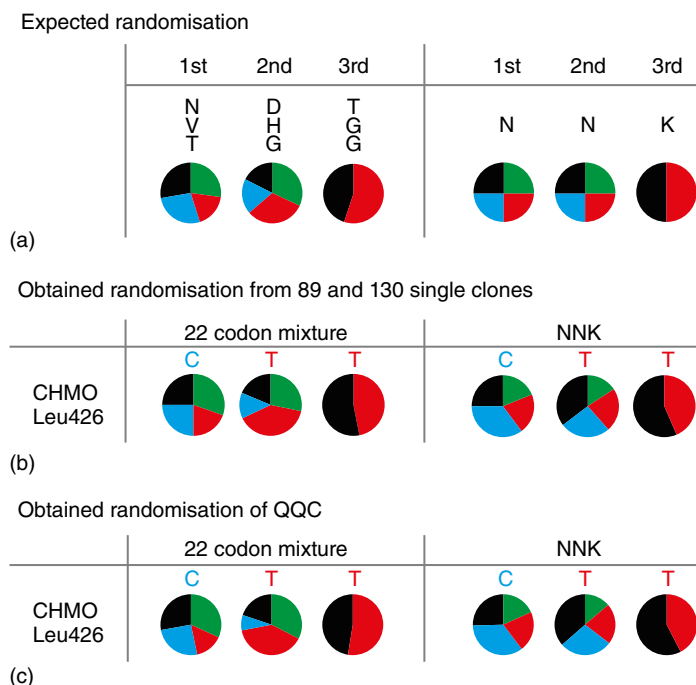


Figure 3.7 Distribution of nucleotide bases in the randomized residue Leu426 of CHMO [75b]. The percentual distribution of nucleotides is shown in pie diagrams for each of the three randomized bases using the 22c-trick (left) and NNK (right) degeneracies. (a) Theoretical expected distribution. (b) Experimental distribution calculated from the sequencing of 89 and 130 individual clones

from the 22c-trick and NNK libraries, respectively. (c) Experimental quick quality control from colony pooling. The nucleotide base guanidine (G) is depicted in black, adenosine (A) in green, threonine (T) in red, and cytosine (C) in blue. (Kille *et al.* [75b]. Reproduced with permission of American Chemical Society.)

3.4

Recombinant Gene Mutagenesis Methods

Methods that cause the breaking and rejoining of DNA in new combinations are called *recombinant techniques*. Numerous embodiments of this general approach to gene mutagenesis have been developed, the most prominent being DNA shuffling [78]. Three general versions are possible: (i) single-gene shuffling; (ii) shuffling of a set of mutants generated from a single gene as illustrated in Scheme 1.8; and (iii) family shuffling using two or more homologous enzymes. Briefly, genes are digested with a DNase with random production of double-stranded oligonucleotide fragments of 10–50 bp, which are then PCR-amplified and purified from agarose gel. Reassembly by a DNA polymerase is achieved in a primerless PCR, followed by amplification in a second PCR employing a pair of flanking

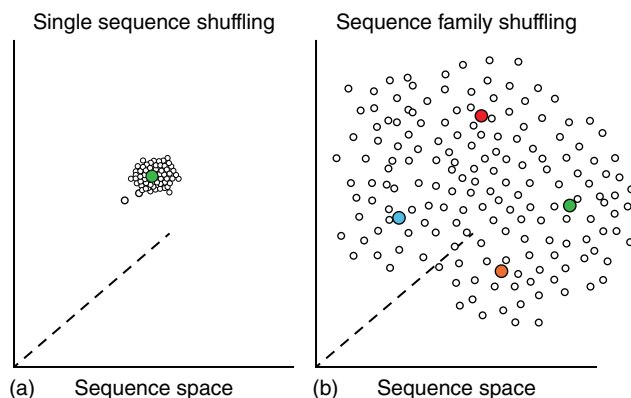


Figure 3.8 Searching sequence space by single-gene shuffling versus family shuffling [78b]. (Cramer *et al.* [78b]. Reproduced with permission of Nature Publishing Group.)

primers, leading to full-length mutant genes. An experimental comparison has shown that family shuffling ensures dramatically higher diversity than single-gene DNA shuffling (Figure 3.8) [78b]. In fact, single-gene shuffling provides libraries having >97% identical enzymes.

DNA shuffling is a multi-step process, which is fairly easy to perform, but as in any mutagenesis method, such issues as optimization, limitations, and potential pitfalls should be considered. For best performance, several variables need to be optimized, including fragment size for controlling the frequency of crossovers [78], temperature cycle during reassembly, amount of assembly, and the number of cycles in amplification [12, 78]. Thus, some experience is necessary for optimal performance. In the case of family shuffling, homology should amount to >70% for acceptable library quality; self-hybridization with formation of WT enzymes also needs to be minimized [78]. Point mutations generally occur during the PCR step, which can be increased by $MnCl_2$ addition. It is also possible to add designed synthetic nucleotides in a type of saturation mutagenesis during the reassembly process, which is a useful technique called combinatorial multiple-cassette mutagenesis (CMCM) [78c]. In order to assess crossover efficiency and to improve DNA shuffling, probe hybridization in macroarray format has been developed, which allows the analysis of chimeric DNA libraries [79a]. This practical procedure also reveals any bias that may occur in shuffling experiments. User-friendly algorithms for estimating diversity and library completeness in DNA shuffling have been developed [79b].

Studies focused on improving DNA shuffling protocols have been summarized in several reviews [2, 12] and detailed experimental procedures are available [79b]. An improved version of family shuffling using ssDNA has also been reported in which the gene is cleaved by restriction enzymes instead of DNase I [80]. Accordingly, ssDNA templates in place of double-stranded templates are employed in the DNase I fragmentation. ssDNA templates reduce the probability of homo-duplex

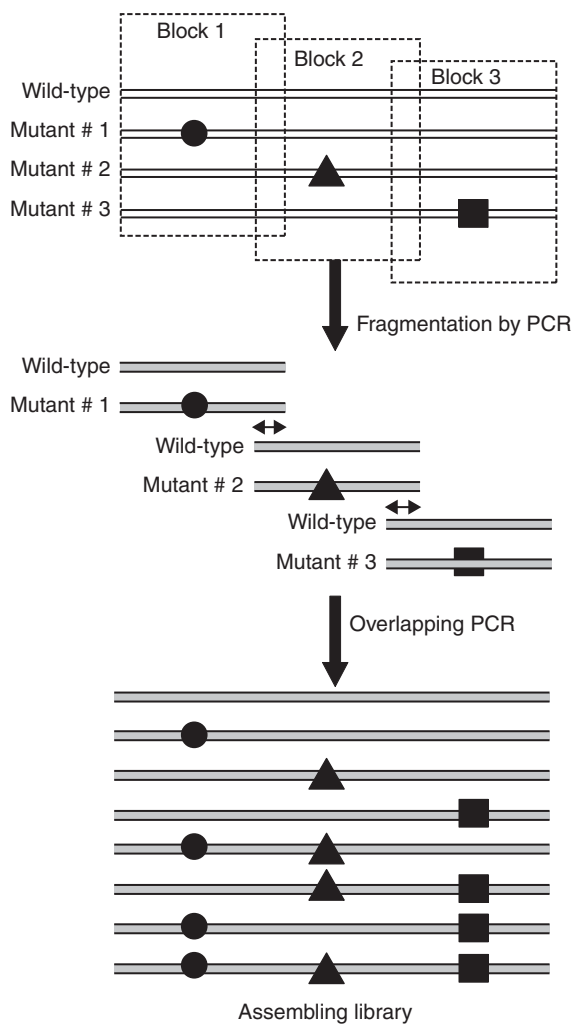
formation, thereby increasing the quality of DNA shuffling libraries as demonstrated in a comparative study using dsDNA versus ssDNA templates [80c]. In a recent more comprehensive comparative study that includes practical experimental protocols, “restriction enzyme-mediated” family shuffling was shown to be superior [81].

In a different shuffling approach, staggered extension process (StEP) was developed based on cross-hybridization of growing gene fragments as the DNA polymerase-catalyzed primer extension process occurs [82]. Following denaturation, the primers anneal and extend under conditions that limit extension, which ensures that the primers re-anneal to different parent sequences throughout the multiple cycles randomly. Then the recombinant full-length gene products are PCR amplified.

In addition to these advancements, a number of alternative and/or improved recombinant procedures have been reported, which focus on various issues such as the problem of low homology [6]. These include random chimeragenesis on transient templates (RACHITT) [83], oligonucleotide gene shuffling (DOGS) [84] in comparison with random drift mutagenesis (RNDM) [85], degenerate incremental truncation for the creation of hybrid enzymes (ITCHY) [86], Thio-ITCHY [87], SCRATCHY (a combination of ITCHY and DNA shuffling) [88], sequence homology-independent protein recombination (SHIPREC) [89], sequence-independent site-directed chimeragenesis (SISDC) [90], recombined extension on truncated templates (RETT) [91], recombination-dependent exponential amplification polymerase chain reaction (RDA-PCR) [92], and structure-based combinatorial protein engineering (SCOPE) [93]. It should be mentioned that some of these techniques have not been used very often following their initial publication, but the future will reveal whether any of the ideas will reappear in future studies.

Yet another recombinant method is “biased mutation-assembly,” according to which a library is created by overlap extension PCR with DNA fragments from a WT protein and phenotypically advantageous mutant genes [94]. The number of assembled mutations in the WT gene is controlled stochastically by the mixing ratio of the WT fragments to the mutant DNA fragments (Scheme 3.10).

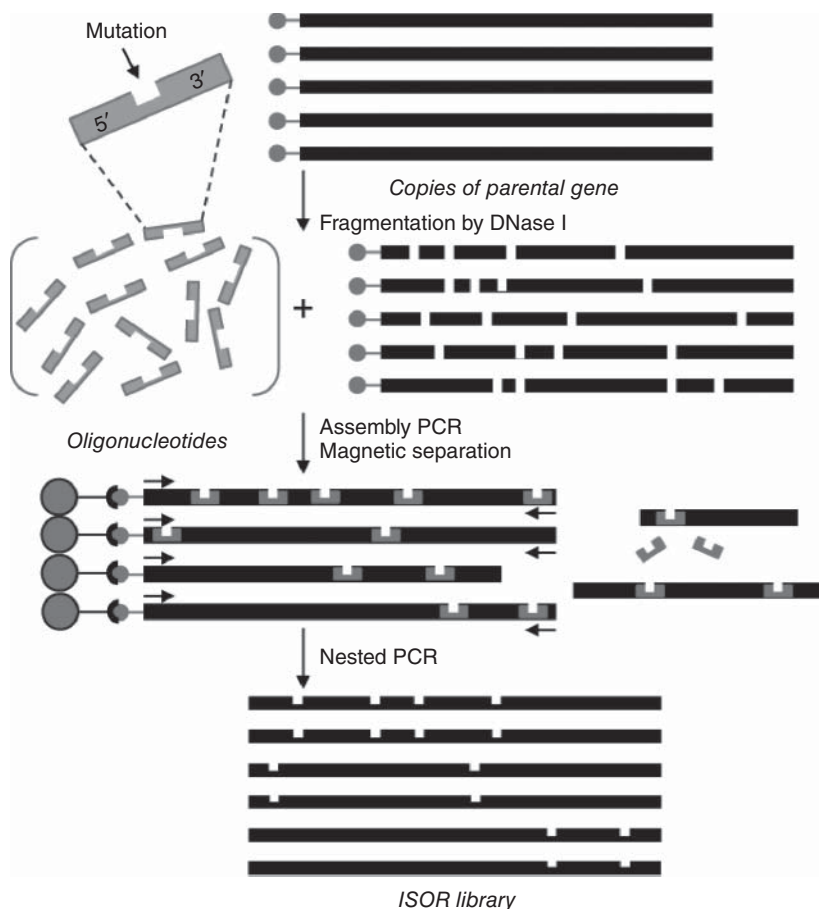
Adding designed synthetic oligonucleotides to a mixture of gene fragments prior to reassembly as a form of semisynthetic DNA shuffling is a straightforward technique, as, for example, in the engineering of the biosynthetic pathway of the natural product Avermectin for pharmaceutical production [95]. It is reminiscent of the Stemmer-method of CMCM [78c], which had been applied in the directed evolution of an enantioselective lipase [26]. The fundamentally important method [95] was later optimized and dubbed “incorporating synthetic oligonucleotides via gene reassembly” (ISOR) (Scheme 3.11) [96]. A biotinylated PCR product of the target gene is subjected to DNase-I-mediated fragmentation, followed by mixing the fragments with a set of synthetic oligonucleotides similar to the CMCM-procedure. After reassembly by self-primed extension catalyzed by Taq-polymerase, the genes are enriched by capture on streptavidin-coated magnetic beads. This is important for maintaining the diversity in the



Scheme 3.10 A schematic example of biased mutation-assembly, assuming a basis set of three mutations [94]. The circle, triangle, and square each represent one mutation. A block represents a portion of the gene containing one mutation and represents a recombination unit. The double-headed

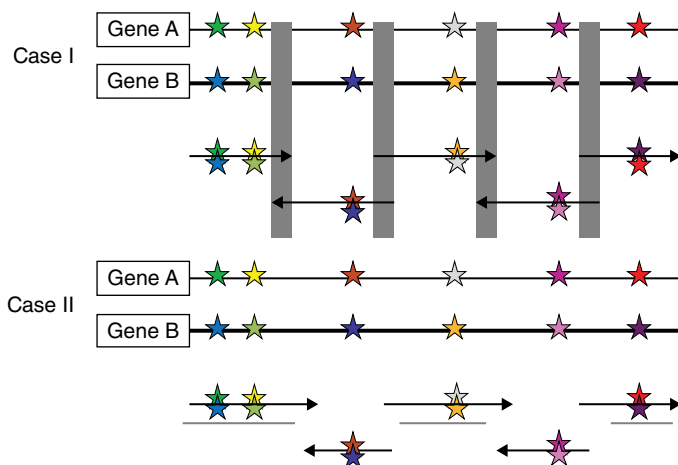
arrows represent overlapping sequences between adjacent blocks and these overlapping sequences hybridize during PCR recombination. (Hamamatsu *et al.* [94]. Reproduced with permission of Oxford University Press.)

assembly process by minimizing mispriming and reducing amplification of short products. ISOR was originally applied to a cytosine-C5 methyltransferase, with 45 individual positions being randomized, and to serum paraoxonase PON1, with insertions and deletions at different sites surrounding the binding pocket [96].



Scheme 3.11 Illustration of ISOR; the use of biotinylated DNA and purification by capture onto streptavidin-coated beads is optional [96]. (Herman and Tawfik [96]. Reproduced with permission of Oxford University Press.)

Hundreds of DNA shuffling patents have been applied for or issued, making collaborative efforts between academic and industrial groups difficult. A promising alternative to DNA gene shuffling is “artificial shuffling,” a process which starts not by fragmenting genes, but by assembling DNA fragments previously designed and generated by other means. Sequence information is used as a guide. Three independent studies initiated this new approach that tolerates low homology and minimizes undesired self-hybridization of parental genes (low WT-appearance in the mutant libraries) [97–99]. In one embodiment called “assembly of designed oligonucleotides” (ADO), two strategies for linking fragments are possible (Scheme 3.12) [99, 100]. The advantages of this particular version of artificial



Scheme 3.12 General concept of ADO [99] with two strategies for the linking of fragments being possible (cases I and II). In case I the two genes A and B to be virtually shuffled are aligned; the different colored stars refer to information that encoded different amino acids, while oligonucleotide fragments with both colored stars in the same position of the parent gene denote the synthetic oligonucleotide fragment with degenerate

nucleotides. The gray blocks denote conserved regions of sequence that can be used as the linking part with homologous recombination. Case II shows no homology between flanking oligos, which can be assembled by ligation between ssDNA with an unknown terminal sequence. (Zha *et al.* [99]. Reproduced with permission of John Wiley & Sons.)

shuffling, which is not patent-protected, were delineated in subsequent studies that include statistical analyses [100, 101].

Thus far, ADO has not been applied very often in directed evolution projects, but with the rapid development of optimized saturation mutagenesis techniques, a combination of both methods could offer new perspectives. Along a different line, the gene assembly step in ADO offers many possibilities as a tool in synthetic biology for creating high quality DNA libraries [99–101]. For a general discussion of modern gene assembly methods in metabolic engineering and synthetic biology, see a review article [101]. The benefits and limitations of ADO as a gene assembly method can be summarized as follows:

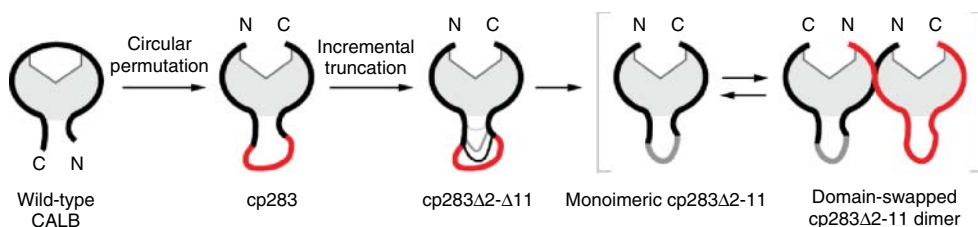
- Fast library construction within hours by changing the primer composition.
- Useful for introducing diversity not just in enzyme-coding genes, but especially when manipulating metabolic pathways and even genomes.
- Rational engineering is possible by controlling the mutagenesis sites.
- Initial costs may be formidable, but it pays off if many libraries are planned.
- Contiguous codons may increase costs if there are no degenerate codons.
- Depending upon the particular system, repetitive sequences may be difficult to assemble without errors.

3.5

Circular Permutation and Other Domain Swapping Techniques

All of the above mutagenesis methods involve random or focused amino acid substitutions. A different possibility is protein engineering by circular permutation, meaning the creation of a different order in the protein sequence in the absence of point mutations [102]. This novel idea was inspired by the structure and function of circularly permuted proteins known to occur in nature, of which more than 2000 have been identified [103]. As a protein engineering technique, two steps are required: (i) covalent linkage of the native N- and C-termini by means of a short and flexible peptide linker, followed by (ii) cleavage of an existing peptide bond, the end result being a shift of the termini elsewhere. Thus, the primary sequence has been rearranged by means of termini relocation. Many studies of this kind have been performed for mechanistic and structural reasons, but few have led to a change in catalytic profile simply because the structure of the binding pocket is generally maintained. In fact, a number of cases are known in which only moderate or even no enhancement of activity occurred [104]. Nevertheless, notable changes in catalytic profile were found in more recent studies [102, 105].

In these model studies, the effect of relocating the termini of the lipase from *Candida antarctica* B (CALB) was investigated [102, 105]. The new locations of the N- and C-termini were designed to occur at positions 283 and 282 of WT CALB, respectively, in hope of influencing local backbone flexibility and possibly active site accessibility. In transesterification reactions using model compounds and vegetable oil, one variant (cp283) showed improved catalyst performance in terms of activity relative to WT [105b]. Using the best variants, the effect of circular permutation on enantioselectivity was also investigated using several chiral substrates [105d]. Stereoselectivity was not compromised, and in some cases slight improvements were actually observed. Thus, it seems that circular permutation is not well suited as a directed evolution method for enhancing or inverting stereoselectivity. Combining the method with loop truncation near the active site may lead to a different picture. This multi-step process was tested with CALB (Scheme 3.13), leading to a change in the enzyme quaternary structure, shifting from a monomeric form to a dimer with domain-swapped N-terminal segment;



Scheme 3.13 Schematic overview of CALB engineering by circular permutation and subsequent incremental truncation of the newly created surface loop in cp283, the most

active variant among the lipase permutants [105e]. (Qian *et al.* [105e]. Reproduced with permission of Elsevier.)

this structural induced enhanced hydrolytic lipase activity, but the effect on enantioselectivity was not reported [105e].

In a comprehensive study that appeared in 2015, the following drawbacks of conventional circular permutation as a gene mutagenesis method for creating libraries of variants were noted [106]: (i) undesired sequence modification at the new termini of circular permutants and (ii) either relatively inefficient blunt-end ligation during library construction or redesign of transposons for optimal expression are required. In their study, the researchers developed an engineered transposon for efficient generation of random circular permutation libraries. The method allows the use of sticky-end ligation and enables possible external tunability for expression of circular permutants. These improvements are significant and go beyond earlier modifications [105], as delineated by the researchers of this study [107].

In summary, traditional circular permutation has been used effectively for more than two decades, *inter alia*, in the study of structure and function of proteins, but its adaptation as a gene mutagenesis method in directed evolution was not realized until much later. Thus far, its role in laboratory evolution in biotechnology has been limited, the reliable control of such parameters as activity, stereoselectivity, regioselectivity, and thermostability being difficult relative to the use of other gene mutagenesis methods such as epPCR, DNA shuffling, or saturation mutagenesis. With the potential emergence of increased predictive power when applying circular permutation in order to change a given catalytic parameter, the method may play a greater role in future directed evolution studies.

Domain swapping (including 3D domain swapping) is the more general approach as evidenced by a number of studies that traditionally focused mainly on questions regarding natural evolution as well as structural and mechanistic aspects of protein science [108]. The technique has also been used to generate hybrid enzymes characterized by novel catalytic profiles. Progress until 2000 was summarized in a review article [109]. Incremental truncation has been included [87]. Since then, further studies have appeared concerning fundamental questions in evolutionary biology, but also aimed at metabolic engineering, where the method is sometimes combined with other mutagenesis techniques such as epPCR or saturation mutagenesis [110a–i]. Domain swapping has also been applied to directed evolution of enantioselective Baeyer–Villiger monooxygenases [110j]. Notwithstanding these important contributions, it remains to be seen whether domain swapping can be developed into a general and reliable method for creating useful enzymes as catalysts in organic chemistry and biotechnology.

3.6

Solid-Phase Combinatorial Gene Synthesis for Library Creation

Two major types of chemical gene syntheses have been developed, polymerase cycling assembly (PCA)-dependent and PCA-independent approaches. Both methods involve recursive cycles of traditional phosphoramidite chemistry

comprising base deprotection, coupling, capping, and oxidation; however, PCA uses PCR [111]. These gene synthesis methods play a central role in synthetic biology and biotechnology in general, especially in the manipulation of proteins, metabolic pathways, and entire genomes [112]. PCA-dependent methods involve, inter alia, PCR-extension, which has some disadvantages as summarized in a review on gene synthesis [113]. PCA-independent techniques utilize the synthesis of oligonucleotides on solid supports followed by assembly catalyzed by enzymes involved in DNA repair and/or ligation rather than polymerases [113]. DNA microarrays have been developed in this respect, allowing the high-throughput synthesis of defined sequences [113, 114]. Most of these advancements utilize the Agilent technology [115]. Unfortunately, these microarray techniques limit the size of fragments to about 200 bases. Two other approaches have emerged, the so-called Blue Heron solid support technology, which has not been used in library construction to date [116] and the Sloning building block technique [117], which was recently exploited in order to create high-quality saturation mutagenesis libraries [73].

The Sloning approach is based on solid phases with bound biotin-modified oligonucleotides. In order to generate a defined gene sequence, a certain number of chemically synthesized building blocks ("splinkers"), which contain self-complementary regions are ligated to the anchor, immobilized, washed, and cleaved with formation of sub-fragments of 18 bp in each cycle. Full-length or large gene fragments are thus accessible in a fully automated manner. Sloning libraries have been prepared in order to incorporate the non-natural amino acid *O*-methyl-L-tyrosine in proteins [118] and in engineering the binding affinity in anticalins [119] and in antibodies [120]. Therefore, it was logical to test Sloning in the directed evolution of enzymes as selective biocatalysts in organic chemistry and biotechnology as an alternative to the traditional PCR-based methods [73].

In this study [73], P450-BM3 was chosen as the enzyme to be optimized by directed evolution, the previously studied regio- and stereoselective oxidative hydroxylation of testosterone [121] serving as the model reaction (Figure 3.9a). Three CAST libraries were designed at Site A (R47/T49/Y51), B (V78/A82), and C (M185/L188) lining or near the binding pocket (Figure 3.9b). The goal was to compare the traditional PCR-based saturation mutagenesis libraries with the respective Sloning libraries. Sites A and C were randomized conventionally using NDC codon degeneracy encoding 12 amino acids (R, D, N, C, G, H, I, L, S, V, F, and Y) (Figure 3.9c). In the Sloning experiments, the codon choice for the same 12 amino acids differed only in the last nucleotide of 5 codons due to a codon optimization algorithm for *E. coli* (Figure 3.9d). The randomization scheme for library B includes all 20 canonical amino acids, which in the case of PCR-based saturation mutagenesis is covered by NNK codon degeneracy. It encodes 32 defined codons, including 1 stop codon, but shows redundancy for amino acids A, G, P, T, V and R, L, S with 2 and 3 codons, respectively. In contrast, Sloning gene synthesis does not involve degenerate codons, which means that full randomization is ensured with a non-redundant set of 20 codons (Figure 3.9d) [73].

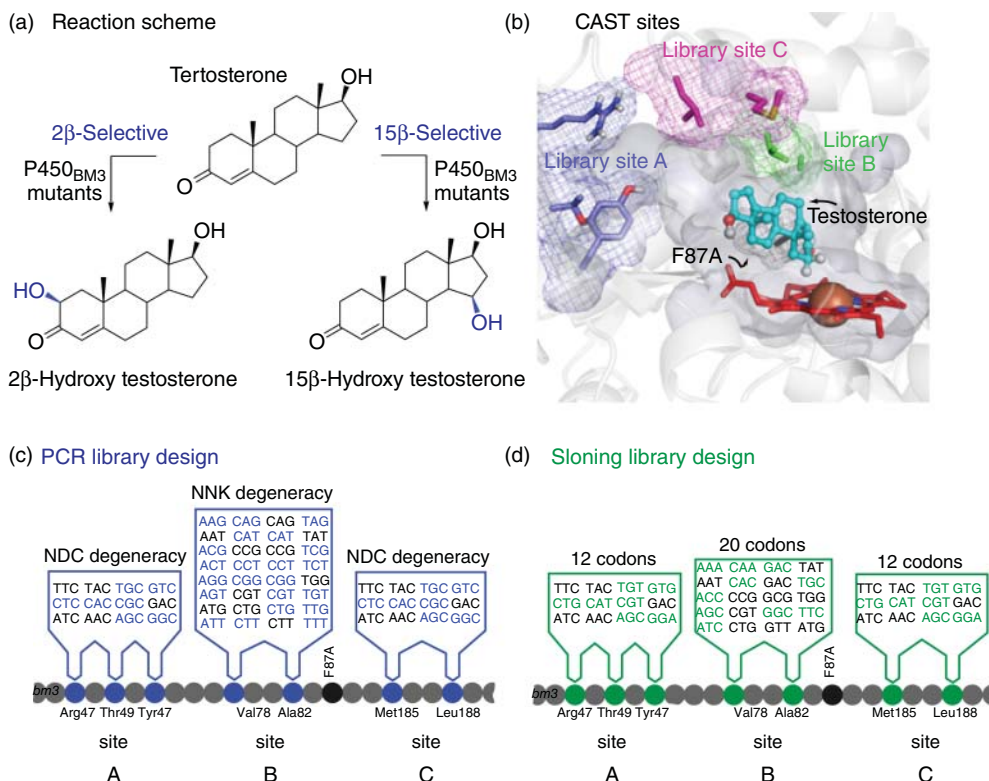


Figure 3.9 Model reaction and library design for comparing traditional PCR-based saturation mutagenesis libraries with Slone libraries [73]. (a) Testosterone hydroxylation by P450_{BM3} mutants. (b) Active site of P450_{BM3} mutant F87A. The three CAST sites and the F87A residue are highlighted. The structure was modeled by docking computations using the Schrödinger software and the picture was created with PyMol. (c) Diversity design of the combinatorial P450_{BM3}-F87A

libraries used in this study. Library A consists of three simultaneously randomized positions, whereas library B and C consist of two. PCR-based libraries use either the nonredundant NDC codon (library A + C) or the redundant NNK codon (library B). (d) Sloning-based libraries encode the same set of amino acids using the displayed codon usages. Gray codons are present in both designs. (Höbenreich *et al.* [73]. Reproduced with permission of American Chemical Society.)

All three PCR-based libraries were generated by the improved megaprimer PCR protocol (Scheme 3.8) [56]. Upon adjusting the annealing temperature and increasing the number of amplification cycles, libraries with sufficient amounts of colonies per transformation (more than 10 000) were obtained. Following pooling [74a] and plasmid extraction, sequencing was performed. As judged by the QQC [74a], acceptable quality was observed. In the case of Sloning, all of the designed libraries were obtained commercially; gene fragments of 683 bp (954 bp with flanking sequences) were cloned into the target plasmid, followed by transformation leading to 5000 clones for library A (A-SLO), 11 000 clones

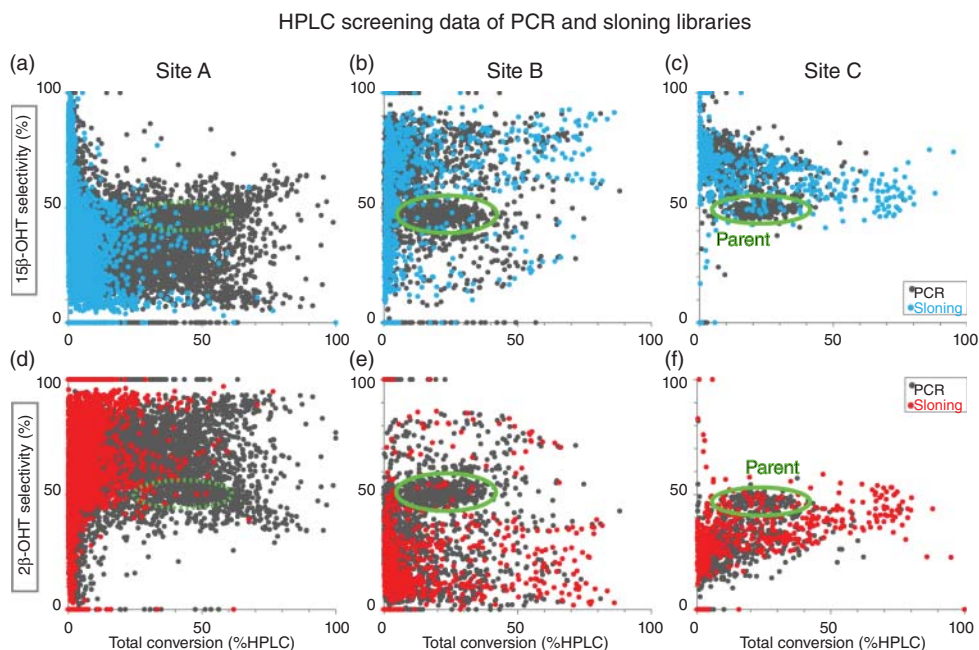


Figure 3.10 Screening results comparing PCR with Sloning libraries [73]. Total testosterone conversion (%HPLC) of the six combinatorial libraries is shown as a function of either 15 β -OHT or 2 β -OHT regioselectivity. Colored entries show the data of the Sloning libraries, while gray entries represent the PCR library results. The green circle highlights a cluster corresponding to parental transformants in PCR libraries. (Höbenreich *et al.* [73]. Reproduced with permission of American Chemical Society.)

for library B (B-SLO), and 5000 clones for library C (C-SLO). All libraries were transformed into *E. coli* BL21-Gold (DE3). Library screening was carried out using automated HPLC. Following random sequencing, statistical analyses were performed. The screening results are shown in Figure 3.10 (for details including regio- and stereoselectivity data, see [73]).

This study shows for the first time that the quality of the synthetic Sloning libraries at the DNA and protein level is notably higher than that of the conventional PCR-based saturation mutagenesis libraries. This result was anticipated, but it had to be proven experimentally. Importantly, due to the higher quality, it correlates with less screening effort. Combined with guidance by the Patrick/Firth [23a,b] or Nov statistical metric [23e, 70], it was demonstrated that the speed of directed evolution can be increased to a significant degree [73]. This gain in efficiency is achieved by the purchase of designed Sloning libraries. Other companies are also engaged in gene synthesis. Currently, the prices are still high, but if the costs continue to go down as expected in the coming years, then the strategy outlined in this study may prove to be a general way to proceed in future directed evolution projects [73]. Optimism along this line is warranted by the recent development of high-throughput on-microchip gene synthesis for the

rapid preparation of potential combinatorial mutant libraries, as, for example, offered by LabGenius Ltd (www.labgeni.us).

3.7

Computational Tools

Achieving maximum efficiency in directed evolution not only requires the use of advanced mutagenesis methods, but also guidance by appropriate computational tools that accompany the experimental work [2, 122]. Depending upon the specific goal of a directed evolution project, different types of computational methods and computer aids are available [122]. Two types of tools have been developed, those that concentrate on the optimal use of degenerate codons, library size, and degree of oversampling, and those that analyze protein structures in the quest to identify functionally important residues (hot spots), which can then be varied by site-specific mutagenesis or randomization by saturation mutagenesis. When designing library diversity and estimating completeness as a function of oversampling, especially in the case of saturation mutagenesis, several metrics are available, including:

- GLUE and GLUE-IT [23a,b] (<http://guinevere.otago.ac.nz/cgi-bin/aef/glue-IT.pl>)
- CASTER [51b] (<http://www.kofo.mpg.de/en/research/biocatalysis>)
- TopLib [70c] (<http://stat.haifa.ac.il/~yuval/toplib/>)
- SwiftLib [123].

CASTER is a user-friendly computer aid which is based on the Patrick/Firth metric GLUE, but which also provides the user with a great deal of additional helpful information such as electronic and steric properties of amino acids, the genetic code, and other useful tips for fast guidance [51b]. It does not tell the user which degenerate codons to use, in contrast to SwiftLib. In the case of the second type of computational tools, several programs are available, which help in identifying and evaluating residues at the protein level that could influence catalytic properties, these hot spots then being positions at which site-specific mutagenesis or saturation mutagenesis can be attempted. Typical software packages include:

- HotSpot Wizzard [124] (<http://loschmidt.chemi.muni.cz/hotspotwizzard/>)
- CASTp [125] (<http://cast.engr.uic.edu/>)
- 3DM [126] (<http://3dmcsis.systemsbiology.nl/>)
- ConSurf [127] (<http://consurf.tau.ac.il/>)
- MBLOSUM [128] (<http://apps.cbu.uib.no/mblosum>)
- Scorecons [129] (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bib/valdar/scorecons_server.pl)
- FamClash [130]
- SIRCH [131]
- IPRO [132]

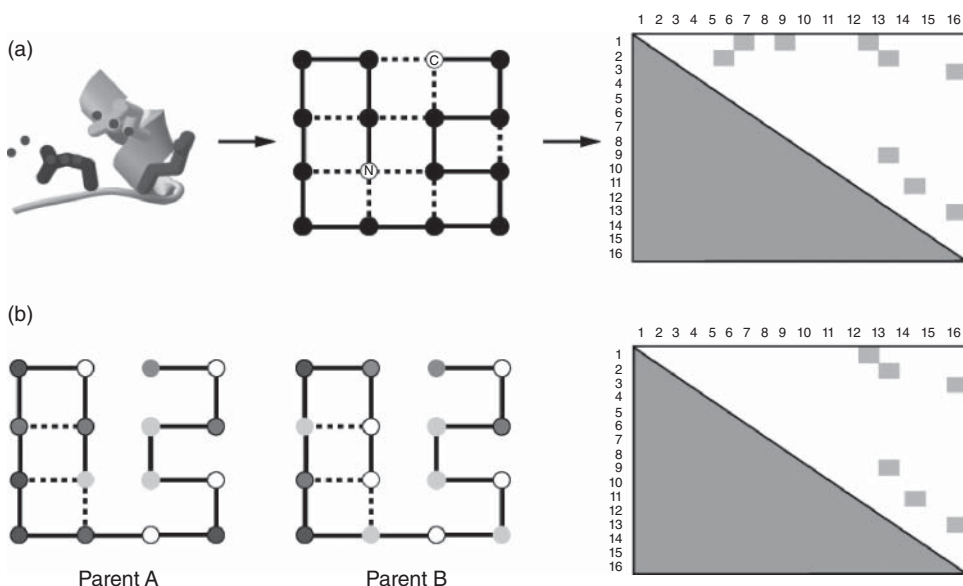
- CorNet [133a,b]
- ANT [133c].

When aiming for enhanced protein thermostability (Chapter 6), the versatile HotSpot Wizzard can be used in addition to the following computational guides:

- B-FITTER [51b] (<http://www.kofo.mpg.de/en/research/biocatalysis>)
- CUPSAT [134] (<http://cupsat.tu-bs.de/>)
- CAVER [135] (<http://www.caver.cz/>)
- FoldXR [136] (<http://foldx.crg.es/>)
- PoPMuSiC [137] (<http://babylone.ulb.ac.be/popmusic/>)
- FRESCO [138]
- FireProt [139].

The optimal choice of the computational tool depends upon the particular problem at hand. Other computational tools for different purposes such as identifying binding sites in proteins have been summarized in reviews [122d,e].

Several other approaches are presented here in more detail because they are characterized by special features, one of them being SCHEMA [140]. It is a structure-guided method utilizing recombinant processes (Scheme 3.14). Blocks of sequences are first identified, which minimize structural disruption when recombination with formation of chimeric proteins occurs. Pairs of interacting

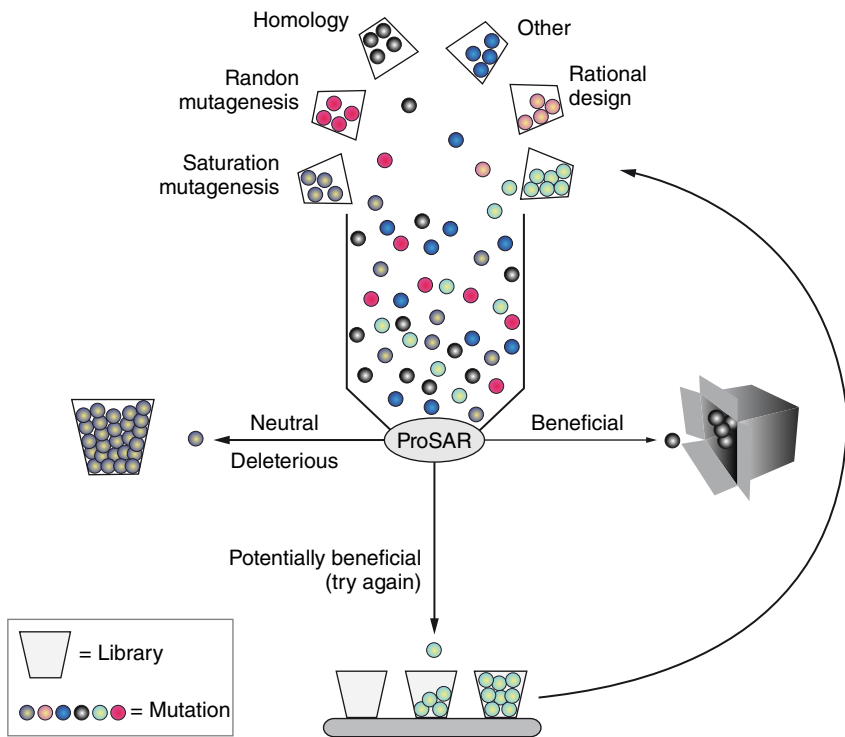


Scheme 3.14 SCHEMA disruption based upon a contact matrix representing interactions between amino acids in the three-dimensional structure of a protein (illustrated here with a simplified model) [140a].

(a) Disruptions in a simplified model and (b) contact matrix to be adjusted for the sequence identity of the parent enzymes. (From [143] with permission; ©2011 Wiley-VCH Verlag.)

residues within 4.5 Å of each other are identified, which are then utilized as a basis for constructing contact matrices. The program includes an optimization algorithm that selects optimal crossovers that minimize the average disruption of the library. All interactions that break upon recombination contribute to a disruption score, which is used in designing shuffling experiments. In order to compute the average disruption, a high-resolution structural data of at least one of the proteins is required. SCHEMA has been applied to P450 monooxygenases [140a], β -lactamases [140b], and cellulases [140c]. A type of hybrid of FamClash and SCHEMA has been claimed to be particularly efficient [141].

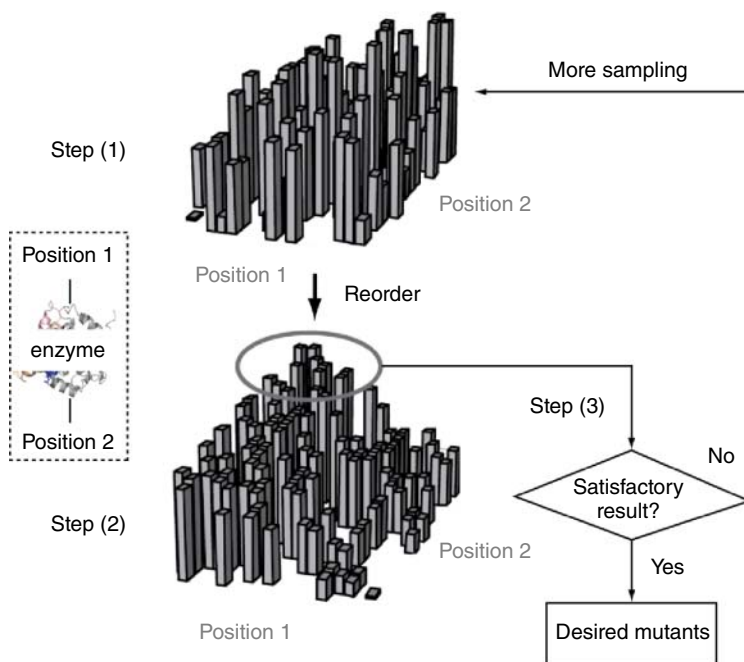
In analogy to quantitative structure-activity relationships (QSARs) used in therapeutic drug discovery, an algorithm based on protein sequence-activity relationships (ProSAR) has been developed by an industrial group for application in directed evolution [142]. As each evolutionary cycle is traversed, information from sequence-activity data is utilized. The best mutant is then used as a template for programming diversity in the subsequent round by inferring the contributions of mutational effects on the catalytic profile of the enzyme (Scheme 3.15). At any point, about 50 mutations as variables are evaluated in the combinatorial libraries, and the hits as well as some less improved mutants are sequenced. As part of the ProSAR analysis, mutations are assigned to four classes: (i) beneficial,



Scheme 3.15 Formal representation of ProSAR [142, 143].

which are fixed in the population in the next round; (ii) potentially beneficial, which are sent back for retesting; (iii) deleterious, which are discarded; and (iv) neutral, which are also discarded. Diversity in addition to shuffling needs to be increased by applying other mutagenesis methods such as rational design or saturation mutagenesis. ProSAR has been applied in the directed evolution of several enzymes, activity, and stereoselectivity being the catalytic parameters of interest. Examples include a halohydrin dehalogenase as the catalyst in the production of the cholesterol-lowering drug Lipitor [142] and a transaminase in the production of the anti-diabetic drug Sitagliptin [143]. The latter study utilizes, *inter alia*, ISM; the multi-step evolutionary process is featured in detail in Chapter 4.

Another computational tool for directed evolution is based on the adaptive substituent reordering algorithm (ASRA) [144], specifically when applied to ISM [145]. It was tested experimentally in the enhancement of enantioselectivity of the ANEH as the catalyst in the hydrolytic kinetic resolution of a chiral epoxide. Two substitution positions, *inter alia*, were chosen as mutation targets and each amino acid on every position was assigned a random distinct integer between 1 and 20, the total number of possible mutants being 400 in this case. In step 1, a small subset of the 400 mutants containing substitutions on both positions were synthesized and their respective enantioselectivity (*E*-value) measured (Scheme 3.16). Due to the random integer assignment, the initial selectivity



Scheme 3.16 Steps when applying ASRA to directed evolution [145]. (Feng *et al.* [145]. Reproduced with permission of John Wiley & Sons.)

landscape is irregular and provides no predictive power. In step 2, the optimal integer assignment for each amino acid at each position was identified so that the property landscape is as regular as possible. It should be noted that when an amino acid at position 1 (or position 2) “moves,” meaning that its integer assignment is changed from a certain value to another, all 20 amino acids on the other position will move along with it in order to maintain consistent indexing. In step 3, the location of the best mutants was predicted based on the geometric features of the re-ordered selectivity landscape. For example, the circle should be a desired

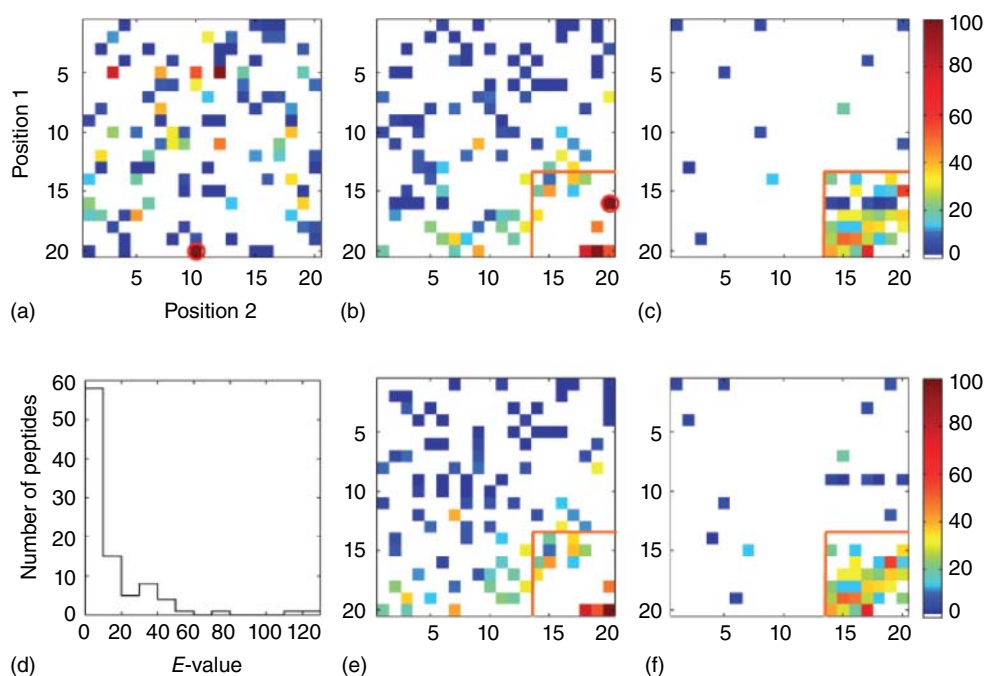


Figure 3.11 Optimal reordering of the *E*-value enantioselectivity landscapes with 60 min reaction time [145]. (a) Color heat map for the enantioselectivity landscape (*E*-values) of 95 randomly sampled mutants plotted with a random amino acid ordering. Each color square represents one mutant with red indicating a high *E*-value and blue corresponding to a low *E*-value (see color bar on the far right). White squares are unsampled proteins. (b) *E*-value landscape of the 95 mutants using the ASRA-identified optimal amino acid ordering. The result predicts that proteins with high *E*-values are most likely located in the lower right corner. The mutant

at position 16/20 (circled in red in both (a) and (b)) of the reordered landscape turned out to be the same as the mutant at position 20/19; the wrong protein was accidentally placed in this position in the experiment. (c) *E*-value landscape for 45 newly sampled mutants, guided by the ordering in (b). (d) *E* value distribution for the 95 initial random mutants. (e) Reordered *E*-value landscape for the 94 mutants (excluding the erroneous mutant at position 16/20 in (b)). (f) *E*-value landscape for the 45 newly sampled mutants, based on the ordering in the enantioselectivity factor *E*. (Feng *et al.* [145]. Reproduced with permission of John Wiley & Sons.)

area because of the monotonic landscape geometry. The identified mutants are relevant in the next ISM round (return to step 1) [145].

A typical result of re-ordering is shown in Figure 3.11. When applying ASRA, no assumptions regarding linearity, additivity or any structure-property relationship are made. Rather, the algorithm allows the identification of the underlying regularity of the protein-property landscape, in the present case stereoselectivity.

Finally, the Rosetta algorithms [146] provide a computational means to access, inter alia, rationally designed enzymes showing promiscuous catalytic behavior which can be improved by directed evolution in subsequent steps (see Chapter 7).

References

1. (a) Selifonova, O. and Schellenberger, V. (2003) in *Directed Evolution Library Creation: Methods and Protocols*, Methods in Molecular Biology, vol. **231** (eds F.H. Arnold and G. Georgiou), Humana Press Inc., Totowa, NJ, pp. 45–52; (b) Muteeb, G. and Sen, R. (2010) Random mutagenesis using a mutator strain. *Methods Mol. Biol.*, **634**, 411–419; (c) Carr, R., Alexeeva, M., Enright, A., Eve, T.S.C., Dawson, M.J., and Turner, N.J. (2003) Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angew. Chem. Int. Ed.*, **42** (39), 4807–4810; (d) Serero, A., Jubin, C., Loeillet, S., Legoix-Né, P., and Nicolas, A.G. (2014) Mutational landscape of yeast mutator strains. *Proc. Natl. Acad. Sci. U.S.A.*, **111** (5), 1897–1902; (e) Agilent Technologies, Inc. (2015) XL1-Red Competent Cells, Agilent Instruction Manual, Catalog 200129.
2. Recent reviews of directed evolution of enzymes: (a) Gillam, E.M.J., Copp, J.N., and Ackersley, D.F. (eds) (2014) *Directed Evolution Library Creation*, Methods in Molecular Biology, Humana Press, Totowa, NJ; (b) Bommarius, A.S. (2015) Biocatalysis, a status report. *Annu. Rev. Chem. Biol. Eng.*, **6**, 319–345; (c) Jäckel, C. and Hilvert, D. (2010) Biocatalysts by evolution. *Curr. Opin. Biotechnol.*, **21** (6), 753–759; (d) Brustad, E.M. and Arnold, F.H. (2011) Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.*, **15** (2), 201–210; (e) Goldsmith, M. and Tawfik, D.S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.*, **22** (4), 406–412; (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) Widersten, M. (2014) Protein engineering for development of new hydrolytic biocatalysts. *Curr. Opin. Chem. Biol.*, **21**, 42–47; (h) Denard, C.A., Ren, H., and Zhao, H. (2015) Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.*, **25**, 55–64; (i) 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; (j) Kumar, A. and Singh, S. (2013) Directed evolution: tailoring biocatalysts for industrial applications. *Crit. Rev. Biotechnol.*, **33**, 356–378; (k) Lutz, S. and Bornscheuer, U.T. (eds) (2009) *Protein Engineering Handbook*, Wiley-VCH Verlag GmbH, Weinheim.
3. (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**, S136–S140.
4. Jiang, W., Zhuang, Y., Wang, S., and Fang, B. (2015) Directed evolution

- and resolution mechanism of 1,3-propanediol oxidoreductase from *Klebsiella pneumoniae* toward higher activity by error-prone PCR and bioinformatics. *PLoS One*, **10**, e0141837.
5. McCullum, E., Williams, B.R., Zhang, J., and Chaput, J. (2010) Random mutagenesis by error-prone PCR. *Methods Mol. Biol.*, **634**, 103–109.
 6. Kaur, J. and Sharma, R. (2006) Directed evolution: an approach to engineer enzymes. *Crit. Rev. Biotechnol.*, **26** (3), 165–199.
 7. Zacco, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.*, **255** (4), 589–603.
 8. Abou-Nader, M. and Benedik, M.J. (2010) Rapid generation of random mutant libraries. *Bioeng. Bugs*, **1** (5), 337–340.
 9. Tee, K.L. and Wong, T.S. (2013) Polishing the craft of genetic diversity creation in directed evolution. *Biotechnol. Adv.*, **31** (8), 1707–1721.
 10. Cirino, P., Mayer, K., and Umeno, D. (2003) in *Directed Evolution Library Creation*, vol. **231** (eds F. Arnold and H. Georgiou), Humana Press, pp. 3–9.
 11. 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.
 12. (a) Neylon, C. (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res.*, **32** (4), 1448–1459; (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.
 13. (a) Weiss, G. and von Haeseler, A. (1995) Modeling the polymerase chain reaction. *J. Comput. Biol.*, **2**, 49–61; (b) Moore, G.L. and Maranas, C.D. (2000) Modeling DNA mutation and recombination for directed evolution experiments. *J. Theor. Biol.*, **205** (3), 483–503; (c) Pritchard, L., Corne, D., Kell, D., Rowland, J., and Winson, M. (2005) A general model of error-prone PCR. *J. Theor. Biol.*, **234** (4), 497–509; (d) Sylvestre, J., Chautard, H., Cedrone, F., and Delcourt, M. (2006) Directed evolution of biocatalysts. *Org. Process Res. Dev.*, **10**, 562–571; (e) Wong, T.S., Roccato, D., and Schwaneberg, U. (2007) Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatal. Biotransform.*, **25**, 229–241; (f) Zhao, J., Kardashliev, T., Joëlle Ruff, A., Bocola, M., and Schwaneberg, U. (2014) Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations. *Biotechnol. Bioeng.*, **111** (12), 2380–2389.
 14. Claveau, S., Sasseville, M., and Beauregard, M. (2004) Alcohol-mediated error prone PCR. *DNA Cell Biol.*, **23**, 789–795.
 15. Minamoto, T., Wada, E., and Shimizu, I. (2012) A new method for random mutagenesis by error-prone polymerase chain reaction using heavy water. *J. Biotechnol.*, **157** (1), 71–74.
 16. Vanhercke, T., Ampe, C., Tirry, L., and Denolf, P. (2005) Reducing mutational bias in random protein libraries. *Anal. Biochem.*, **339**, 9–14.
 17. (a) Biles, B.D. and Connolly, B.A. (2004) Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res.*, **32** (22), e176; (b) Kardashliev, T., Ruff, A., Zhao, J., and Schwaneberg, U. (2014) A high-throughput screening method to reengineer DNA polymerases for random mutagenesis. *Mol. Biotechnol.*, **56** (3), 274–283.
 18. Vartanian, J.P., Henry, M., and Wain-Hobson, S. (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.*, **24** (14), 2627–2631.
 19. (a) Miyazaki, K. and Takenouchi, M. (2002) Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. *Biotechniques*, **33** (5), 1033–1034, 1036–1038; (b) Miyazaki, K. (2003) Creating random mutagenesis libraries by megaprimer PCR of whole

- plasmid (MEGAWHOP). *Methods Mol. Biol.*, **231**, 23–28; (c) Miyazaki, K. (2011) MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods Enzymol.*, **498**, 399–406.
20. (a) Yu, H., Li, J., Zhang, D., Yang, Y., Jiang, W., and Yang, S. (2009) Improving the thermostability of N-carbamyl-D-amino acid amidohydrolase by error-prone PCR. *Appl. Microbiol. Biotechnol.*, **82** (2), 279–285; (b) Gleichmann, T., Diensthuber, R.P., and Möglich, A. (2013) Charting the signal trajectory in a light-oxygen-voltage photoreceptor by random mutagenesis and covariance analysis. *J. Biol. Chem.*, **288** (41), 29345–29355; (c) Liang, C., Gui, X., Zhou, C., Xue, Y., Ma, Y., and Tang, S.-Y. (2015) Improving the thermoactivity and thermostability of pectate lyase from *Bacillus pumilus* for ramie degumming. *Appl. Microbiol. Biotechnol.*, **99** (6), 2673–2682; (d) Jiang, P., Mu, S., Li, H., Li, Y., Feng, C., Jin, J.-M., and Tang, S.-Y. (2015) Design and application of a novel high-throughput screening technique for 1-deoxynojirimycin. *Sci. Rep.*, **5**, 8563.
 21. (a) Zhang, D., Zhu, F., Fan, W., Tao, R., Yu, H., Yang, Y., Jiang, W., and Yang, S. (2011) Gradually accumulating beneficial mutations to improve the thermostability of N-carbamoyl-D-amino acid amidohydrolase by step-wise evolution. *Appl. Microbiol. Biotechnol.*, **90** (4), 1361–1371; (b) Sandström, A.G., Wikmark, Y., Engström, K., Nyhlén, J., and Bäckvall, J.-E. (2012) Combinatorial reshaping of the *Candida antarctica* lipase a substrate pocket for enantioselectivity using an extremely condensed library. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (1), 78–83; (c) Saß, S., Kadow, M., Geitner, K., Thompson, M.L., Talmann, L., Böttcher, D., Schmidt, M., and Bornscheuer, U.T. (2012) A high-throughput assay method to quantify Baeyer–Villiger monooxygenase activity. *Tetrahedron*, **68** (37), 7575–7580; (d) Agudo, R., Roiban, G.D., and Reetz, M.T. (2012) Achieving regio- and enantioselectivity of P450-catalyzed oxidative CH activation of small functionalized molecules by structure-guided directed evolution. *ChemBioChem*, **13** (10), 1465–1473.
 22. (a) Cirino, P.C., Mayer, K.M., and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods Mol. Biol.*, **231**, 3–22; (b) 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; (c) Bill, A., Rosethorne, E.M., Kent, T.C., Fawcett, L., Burchell, L., van Diepen, M.T., Marelli, A., Batalov, S., Miraglia, L., Orth, A.P., Renaud, N.A., Charlton, S.J., Gosling, M., Gaither, L.A., and Groot-Kormelink, P.J. (2014) High throughput mutagenesis for identification of residues regulating human prostacyclin (hiP) receptor expression and function. *PLoS One*, **9** (6), e97973.
 23. (a) Firth, A.E. and Patrick, W.M. (2005) Statistics of protein library construction. *Bioinformatics*, **21**, 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; (c) 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; (d) Bosley, A.D. and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.*, **22**, 57–61; (e) Nov, Y. (2012) When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.*, **78** (1), 258–262.
 24. Reetz, M.T., Zonta, A., Schimossek, K., Jaeger, K.-E., and Liebeton, K. (1997) Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. *Angew. Chem. Int. Ed. Engl.*, **36** (24), 2830–2832.
 25. (a) Arnold, F.H. (1998) Enzyme engineering reaches the boiling point.

- Proc. Natl. Acad. Sci. U.S.A.*, **95** (5), 2035–2036; (b) Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., and Arnold, F.H. (2005) Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.*, **102** (3), 606–611; (c) Tracewell, C.A., and Arnold, F.H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.*, **13** (1), 3–9.
26. Reetz, M.T., Wilensek, S., Zha, D., and Jaeger, K.-E. (2001) Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angew. Chem. Int. Ed.*, **40** (19), 3589–3591.
 27. Zha, S., Wilensek, S., Hermes, M., and Jaeger, K.-E. (2001) Complete reversal of an enzyme-catalyzed reaction by directed evolution. *Chem. Commun.*, 2664–2665.
 28. (a) Zacco, M. and Gherardi, E. (1999) The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 β -lactamase. *J. Mol. Biol.*, **285** (2), 775–783; (b) Daugherty, P.S., Chen, G., Iverson, B.L., and Georgiou, G. (2000) Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl. Acad. Sci. U.S.A.*, **97** (5), 2029–2034.
 29. Drummond, D.A., Iverson, B.L., Georgiou, G., and Arnold, F.H. (2005) Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.*, **350** (4), 806–816.
 30. (a) Holland, E.G., Buhr, D.L., Acca, F.E., Alderman, D., Bovat, K., Busygina, V., Kay, B.K., Weiner, M.P., and Kiss, M.M. (2013) AXM mutagenesis: an efficient means for the production of libraries for directed evolution of proteins. *J. Immunol. Methods*, **394**, 55–61; (b) Holland, E.G., Acca, F.E., Belanger, K.M., Bylo, M.E., Kay, B.K., Weiner, M.P., and Kiss, M.M. (2015) In vivo elimination of parental clones in general and site-directed mutagenesis. *J. Immunol. Methods*, **417**, 67–75.
 31. (a) Fujii, R., Kitaoka, M., and Hayashi, K. (2006) Error-prone rolling circle amplification: the simplest random mutagenesis protocol. *Nat. Protoc.*, **1**, 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**, e145; (c) Fujii, W., Kano, K., Sugiura, K., and Naito, K. (2013) Repeatable construction method for engineered zinc finger nuclease based on overlap extension PCR and TA-cloning. *PLoS One*, **8** (3), e59801; (d) Fujii, R., Kitaoka, M., and Hayashi, K. (2014) Error-prone rolling circle amplification greatly simplifies random mutagenesis. *Methods Mol. Biol.*, **1179**, 23–29.
 32. (a) Wong, T.S., Tee, K.L., Hauer, B., and Schwaneberg, U. (2004) Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.*, **32** (3), e26; (b) Wong, T.S., Roccatano, D., Loakes, D., Tee, K.L., Schenk, A., Hauer, B., and Schwaneberg, U. (2008) Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. *Biotechnol. J.*, **3**, 74–82; (c) Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. (2009) Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.*, **13**, 19–25; (d) Ruff, A.J., Kardashliev, T., Dennig, A., and Schwaneberg, U. (2014) The sequence saturation mutagenesis (SeSaM) method. *Methods Mol. Biol.*, **1179**, 45–68.
 33. (a) Ali, M.M., Li, F., Zhang, Z., Zhang, K., Kang, D.-K., Ankrum, J.A., Le, X.C., and Zhao, W. (2014) Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.*, **43** (10), 3324–3341; (b) Kobori, T. and Takahashi, H. (2014) Expanding possibilities of rolling circle amplification as a biosensing platform. *Anal. Sci.*, **30** (1), 5–64.
 34. Shahsavarian, M.A., Le Minoux, D., Matti, K.M., Kaveri, S.,

- Lacroix-Desmazes, S., Boquet, D., Friboulet, A., Avasle, B., and Padiolleau-Lefèvre, S. (2014) Exploitation of rolling circle amplification for the construction of large phage-display antibody libraries. *J. Immunol. Methods*, **407**, 26–34.
35. Huovinen, T., Brockmann, E.-C., Akter, S., Perez-Gamarra, S., Ylä-Pelto, J., Liu, Y., and Lamminmäki, U. (2012) Primer extension mutagenesis powered by selective rolling circle amplification. *PLoS One*, **7** (2), e31817.
 36. Koyanagi, T., Yoshida, E., Minami, H., Katayama, T., and Kumagai, H. (2008) A rapid, simple, and effective method of constructing a randomly mutagenized plasmid library free from ligation. *Biosci. Biotechnol., Biochem.*, **72** (4), 1134–1137.
 37. Sondek, J. and Shortle, D. (1992) A general strategy for random insertion and substitution mutagenesis: substoichiometric coupling of trinucleotide phosphoramidites. *Proc. Natl. Acad. Sci. U.S.A.*, **89** (8), 3581–3585.
 38. Murakami, H., Hohsaka, T., and Sisido, M. (2002) Random insertion and deletion of arbitrary number of bases for codonbased random mutation of DNAs. *Nat. Biotechnol.*, **20**, 76–81.
 39. (a) Heim, R. and Tsien, R.Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.*, **6** (2), 178–182; (b) Tsien, R.Y. (2009) Constructing and exploiting the fluorescent protein paintbox (Nobel lecture). *Angew. Chem. Int. Ed.*, **48** (31), 5612–5626.
 40. Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, **173** (1), 33–38.
 41. (a) Budisa, N. (2013) Expanded genetic code for the engineering of ribosomally synthesized and post-translationally modified peptide natural products (RIPPS). *Curr. Opin. Biotechnol.*, **24** (4), 591–598; (b) Lang, K. and Chin, J.W. (2014) Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.*, **114** (9), 4764–4806; (c) Sun, S.B., Schultz, P.G., and Kim, C.H. (2014) Therapeutic applications of an expanded genetic code. *ChemBioChem*, **15** (12), 1721–1729.
 42. Zheng, S. and Kwon, I. (2012) Manipulation of enzyme properties by noncanonical amino acid incorporation. *Biotechnol. J.*, **7** (1), 47–60.
 43. (a) Pikkemaat, M.G. and Janssen, D.B. (2002) Generating segmental mutations in haloalkane dehalogenase: a novel part in the directed evolution toolbox. *Nucleic Acids Res.*, **30**, e35; (b) Jones, D.D. (2005) Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β -lactamase to an amino acid deletion. *Nucleic Acids Res.*, **33**, e80.
 44. Bocola, M., Schulz, F., Leca, F., Vogel, A., Fraaije, M.W., and Reetz, M.T. (2005) Converting phenylacetone monooxygenase into phenylcyclohexanone monooxygenase by rational design: towards practical Baeyer–Villiger monooxygenases. *Adv. Synth. Catal.*, **347** (7–8), 979–986.
 45. Erdogan, A., Jones, R.J., Matzlin, P., Hanna, M.H., Smith, S.M.E., and Salerno, J.C. (2005) A novel mutagenesis method generating high yields of closed circular mutant DNA with one primer per mutant. *Mol. Biotechnol.*, **30**, 21–30.
 46. Miyazaki, K. and Arnold, F.H. (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J. Mol. Evol.*, **49** (6), 716–720.
 47. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., and Jaeger, K.-E. (2000) Directed evolution of an enantioselective lipase. *Chem. Biol.*, **7** (9), 709–718.
 48. 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.
 49. (a) Steipe, B. (2004) Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. *Methods Enzymol.*, **388**,

- 176–186; (b) Lehmann, M., Loch, C., Middendorff, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng. Des. Sel.*, **15** (5), 403–411; (c) Polizzi, K.M., Chaparro-Riggers, J.F., Vazquez-Figueroa, E., and Bommarius, A.S. (2006) Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Biotechnol. J.*, **1** (5), 531–536; (d) Reetz, M.T. and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chem. Commun.*, (43), 5499–5501.
50. Wu, S., Acevedo, J.P., and Reetz, M.T. (2010) Induced allostery in the directed evolution of an enantioselective Baeyer–Villiger monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.*, **107** (7), 2775–2780.
 51. (a) Reetz, M.T., Carballeira, J.D., and Vogel, A. (2006) Iterative saturation mutagenesis on the basis of B-factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.*, **45** (46), 7745–7751; (b) 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; (c) Reetz, M.T., Soni, P., Fernandez, L., Gumulya, Y., and Carballeira, J.D. (2010) Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem. Commun.*, **46**, 8657–8658.
 52. 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, 1162, 1164–1155.
 53. Sarkar, G. and Sommer, S.S. (1990) The “megaprimer” method of site-directed mutagenesis. *Biotechniques*, **8** (4), 404–407.
 54. (a) 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; (b) 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; (c) Heckman, K.L. and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.*, **2** (4), 924–932.
 55. Reikofski, J. and Tao, B.Y. (1992) Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.*, **10** (4), 535–547.
 56. Sanchis, J., Fernández, L., Carballeira, J.D., Drone, J., Gumulya, Y., Höbenreich, H., Kahakeaw, D., Kille, S., Lohmer, R., Peyralans, J.P., Podtetenieff, J., Prasad, S., Soni, P., Taglieber, A., Wu, S., Zilly, F., and Reetz, M. (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. *Appl. Microbiol. Biotechnol.*, **81** (2), 387–397.
 57. (a) Kirsch, R.D. and Joly, E. (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.*, **26** (7), 1848–1850; (b) Zheng, L., Baumann, U., and Reymond, J.-L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32** (14), e115.
 58. (a) Ono, A., Matsuda, A., Zhao, J., and Santi, D.V. (1995) The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis. *Nucleic Acids Res.*, **23** (22), 4677–4682; (b) Neuner, P., Cortese, R., and Monaci, P. (1998) Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res.*, **26** (5), 1223–1227; (c) Gaytán, P. and Roldán-Salgado, A. (2013) Elimination of redundant and stop codons during the chemical synthesis of degenerate oligonucleotides. Combinatorial testing on the chromophore region of the red fluorescent protein mKate. *ACS Synth. Biol.*, **2** (8), 453–462.

59. Hughes, M.D., Nagel, D.A., Santos, A.F., Sutherland, A.J., and Hine, A.V. (2003) Removing the redundancy from randomised gene libraries. *J. Mol. Biol.*, **331** (5), 973–979.
60. Ashraf, M., Frigotto, L., Smith, M.E., Patel, S., Hughes, M.D., Poole, A.J., Hebaishi, H.R., Ullman, C.G., and Hine, A.V. (2013) Proximax randomization: a new technology for non-degenerate saturation mutagenesis of contiguous codons. *Biochem. Soc. Trans.*, **41** (Pt. 5), 1189–1194.
61. Dennig, A., Shivange, A.V., Marienhagen, J., and Schwaneberg, U. (2011) OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One*, **6**, e26222.
62. Ke, S.H. and Madison, E.L. (1997) Rapid and efficient site-directed mutagenesis by single-tube ‘megaprimer’ PCR method. *Nucleic Acids Res.*, **25** (16), 3371–3372.
63. (a) Hidalgo, A., Schließmann, A., Molina, R., Hermoso, J., and Bornscheuer, U.T. (2008) A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution. *Protein Eng. Des. Sel.*, **21** (9), 567–576; (b) Hidalgo, A., Schließmann, A., and Bornscheuer, U.T. (2014) One-pot simple methodology for cassette randomization and recombination for focused directed evolution. *Methods Mol. Biol.*, **1179**, 207–212.
64. Ema, T., Kamata, S., Takeda, M., Nakano, Y., and Sakai, T. (2010) Rational creation of mutant enzyme showing remarkable enhancement of catalytic activity and enantioselectivity toward poor substrates. *Chem. Commun.*, **46** (30), 5440–5442.
65. 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.
66. Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D., and Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed.*, **44** (27), 4192–4196.
67. (a) Reetz, M.T., Wang, L.-W., and Bocola, M. (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. Ed.*, **45** (8), 1236–1241; (b) Reetz, M.T. (2005) Evolution im Reagenzglas: Neue Perspektiven für die Weiße Biotechnologie. *Tätigkeitsberichte der Max-Planck-Gesellschaft*, pp. 327–331.
68. Acevedo-Rocha, C., Höbenreich, S., and Reetz, M.T. (2014) Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol. Biol.*, **1179**, 103–128.
69. (a) Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution. *ChemBioChem*, **9** (11), 1797–1804; (b) Clouthier, C.M., Kayser, M.M., and Reetz, M.T. (2006) Designing new Baeyer–Villiger monooxygenases using restricted CASTing. *J. Org. Chem.*, **71** (22), 8431–8437.
70. (a) Nov, Y. (2013) Fitness loss and library size determination in saturation mutagenesis. *PLoS One*, **8** (7), e68069; (b) Nov, Y. (2014) Probabilistic methods in directed evolution: library size, mutation rate, and diversity. *Methods Mol. Biol.*, **1179**, 261–278; (c) Nov, Y., Fulton, A., and Jaeger, K.-E. (2013) Optimal scanning of all single-point mutants of a protein. *J. Comput. Biol.*, **20**, 990–997.
71. Review of directed evolution of stereoselective enzymes with emphasis on iterative saturation mutagenesis: 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.
72. (a) Regan, L. and DeGrado, W. (1988) Characterization of a helical protein designed from first principles. *Science*, **241** (4868), 976–978; (b) Osawa, S., Jukes, T.H., Watanabe, K., and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, **56**

- (1), 229–264; (c) Davidson, A.R., Lumb, K.J., and Sauer, R.T. (1995) Cooperatively folded proteins in random sequence libraries. *Nat. Struct. Mol. Biol.*, **2** (10), 856–864; (d) Walter, K.U., Vamvaca, K., and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.*, **280** (45), 37742–37746.
73. Höbenreich, S., Zilly, F.E., Acevedo-Rocha, C.G., Zilly, M., and Reetz, M.T. (2015) Speeding up directed evolution: combining the advantages of solid-phase combinatorial gene synthesis with statistically guided reduction of screening effort. *ACS Synth. Biol.*, **4** (3), 317–331.
74. (a) Bougioukou, D.J., Kille, S., Taglieber, A., and Reetz, M.T. (2009) Directed evolution of an enantioselective enoate-reductase: testing the utility of iterative saturation mutagenesis. *Adv. Synth. Catal.*, **351** (18), 3287–3305; (b) Sullivan, B., Walton, A.Z., and Stewart, J.D. (2013) Library construction and evaluation for site saturation mutagenesis. *Enzyme Microb. Technol.*, **53** (1), 70–77.
75. (a) Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M., and Jiang, R. (2012) Construction of “small-intelligent” focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques*, **52** (3), 149–158; (b) Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.G., Opperman, D.J., Reetz, M.T., and Acevedo, J.P. (2013) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.*, **2** (2), 83–92.
76. Acevedo-Rocha, C.G., Reetz, M.T., and Nov, Y. (2015) Economical analysis of saturation mutagenesis experiments. *Sci. Rep.*, **5**, 10654.
77. (a) Xiao, H., Nasertorabi, F., Choi, S.-H., Han, G.W., Reed, S.A., Stevens, R.C., and Schultz, P.G. (2015) Exploring the potential impact of an expanded genetic code on protein function. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6961–6966; (b) Hoesl, M.G. and Budisa, N. (2011) In vivo incorporation of multiple noncanonical amino acids into proteins. *Angew. Chem. Int. Ed.*, **50**, 2896–2902; (c) Chin, J.W. (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.*, **83**, 379–408; (d) Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L., Huber, R., and Budisa, N. (2003) Expansion of the genetic code enables design of a novel “gold” class of green fluorescent proteins. *J. Mol. Biol.*, **328** (5), 1071–1081; (e) Kwon, I. and Lim, S.I. (2013) Non-natural amino acids for protein engineering and new protein chemistries. *Macromol. Chem. Phys.*, **214**, 1295–1301.
78. (a) Stemmer, W.P.C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370** (6488), 389–391; (b) 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; (c) Crameri, A. and Stemmer, W.P.C. (1995) Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wild-type sequences. *Biotechniques*, **18**, 194–196.
79. (a) Joern, J.M., Meinhold, P., and Arnold, F.H. (2002) Analysis of shuffled gene libraries. *J. Mol. Biol.*, **316**, 643–656; (b) Joern, J. (2003) in *Directed Evolution Library Creation*, vol. **231** (eds F. Arnold and G. Georgiou), Humana Press, pp. 85–89.
80. (a) Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) Novel family shuffling methods for the in vitro evolution of enzymes. *Gene*, **236** (1), 159–167; (b) Kikuchi, M., Ohnishi, K., and Harayama, S. (2000) An effective family shuffling method using single-stranded DNA. *Gene*, **243** (1–2), 133–137; (c) Zha, W., Zhu, T., and Zhao, H. (2003) in *Directed Evolution Library Creation*, vol. **231** (eds F. Arnold and G. Georgiou), Humana Press, pp. 91–97.
81. Behrendorff, J.Y.H., Johnston, W., and Gillam, E.J. (2014) in *Directed Evolution Library Creation*, vol. **1179** (eds E.M.J.

- Gillam, J.N. Copp, and D. Ackerley), Springer, New York, pp. 175–187.
82. (a) Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.*, **16**, 258–261; (b) Garcia-Ruiz, E., Mate, D., Ballesteros, A., Martinez, A., and Alcalde, M. (2010) Evolving thermostability in mutant libraries of ligninolytic oxidoreductases expressed in yeast. *Microb. Cell Fact.*, **9** (1), 17.
 83. Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., and Monticello, D.J. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.*, **19** (4), 354–359.
 84. Gibbs, M.D., Nevalainen, K.M.H., and Bergquist, P.L. (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene*, **271**, 13–20.
 85. Bergquist, P.L., Reeves, R.A., and Gibbs, M.D. (2005) Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution. *Biomol. Eng.*, **22**, 63–72.
 86. Ostermeier, M., Shim, J.H., and Benkovic, S.J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.*, **17**, 1205–1209.
 87. Lutz, S., Ostermeier, M., and Benkovic, S.J. (2001) Rapid generation of incremental truncation libraries for protein engineering using α -phosphothioate nucleotides. *Nucleic Acids Res.*, **29**, e16.
 88. Kawarasaki, Y., Griswold, K.E., Stevenson, J.D., Selzer, T., Benkovic, S.J., Iverson, B.L., and Georgiou, G. (2003) Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. *Nucleic Acids Res.*, **31**, e126.
 89. Sieber, V., Martinez, C.A., and Arnold, F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.*, **19**, 456–460.
 90. Higara, K. and Arnold, F.H. (2003) General method for sequence-independent site-directed chimera-genesis. *J. Mol. Biol.*, **330**, 287–296.
 91. Lee, S.H., Ryu, E.J., Kang, M.J., Wang, E.-S., Piao, Z., Choi, Y.J., Jung, K.H., Jeon, J.Y.J., and Shin, Y.C. (2003) A new approach to directed gene evolution by recombined extension on truncated templates (RETT). *J. Mol. Catal. B: Enzym.*, **26**, 119–129.
 92. Ikeuchi, A., Kawarasaki, Y., Shinbata, T., and Yamane, T. (2003) Chimeric gene library construction by a simple and highly versatile method using recombination-dependent exponential amplification. *Biotechnol. Progr.*, **19**, 1460–1467.
 93. O'Maille, P.E., Bakhtina, M., and Tsai, M.D. (2002) Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.*, **321**, 677–691.
 94. Hamamatsu, N., Aita, T., Nomiyama, Y., Uchiyama, H., Nakajima, M., Husimi, Y., and Shibana, Y. (2005) Biased mutation-assembling: an efficient method for rapid directed evolution through simultaneous mutation accumulation. *Protein Eng. Des. Sel.*, **18**, 265–271.
 95. Stutzman-Engwall, K., Conlon, S., Fedechko, R., McArthur, H., Pekrun, K., Chen, Y., Jenne, S., La, C., Trinh, N., Kim, S., Zhang, Y.-X., Fox, R., Gustafsson, C., and Krebber, A. (2005) Semi-synthetic DNA shuffling of *aveC* leads to improved industrial scale production of doramectin by *Streptomyces avermitilis*. *Metab. Eng.*, **7**, 27–37.
 96. (a) Herman, A. and Tawfik, D.S. (2007) Incorporating synthetic oligonucleotides via gene reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.*, **20**, 219–226; (b) Rockah-Shmuel, L., Tawfik, D.S., and Goldsmith, M. (2014) Generating targeted libraries by the combinatorial incorporation of synthetic oligonucleotides during gene shuffling (ISOR). *Methods Mol. Biol.*, **1179**, 129–137.
 97. Coco, W.M., Encell, L.P., Levinson, W.E., Crist, M.J., Loomis, A.K.,

- Licato, L.L., Arensdorf, J.J., Sica, N., Pienkos, P.T., and Monticello, D.J. (2002) Growth factor engineering by degenerate homoduplex gene family recombination. *Nat. Biotechnol.*, **20**, 1246–1250.
98. Ness, J.E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T.V., Govindarajan, S., Mundorff, E.C., and Minshull, J. (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.*, **20**, 1251–1255.
 99. Zha, D., Eipper, A., and Reetz, M.T. (2003) Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. *ChemBioChem*, **4** (1), 34–39.
 100. Acevedo-Rocha, C.G. and Reetz, M.T. (2014) Assembly of designed oligonucleotides: a useful tool in synthetic biology for creating high quality combinatorial DNA libraries. *Methods Mol. Biol.*, **1179**, 189–206.
 101. Chao, R., Yuan, Y., and Zhao, H. (2015) Recent advances in DNA assembly technologies. *FEMS Yeast Res.* doi: 10.1111/1567-1364.12171
 102. Yu, Y. and Lutz, S. (2011) Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.*, **29**, 18–25.
 103. (a) Cunningham, B.A., Hemperly, J.J., Hopp, T.P., and Edelman, G.M. (1979) Favin versus concanavalin A: circularly permuted amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.*, **76** (7), 3218–3222; (b) Lo, W.-C., Lee, C.-C., Lee, C.-Y., and Lyu, P.-C. (2009) Cpdb: a database of circular permutation in proteins. *Nucleic Acids Res.*, **37** (Database issue), D328–D332.
 104. (a) Hennecke, J., Sebbel, P., and Glockshuber, R. (1999) Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J. Mol. Biol.*, **286** (4), 1197–1215; (b) Butler, J.S., Mitrea, D.M., Mitrousis, G., Cingolani, G., and Loh, S.N. (2009) Structural and thermodynamic analysis of a conformationally strained circular permutant of barnase. *Biochemistry*, **48** (15), 3497–3507; (c) Cheltsov, A.V., Barber, M.J., and Ferreira, G.C. (2001) Circular permutation of 5-aminolevulinate synthase: mapping the polypeptide chain to its function. *J. Biol. Chem.*, **276** (22), 19141–19149.
 105. (a) Qian, Z. and Lutz, S. (2005) Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. *J. Am. Chem. Soc.*, **127** (39), 13466–13467; (b) Yu, Y. and Lutz, S. (2010) Improved triglyceride transesterification by circularly permuted *Candida antarctica* lipase B. *Biotechnol. Bioeng.*, **105** (1), 44–50; (c) Reitingner, S., Yu, Y., Wicki, J., Ludwiczek, M., D'Angelo, I., Baturin, S., Okon, M., Strynadka, N.C.J., Lutz, S., Withers, S.G., and McIntosh, L.P. (2010) Circular permutation of *Bacillus circulans* xylanase: a kinetic and structural study. *Biochemistry*, **49** (11), 2464–2474; (d) Qian, Z., Fields, C.J., and Lutz, S. (2007) Investigating the structural and functional consequences of circular permutation on lipase B from *Candida antarctica*. *ChemBioChem*, **8** (16), 1989–1996; (e) Qian, Z., Horton, J.R., Cheng, X., and Lutz, S. (2009) Structural redesign of lipase B from *Candida antarctica* by circular permutation and incremental truncation. *J. Mol. Biol.*, **393** (1), 191–201.
 106. Pierre, B., Shah, V., Xiao, J., and Kim, J.R. (2015) Construction of a random circular permutation library using an engineered transposon. *Anal. Biochem.*, **474**, 16–24.
 107. Mehta, M.M., Liu, S., and Silberg, J.J. (2012) A transposase strategy for creating libraries of circularly permuted proteins. *Nucleic Acids Res.*, **40** (9), e71.
 108. (a) Bennett, M.J., Choe, S., and Eisenberg, D. (1994) Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **91** (8), 3127–3131; (b) Wulff, B.B.H., Thomas, C.M., Smoker, M., Grant, M., and Jones, J.D.G. (2001) Domain swapping and gene shuffling identify sequences required for induction of an Avr-dependent hypersensitive response by the tomato Cf-4 and Cf-9

- proteins. *Plant Cell*, **13** (2), 255–272;
- (c) Park, S.-H., Park, H.-Y., Sohng, J.K., Lee, H.C., Liou, K., Yoon, Y.J., and Kim, B.-G. (2009) Expanding substrate specificity of GT-B fold glycosyl-transferase via domain swapping and high-throughput screening. *Biotechnol. Bioeng.*, **102** (4), 988–994.
109. Ostermeier, M. and Benkovic, S.J. (2000) Evolution of protein function by domain swapping. *Adv. Protein Chem.*, **55**, 29–77.
110. (a) Golczak, M., Sears, A.E., Kiser, P.D., and Palczewski, K. (2015) LRAT-specific domain facilitates vitamin a metabolism by domain swapping in HRASLS3. *Nat. Chem. Biol.*, **11** (1), 26–32; (b) Zhou, X., Wang, H., Zhang, Y., Gao, L., and Feng, Y. (2012) Alteration of substrate specificities of thermophilic α/β hydrolases through domain swapping and domain interface optimization. *Acta Biochim. Biophys. Sin.*, **44** (12), 965–973; (c) Yamanaka, M., Nagao, S., Komori, H., Higuchi, Y., and Hirota, S. (2015) Change in structure and ligand binding properties of hyperstable cytochrome c555 from *Aquifex aeolicus* by domain swapping. *Protein Sci.*, **24** (3), 366–375; (d) Roach, C.R., Hall, D.E., Zerbe, P., and Bohlmann, J. (2014) Plasticity and evolution of (+)-3-carene synthase and (–)-sabinene synthase functions of a sitka spruce monoterpene synthase gene family associated with weevil resistance. *J. Biol. Chem.*, **289** (34), 23859–23869; (e) Pardo, I. and Camarero, S. (2015) Laccase engineering by rational and evolutionary design. *Cell. Mol. Life Sci.*, **72** (5), 897–910; (f) Kang, J.-Y., Ryu, S.H., Park, S.-H., Cha, G.S., Kim, D.-H., Kim, K.-H., Hong, A.W., Ahn, T., Pan, J.-G., Joung, Y.H., Kang, H.-S., and Yun, C.-H. (2014) Chimeric cytochromes p450 engineered by domain swapping and random mutagenesis for producing human metabolites of drugs. *Biotechnol. Bioeng.*, **111** (7), 1313–1322; (g) Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011) Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chem. Biol.*, **18** (5), 601–607; (h) Shingate, P. and Sowdhamini, R. (2012) Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. *PLoS One*, **7** (7), e39305; (i) Chang, C., Huang, R., Yan, Y., Ma, H., Dai, Z., Zhang, B., Deng, Z., Liu, W., and Qu, X. (2015) Uncovering the formation and selection of benzylmalonyl-CoA from the biosynthesis of splenocin and enterocin reveals a versatile way to introduce amino acids into polyketide carbon scaffolds. *J. Am. Chem. Soc.*, **137** (12), 4183–4190; (j) van Beek, H.L., de Gonzalo, G., and Fraaije, M.W. (2012) Blending Baeyer-Villiger monooxygenases: using a robust BVMO as a scaffold for creating chimeric enzymes with novel catalytic properties. *Chem. Commun.*, **48**, 3288–3290.
111. Xiong, A.S., Peng, R.H., Zhuang, J., Liu, J.G., Gao, F., Chen, J.M., Cheng, Z.M., and Yao, Q.H. (2008) Non-polymerase-cycling assembly-based chemical gene synthesis: strategies, methods, and progress. *Biotechnol. Adv.*, **26**, 121–134.
112. (a) Carothers, J.M., Goler, J.A., and Keasling, J.D. (2009) Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.*, **20**, 498–503; (b) Leprince, A., van Passel, M.W., and dos Santos, V.A. (2012) Streamlining genomes: toward the generation of simplified and stabilized microbial systems. *Curr. Opin. Biotechnol.*, **23**, 651–658.
113. Xiong, A.S., Peng, R.H., Zhuang, J., Gao, F., Li, Y., Cheng, Z.M., and Yao, Q.H. (2008) Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS Microbiol. Rev.*, **32**, 522–540.
114. (a) Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G. Jr., Kinney, J.B., Kellis, M., Lander, E.S., and Mikkelsen, T.S. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277; (b) Kwasniewski,

- J.C., Mogno, I., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2012) Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 19498–19503; (c) Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., and Shendure, J. (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.*, **27**, 1173–1175; (d) Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.*, **30**, 521–530.
115. LeProust, E.M., Peck, B.J., Spirin, K., McCuen, H.B., Moore, B., Namsaraev, E., and Caruthers, M.H. (2010) Synthesis of highquality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.*, **38**, 2522–2540.
 116. Mulligan, J.T. and Parker, H.-Y. (2008) Solid phase methods for polynucleotide production. US Patent 7, 482,119.
 117. van den Brulle, J., Fischer, M., Langmann, T., Horn, G., Waldmann, T., Arnold, S., Fuhrmann, M., Schatz, O., O'Connell, T., O'Connell, D., Auckenthaler, A., and Schwer, H. (2008) A novel solid phase technology for high-throughput gene synthesis. *Biotechniques*, **45**, 340–343.
 118. Kuhn, S.M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010) Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting. *J. Mol. Biol.*, **404**, 70–87.
 119. Gebauer, M., Schiefner, A., Matschiner, G., and Skerra, A. (2013) Combinatorial design of an Anticalin directed against the extradomain B for the specific targeting of oncofetal fibronectin. *J. Mol. Biol.*, **425**, 780–802.
 120. Bowers, P.M., Neben, T.Y., Tomlinson, G.L., Dalton, J.L., Altobelli, L., Zhang, X., Macomber, J.L., Wu, B.F., Toobian, R.M., McConnell, A.D., Verdino, P., Chau, B., Horlick, R.A., and King, D.J. (2013) Humanization of antibodies using heavy chain complementarity-determining region 3 grafting coupled with in vitro somatic hypermutation. *J. Biol. Chem.*, **288**, 7688–7696.
 121. Kille, S., Zilly, F.E., Acevedo, J.P., and Reetz, M.T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.*, **3** (9), 738–743.
 122. (a) Sebestova, E., Bendl, J., Brezovsky, J., and Damborsky, J. (2014) Computational tools for designing smart libraries. *Methods Mol. Biol.*, **1179**, 291–314; (b) Zaugg, J., Gumulya, Y., Gillam, E.J., and Bodén, M. (2014) Computational tools for directed evolution: a comparison of prospective and retrospective strategies. *Methods Mol. Biol.*, **1179**, 315–333; (c) Swiderek, K., Tunon, I., Moliner, V., and Bertran, J. (2015) Computational strategies for the design of new enzymatic functions. *Arch. Biochem. Biophys.*, **582**, 68–79; (d) Henrich, S., Salo-Ahen, O.M.H., Huang, B., Rippmann, F.F., Cruciani, G., and Wade, R.C. (2009) Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.*, **23**, 209–219; (e) Lin, Y., Yoo, S., and Sanchez, R. (2012) SiteComp. A server for ligand binding site analysis in protein structures. *Bioinformatics*, **28**, 1172–1173.
 123. Jacobs, T.M., Yumerefendi, H., Kuhlman, B., and Leaver-Fay, A. (2015) SwiftLib: rapid degenerate-codon-library optimization through dynamic programming. *Nucleic Acids Res.*, **43** (5), e34.
 124. Pavelka, A., Chovancova, E., and Damborsky, J. (2009) Hotspot wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.*, **37**, W376–W383.
 125. Dundas, J., Ouyang, Z., and Miteva, M.A. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping

- of functionally annotated residues. *Nucleic Acids Res.*, **34**, W116–W118.
126. Kuipers, R.K., Joosten, H.-J., van Berkel, W.J.H., Leferink, N.G.H., Rooijen, E., Ittmann, E., van Zimmeren, F., Jochens, H., Bornscheuer, U., Vriend, G., Martins dos Santos, V.A.P., and Schaap, P.J. (2010) 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins*, **78** (9), 2101–2113.
 127. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38** (Web Server issue), W529–W533.
 128. Ma, B.-G. and Berezovsky, I.N. (2010) The mblosum: a server for deriving mutation targets and position-specific substitution rates. *J. Biomol. Struct. Dyn.*, **28** (3), 415–419.
 129. Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J., and Rousseau, F. (2011) A graphical interface for the FoldX forcefield. *Bioinformatics*, **27** (12), 1711–1712.
 130. Saraf, M.C., Horswill, A.R., Benkovic, S.J., and Maranas, C.D. (2004) FamClash: a method for ranking the activity of engineered enzymes. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4142–4147.
 131. Moore, G.L. and Maranas, C.D. (2003) Identifying residue–residue clashes in protein hybrids by using a second-order mean-field approach. *Proc. Natl. Acad. Sci. U.S.A.*, **100** (9), 5091–5096.
 132. Saraf, M.C., Moore, G.L., Goodey, N.M., Cao, V.Y., and Benkovic, S.J. (2006) IPRO: an iterative computational protein library redesign and optimization procedure. *Biophys. J.*, **90**, 4167–4180.
 133. (a) Kuipers, R.K.P., Joosten, H.-J., Verwiel, E., Paans, S., Akerboom, J., van der Oost, J., Leferink, N.G.H., van Berkel, W.J.H., Vriend, G., and Schaap, P.J. (2009) Correlated mutation analyses on super-family alignments reveal functionally important residues. *Proteins Struct. Funct. Bioinf.*, **76** (3), 608–616; (b) Nobili, A., Tao, Y., Pavlidis, I.V., van den Bergh, T., Joosten, H.-J., Tan, T., and Bornscheuer, U.T. (2015) Simultaneous use of in silico design and a correlated mutation network as a tool to efficiently guide enzyme engineering. *ChemBioChem*, **16** (5), 805–810; (c) Engqvist, M.K.M. and Nielsen, J. (2015) ANT: software for generating and evaluating degenerate codons for natural and expanded genetic codes. *ACS Synth. Biol.*, **4**, 935–938.
 134. Parthiban, V., Gromiha, M.M., and Schomburg, D. (2006) Cupsat: prediction of protein stability upon point mutations. *Nucleic Acids Res.*, **34** (Web Server issue), W239–W242.
 135. Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., and Damborsky, J. (2012) Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.*, **8** (10), e1002708.
 136. Guerois, R., Nielsen, J.E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320** (2), 369–387.
 137. Dehouck, Y., Kwasigroch, J., Gilis, D., and Rooman, M. (2011) Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinf.*, **12** (1), 151.
 138. Wijma, H.J., Floor, R.J., Jekel, P.A., Baker, D., Marrink, S.J., and Janssen, D.B. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.*, **27** (2), 49–58.
 139. Bednar, D., Beerens, K., Sebestova, E., Bendl, J., Khare, S., Chaloupkova, R., Prokop, Z., Brezovsky, J., Baker, D., and Damborsky, J. (2015) Fireprot: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.*, **11** (11), e1004556.
 140. (a) Saab-Rincon, G., Li, Y., Meyer, M., Carbone, M., Landwehr, M., and Arnold, F.H. (2009) in *Protein Engineering Handbook* (eds S. Lutz and U.T. Bornscheuer), Wiley-VCH Verlag GmbH, Weinheim, pp. 481–492;

- (b) Heinzelman, P., Snow, C.D., Smith, M.A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F.H. (2009) SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J. Biol. Chem.*, **284**, 26229–26233;
- (c) Heinzelman, P., Romero, P.A., and Arnold, F.H. (2013) Efficient sampling of SCHEMA chimera families to identify useful elements. *Methods Enzymol.*, **523**, 351–368.
141. Pantazes, R.J., Saraf, M.C., and Maranas, C.D. (2007) Optimal protein library design using recombination or point mutations based on sequence based scoring functions. *Protein Eng. Des. Sel.*, **20**, 361–373.
 142. Fox, R.J., Davis, S.C., Mundorff, E.C., Newman, L.M., Gavrilovic, V., Ma, S.K., Chung, L.M., Ching, C., Tam, S., Muley, S., Grate, J., Gruber, J., Whitman, J.C., Sheldon, R.A., and Huisman, G.W. (2007) Improving catalytic function by ProSAR-driven enzyme evolution. *Nat. Biotechnol.*, **25**, 338–344.
 143. Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., and Hughes, G.J. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to Sitagliptin manufacture. *Science*, **329** (5989), 305–309.
 144. (a) Liang, F., Feng, X.-j., Lowry, M., and Rabitz, H. (2005) Maximal use of minimal libraries through the adaptive substituent reordering algorithm. *J. Phys. Chem. B*, **109** (12), 5842–5854; (b) Carvalho, S.A., da Silva, E.F., Santa-Rita, R.M., de Castro, S.L., and Fraga, C.A.M. (2004) Synthesis and antitrypanosomal profile of new functionalized 1,3,4-thiadiazole-2-arylhydrazone derivatives, designed as non-mutagenic megazol analogues. *Bioorg. Med. Chem. Lett.*, **14** (24), 5967–5970.
 145. Feng, X., Sanchis, J., Reetz, M.T., and Rabitz, H. (2012) Enhancing the efficiency of directed evolution in focused enzyme libraries by the adaptive substituent reordering algorithm. *Chem. Eur. J.*, **18** (18), 5646–5654.
 146. (a) Huang, P.-S., Ban, Y.-E.A., Richter, F., Andre, I., Vernon, R., Schief, W.R., and Baker, D. (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One*, **6**, e24109; (b) Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., and Houk, K.N. (2013) Computational enzyme design. *Angew. Chem. Int. Ed.*, **52** (22), 5700–5725.

4

Strategies for Applying Gene Mutagenesis Methods

4.1

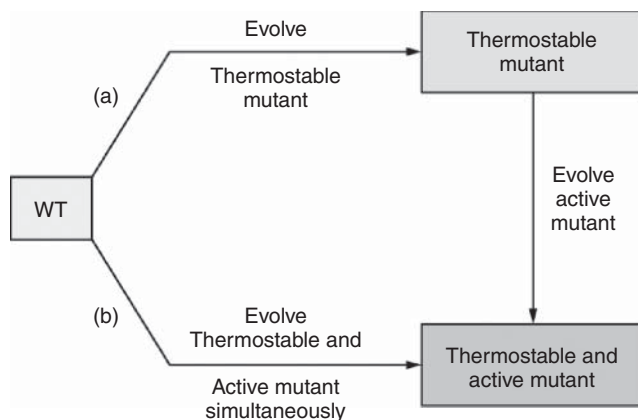
General Guidelines

As already pointed out in Chapter 3, application of any one of the gene mutagenesis techniques is likely to provide improved enzyme variants, but choosing the best method for a given endeavor coupled with the optimal strategy on how to apply it ensures efficiency needed in practical (industrial) applications [1]. The optimal choice(s) will depend on the catalytic parameter(s) to be engineered. Thermostability and robustness in hostile organic solvents are treated in Chapter 6. In the present chapter, the primary focus is on enzymes as catalysts in organic chemistry and biotechnology in which the following parameters are manipulated:

- stereo- and regioselectivity
- activity
- substrate scope.

Whenever more than one enzyme parameter needs to be improved, two strategic options are possible [1]: (i) Engineering the parameters separately in two sequential projects or (ii) optimizing both parameters simultaneously in a single project. It has been recommended that for practical reasons it is useful to start directed evolution from a robust enzyme, since protein stability promotes evolvability [2]. It can, therefore, be concluded that enhancing thermostability should precede the engineering of other parameters such as activity (Scheme 4.1a). However, it should also be possible to improve both parameters simultaneously by designing appropriate experimental platforms in which the microtiter plates are heated at a given temperature for a defined length of time prior to high-throughput screening for activity. All variants of inferior thermostability would denature by the heat treatment, and only those in which thermostabilizing mutations have been introduced will “survive” (Scheme 4.1b). Although the simultaneous optimization of two parameters seems particularly appealing, not enough comparative studies have been performed thus far for a final assessment (see also Chapter 6).

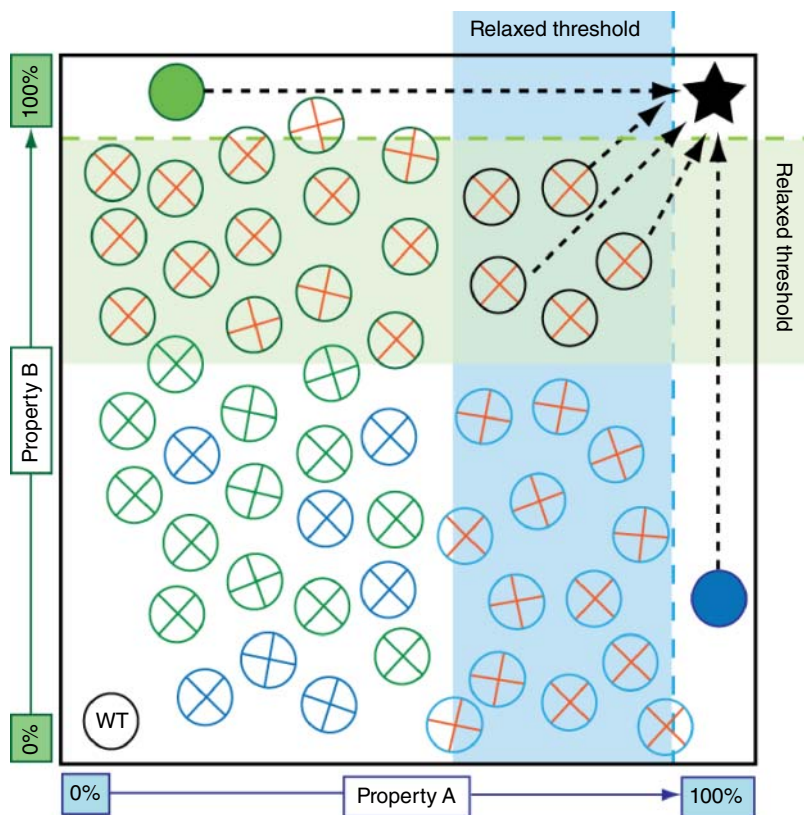
In other directed evolution endeavors, activity and/or enantioselectivity constitute the catalytic parameters of interest. In such cases a given library may contain variants with highest enantioselectivity (property A), but low activity (property B),



Scheme 4.1 Two choices when attempting to optimize thermostability and activity of an enzyme. (a) Engineer thermostability and then activity. (b) Engineer both thermostability and activity simultaneously.

or vice versa, variants with highest activity but low enantioselectivity. It has been demonstrated that the very best hits with respect to one parameter (enantioselectivity) should not be used as templates in the subsequent cycle of mutagenesis/expression/screening when seeking higher activity as well as stereoselectivity, and vice versa [3]. Rather, a compromise is recommended in that several medium- or higher-quality hits, but not the very best ones in terms of stereoselectivity, are not discarded, but used as templates in the next directed evolution round focused on the other catalytic parameter (property B, e.g., activity). This approach is illustrated in Scheme 4.2 [3]. These kinds of non-discarded variants are reminiscent of neutral drift mutants [4], but the term “*lateral hits*” may be more appropriate [3]. A remote relationship to the Eigen/Schuster notion of quasi-species [5] may also be noted, as invoked in other directed evolution studies [6]. An opposing view is to evolve maximally high activity first, and then to tune stereoselectivity. Comparative studies are needed in order to make final conclusions.

Yet another and likewise crucial question concerns the optimal number of point mutations that should be introduced in each mutagenesis/expression/screening cycle, irrespective of the type of genetic method used in a directed evolution project. Different opinions have been voiced on this important issue, and, certainly, successful examples of single point mutations and also of multiple point mutations, have been reported numerous times [1]. In 2009, it was strongly recommended that climbing the hill in a fitness landscape is best achieved when only single mutations are allowed to accumulate one by one in evolutionary cycles [7]. However, this conclusion did not consider previous studies in which the simultaneous introduction of more than one point mutation was shown to be highly efficient [8–10], success being due to cooperative effects (more than traditional additivity) within a set of point mutations and between sets of



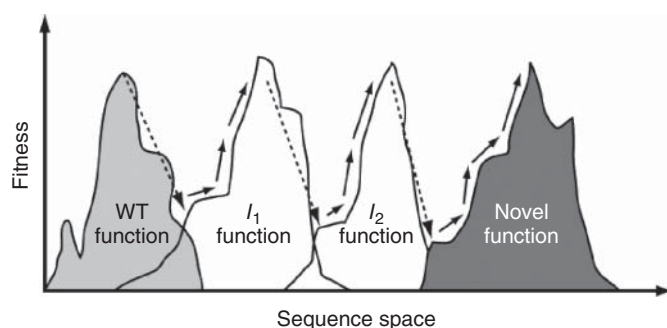
Scheme 4.2 Preferred approach for the simultaneous optimization of two catalyst properties A and B [3]. Black star indicates the desired variant; blue and green dashed lines, stringent thresholds; blue and green rectangles, relaxed thresholds; blue and green filled circles, best mutant for property A and B, respectively, which are not used

in further mutagenesis; red-crossed blue and green circles, variants with improved property A or B; red-crossed black circles, mutants with improved A and B property. Black dashed arrows, second round of mutagenesis. (Bougioukou *et al.* [3]. Reproduced with permission of John Wiley & Sons.)

mutations [11–14]. The occurrence of pronounced mutational cooperativity in directed evolution has been interpreted as a sign of efficacy [14].

Whenever an enzyme's substrate scope needs to be shifted or enlarged because a given compound of interest is not accepted (lack of notable activity), two different directed evolution strategies can be applied: (i) proceed conventionally by using the substrate of interest directly in mutagenesis/screening rounds, possibly aided by structural, consensus sequence data, and/or computational support or (ii) apply the *in vitro* coevolutionary strategy initially developed for manipulating binding affinity [15] and later applied in enzyme catalysis [16]. It is a type of “substrate walking”, in which a compound structurally more closely related to the natural substrate, but reacting slower, is first subjected to directed evolution for enhanced

activity. This is followed by several similar iterative steps using structurally altered compounds until the final substrate of actual interest is targeted (Scheme 4.3). The latter option, although conceptionally intriguing, requires several steps and may not be necessary due to recent developments of dramatically improved directed evolution methods and strategies. However, should such advanced techniques fail, for whatever reason, the approach based on *in vitro* coevolution may be the best way to solve the problem.



Scheme 4.3 The strategy of *in vitro* coevolution (substrate walking) for engineering novel protein functions [15, 16]. The wild-type (WT) protein function and the novel protein function are separated by an inactive region of sequence space, which may

be filled by two intermediate functions (I_1 and I_2) that are amenable to conventional directed evolution. The arrows illustrate a potential evolutionary path leading to the novel protein function. (Chen and Zhao [15]. Reproduced with permission of Elsevier.)

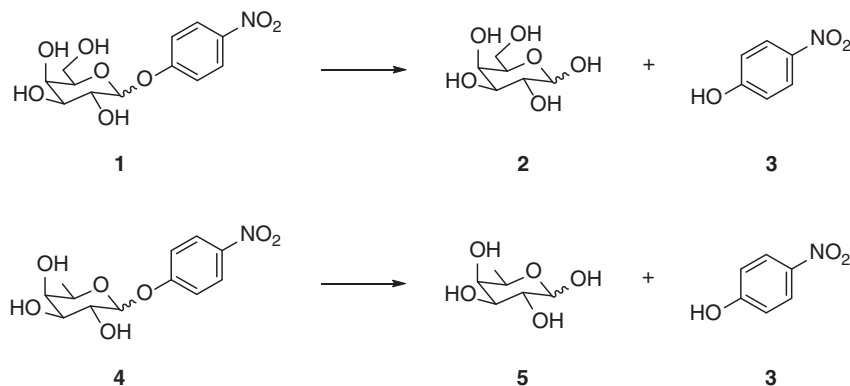
A final and particularly pressing issue concerns the choice of the mutagenesis method. In successful attempts to manipulate stereoselectivity, activity, and/or substrate scope, a variety of different molecular biological techniques such as error-prone polymerase chain reaction (epPCR), saturation mutagenesis, and DNA shuffling have been invoked (Chapter 3), but it took a number of years of research before the optimal strategy emerged [1]. In the following section, rare cases of comparative studies are highlighted and analyzed. These contributions have led to the conclusion that saturation mutagenesis at sites lining the binding pocket is the method of choice, provided structural information serving as a guide in designing mutant libraries is available. Section 4.3 that follows thereafter, contains tips on how to apply saturation mutagenesis and iterative saturation mutagenesis (ISM).

4.2

Rare Cases of Comparative Studies

An early example of comparing two gene mutagenesis methods concerns DNA shuffling [17] versus saturation mutagenesis [18] in the directed evolution of

substrate acceptance, specifically evolving a fucosidase starting from a galactosidase. The model reactions involved the hydrolysis of the “natural” synthetic substrate $1 \rightarrow 2 + 3$ catalyzed by *Escherichia coli lacZ* β -galactosidase (BGAL) and the respective reaction of the “non-natural” substrate $4 \rightarrow 5 + 3$ to be catalyzed by an evolved mutant (Scheme 4.4). The two substrates differ by a single hydroxyl group at position C6 ($-\text{CH}_2\text{OH}$ versus $-\text{CH}_3$), but otherwise the stereochemical features of the compounds are identical. In the reaction of substrate **4**, wild-type β -galactosidase (WTBGAL) is a poor catalyst in a sluggish reaction.

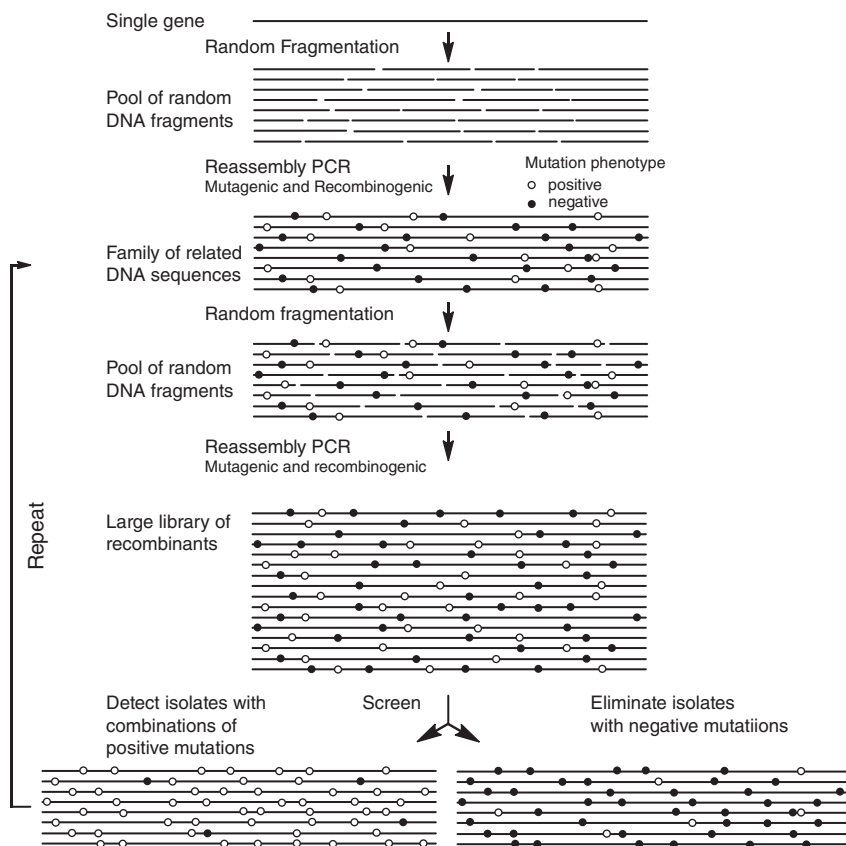


Scheme 4.4 Model reactions in the directed evolution of a fucosidase from a galactosidase [17, 18]. (Zhang *et al.* [17]. Reproduced with permission of National Academy of Sciences.)

The steps involved in the DNA shuffling study are summarized in Scheme 4.5 [17]. A total of seven DNA shuffling cycles were performed, each time about 10 000 colonies being screened using a crude but convenient on-plate color test for activity. About 2–5% of the transformants displayed enhanced activity as judged by the pre-test, and typically 20–40 variants were then isolated and studied separately by kinetics.

The best mutant from the seventh and final round of shuffling showed a 66-fold increase in fucosidase activity and a reduction in galactosidase activity, but it still displayed a preference for the “natural” substrate **1** by a factor of 2.7 [17]. The desired reversal of substrate preference was not achieved. The best variant is characterized by eight point mutations, only two being near the active site. Apparently, subtle effects are involved, which were not unambiguously uncovered. Moreover, since deconvolution studies were not performed, it is unclear whether all eight point mutations are really necessary for the change in substrate acceptance.

The purpose of the second study, this time utilizing saturation mutagenesis, was to compare the two gene mutagenesis methods by employing the identical substrates, the same enzyme BGAL and the previous on-plate screening assay [18]. Based on X-ray data of BGAL [19], saturation mutagenesis was applied at a site comprising three residues 201, 540, and 604 which bind sodium ions (Figure 4.1). It was speculated that simultaneous randomization at these positions as part of the



Scheme 4.5 DNA shuffling process used in the directed evolution of a fucosidase starting from a galactosidase [17]. Zhang *et al.* [17]. Reproduced with permission of National Academy of Sciences.

binding site would provide variants that no longer coordinate Na^+ , nor interact with the hydroxyl group at C6. Rather, interaction directly with the methyl group of substrate **4** was expected, which should shift substrate selectivity in favor of the “fuco-substrate” **4** and less so for the “galacto-substrate” **1**.

Saturation mutagenesis was performed using NNK codon degeneracy encoding all 20 canonical amino acids, followed by screening with the on-plate pre-test. In this way, about 10 000 transformants were assayed, leading to the discovery of a handful of active variants showing pronounced degrees of preference for substrate **4** (Figure 4.2) [18].

The best variant, H540V/N604T, proved to be considerably more active (180-fold increase in $k_{\text{cat}}/K_{\text{M}}$) in the reaction of **4**, while favoring its transformation relative to that of the “natural” synthetic substrate **1**. The observed 7000-fold switch is dramatic. Although the interpretation of the result on a molecular level was difficult because Asp201 was retained, the overall conclusion regarding the pros

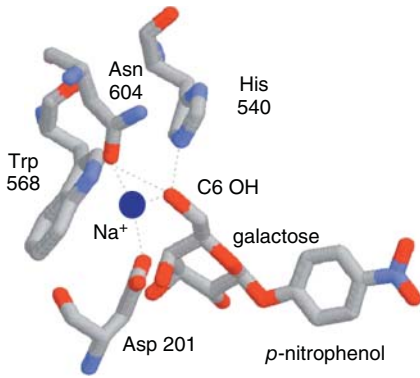


Figure 4.1 Structure of BGAL active site [19] used as a guide in designing saturation mutagenesis at amino acid positions 201, 540, and 604 [18]. (Parikh and Matsumura [18]. Reproduced with permission of Elsevier.)

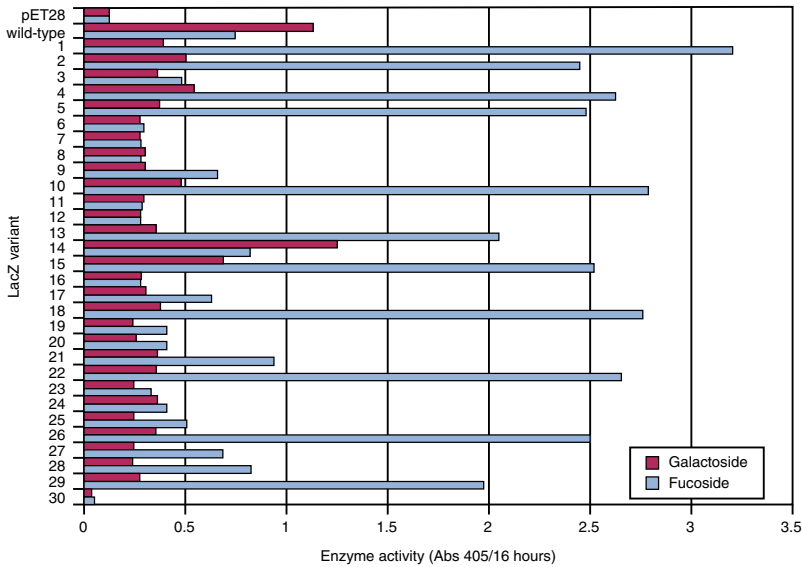


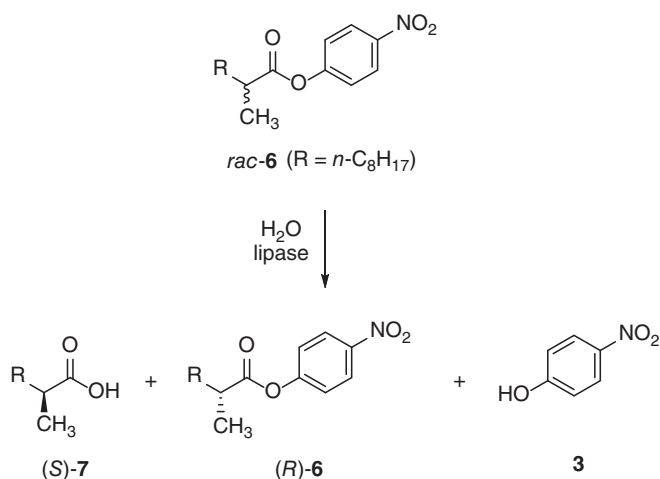
Figure 4.2 Selected BGAL variants resulting from saturation mutagenesis at a site composed of amino acid positions 201, 540, and 604 [18]. (Parikh and Matsumura [18]. Reproduced with permission of Elsevier.)

and cons of DNA shuffling versus saturation mutagenesis was convincing: focused library generation by structure-guided saturation mutagenesis is superior to DNA shuffling. The generation and screening of a single saturation mutagenesis library (10 000 transformants) leading to these excellent results contrasts with the formation and screening of seven DNA shuffling libraries (total 70 000 transformants) with evolution of notably inferior variants [18]. Nevertheless, the authors were careful not to generalize their conclusion [18].

Another aspect of this important study deserves mention. As delineated in Chapter 3, randomization at a three-residue site would require for 95% library

coverage the screening of about 98 000 transformants (Table 3.3). Since only 10 000 were actually assayed, it is clear that essentially full library coverage is not at all necessary for obtaining acceptable results [18]. Indeed, as demonstrated in an earlier study, even lower library coverage in saturation mutagenesis for enhanced stereoselectivity and activity may well suffice [20]. Obviously, the best variants will be missed by such a procedure, but the *n*th best hit may be acceptable for practical applications [21].

The most systematically performed comparison of different mutagenesis methods and strategies in directed evolution is based on a series of studies over a period of more than a decade. The model reaction involves the hydrolytic kinetic resolution of *rac*-6 with preferential formation of (*S*)-7 catalyzed by mutants of the lipase from *Pseudomonas aeruginosa* (PAL) (Scheme 4.6) [13, 20, 22–25]. As shown in the introductory Chapter 1, this system was used in the original proof-of-principle study showing for the first time that directed evolution can be used to control the enantioselectivity of enzymes [22]. Wild-type lipase from *Pseudomonas aeruginosa* (WT PAL) shows a slight preference for (*S*)-7 (*E* = 1.1).



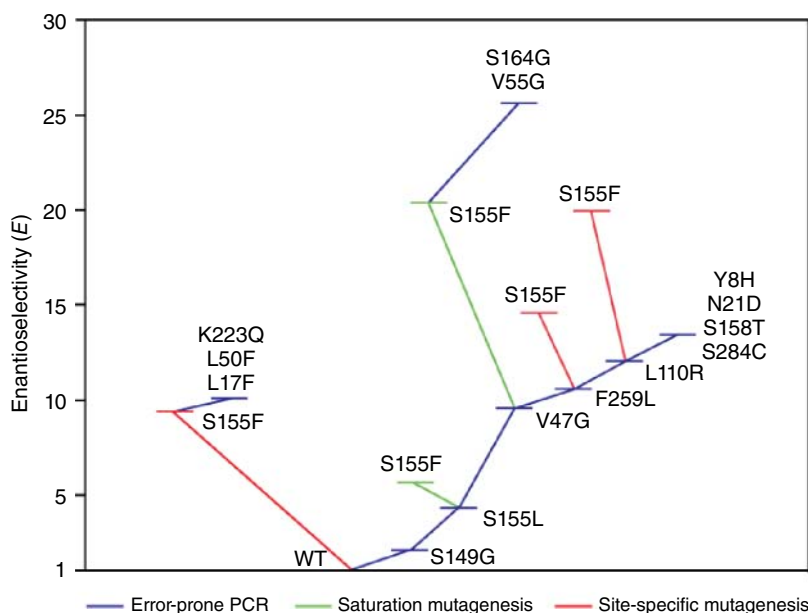
Scheme 4.6 Model reaction used in the directed evolution of PAL [13, 20, 22–25]. (Reetz *et al.* [13]. Reproduced with permission of American Chemical Society.)

Initially, four cycles of epPCR at low mutation averaging one amino acid exchange per enzyme were applied. One point mutation accumulated in each round leading to the (*S*)-selective variant S149G/S155L/V47G/F259L characterized by a selectivity factor of *E* = 11.3 (see Scheme 1.10) [22]. Unfortunately, two further epPCR rounds failed to provide notably improved variants, *E*-values of only 13–15 being achieved [26].

This was a clear indication that iterative epPCR is not well suited for enhancing stereoselectivity, certainly not in this particular enzyme system. Moreover, most

of the hot spots appeared to be remote from the active site as judged by a homology model (the crystal structure of PAL was not published until later). Therefore, methodology development was initiated by exploring other strategies. Assuming that the four mutations S149G/S155L/V47G/F259L occur at hot spots, but that the newly introduced amino acids may not be optimal, saturation mutagenesis was applied at these positions [23].

In doing so, wild-type (WT) or first-, second-, third-, or fourth-generation mutants can be used as templates. In some cases this provided improvements to variants displaying higher enantioselectivity, but in others no improvements were detected. For example, using the third generation mutant S149G/S155L/V47G as the template, saturation mutagenesis using NNK codon degeneracy encoding all 20 canonical amino acids was performed at position 155. Variant S149G/S155F/V47G was identified in which leucine was exchanged for phenylalanine, leading to a notable increase in enantioselectivity ($E \approx 20$). Randomization at this position using other templates likewise led to the same exchange S155F. However, focusing on other positions failed to provide better variants. Therefore, alternating saturation mutagenesis at different positions and epPCR were tested, the best variant V47G/V55G/S149G/S155F/S164G displaying a selectivity factor of $E = 26$ (Scheme 4.7) [23]. This can be viewed as a forerunner of ISM [8] (for a detailed discussion of ISM, see Section 3.3).



Scheme 4.7 Alternating saturation mutagenesis at different positions with epPCR in the quest to enhance the enantioselectivity of PAL in the model reaction $\text{rac-6} \rightarrow (\text{S})\text{-7} + \text{3}$ [23]. (Liebeton *et al.* [23]. Reproduced with permission of Elsevier.)

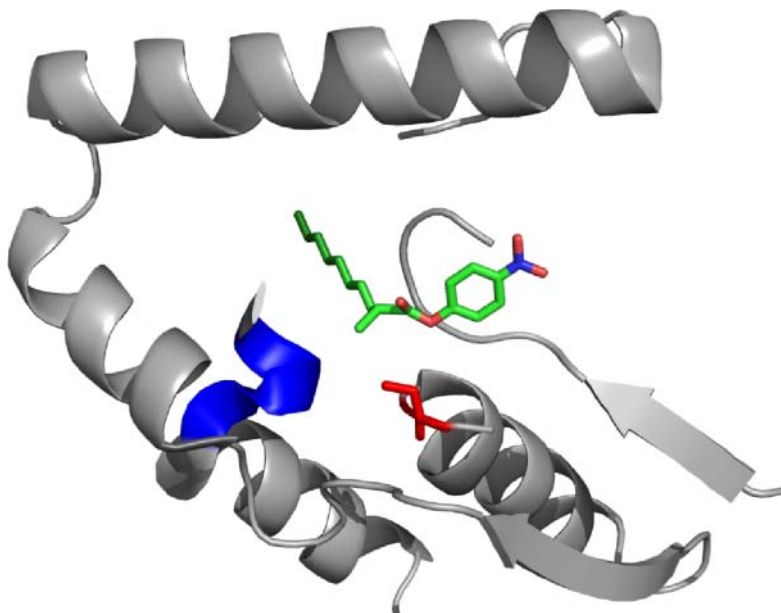


Figure 4.3 Binding pocket of PAL [27] for the acid part of *rac*-6 (green) showing the geometric position of amino acids 160–163 (blue), which were randomized simultaneously by saturation mutagenesis to enhance enantioselectivity [20]. Ser82 (red), as part

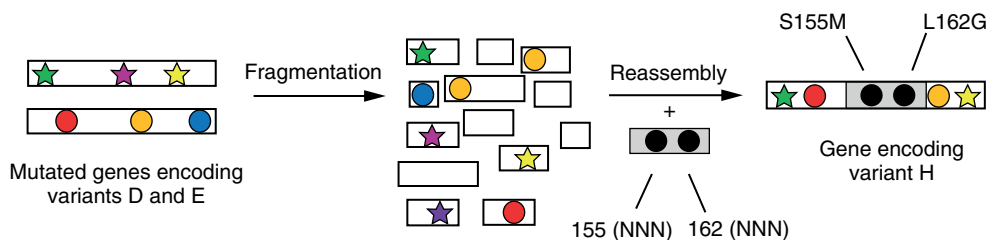
of the catalytic triad Asp/His/Ser, attacks the carbonyl function nucleophilically with rate- and stereoselectivity-determining formation of a short-lived oxyanion. (Reetz *et al.* [20]. Reproduced with permission of John Wiley & Sons.)

While these strategies ensured partial success, the observed enantioselectivities were not high enough for practical applications. Therefore, other approaches were tested, the first involving saturation mutagenesis (NNK codon degeneracy) at a site comprising four amino acid positions 160–163 at the binding pocket of PAL [20] (Figure 4.3). This experiment was guided by the PAL crystal structure [27], and constitutes the first example of focused saturation mutagenesis at a site lining the binding pocket of an enzyme with the aim of enhancing stereoselectivity. It was essentially an attempt to reshape the binding pocket of an enzyme, keeping Emil Fischer's lock-and-key hypothesis in mind. After screening only 5000 transformants, the quadruple mutant G160A/S161D/L162G/N163F showing a selectivity factor of $E = 30$ was identified [20]. At the time, oversampling statistics were not considered, but subsequently it became clear that for 95% library coverage about 3×10^6 transformants should have been screened, far beyond practical possibilities. The result was later interpreted as an indication that aiming for full library coverage in saturation mutagenesis is not necessary, as also demonstrated in the GBAL-study (see above) [18].

In a second saturation mutagenesis experiment in the same study [20], a two-residue site Ser155/Leu162 was randomized, again using NNK codon degeneracy, which provided two variants (labeled VIII and IX) of similar enantioselectivity,

$E = 34$ and $E = 30$, respectively, in favor of (S)-7. Unfortunately, these positive results did not lead to the realization that saturation mutagenesis at sites lining the binding pocket is superior to epPCR and DNA shuffling. This conclusion was finally made several years later with the systematization of saturation mutagenesis at sites lining the binding pocket in the form of the combinatorial active-site saturation test (CAST) [25] and the emergence of ISM [8] (see below).

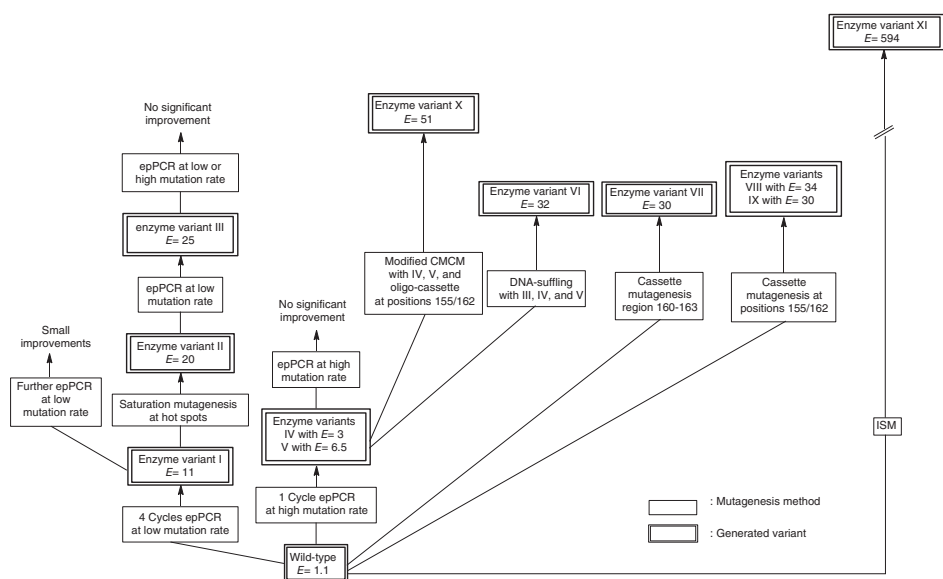
In the PAL study describing saturation mutagenesis at the two- and four-residue sites, a second strategy was tested in which the genes of two previous mutants (labeled IV and V) were subjected to DNA shuffling [20]. In order to increase diversity in the DNA shuffling experiment, a modified form called combinatorial multiple-cassette mutagenesis (CMCM) [28] was applied (for a detailed description, see Section 3.3). Accordingly, a mutagenic oligocassette was included in the shuffling, which allows parallel randomization at positions 155 and 162 (Scheme 4.8) [20]. This procedure provided an enzyme (PAL variant X) characterized by six point mutations D20N/S53P/S155M/L162G/180I/T234S and highest enantioselectivity ($E = 51$) in the model reaction observed up to that point (Scheme 4.9) [20]. Interestingly, only L162G was found to be next to the binding pocket, the other five point mutations being more or less remote.



Scheme 4.8 Extended CMCM in the evolution of an (S)-selective variant X in the hydrolytic kinetic resolution of *rac*-6 [20]. Green star, position 20; purple star, position 161; yellow star, position 234; red circle,

position 53; orange circle, position 180; and blue circle, position 272. (Reetz *et al.* [20]. Reproduced with permission of John Wiley & Sons.)

On the basis of these exploratory experiments, it was concluded that the combination of epPCR and DNA shuffling is the optimal strategy for evolving stereoselective enzymes. Indeed, other groups joined efforts in generalizing directed evolution of stereoselectivity using other types of enzymes along similar lines as summarized in an early review [24]. However, a subsequent mechanistic and quantum mechanics/molecular mechanics (QM/MM) study suggested that only two of the six point mutations in the best variant X are necessary for high enantioselectivity. Indeed, the predicted double mutant S53P/L162G was generated and shown to have a selectivity factor of $E = 63$ in the model reaction, demonstrating that four of the six point mutations are superfluous and actually reduce stereoselectivity [29]. This was a clear signal that the strategy that includes epPCR and DNA shuffling is successful, but not efficient. A total of 50 000 transformants were screened in all



Scheme 4.9 Summary of all comparative studies of PAL as a catalyst in the hydrolytic kinetic resolution of *rac*-6, including the result of the final study based on ISM (far right) [13]. (Reetz *et al.* [13]. Reproduced with permission of American Chemical Society.)

of these experiments as summarized in Scheme 4.9. Similar screening efforts had to be invested in order to invert stereoselectivity favoring (*R*)-7 [30].

Following two advances in developing more efficient saturation mutagenesis, namely the structure-guided use of reduced amino acid alphabets [10] and ISM [8, 9] (see Chapter 3), the PAL model system was re-visited in order to ensure a more rigorous comparison of mutagenesis methods and strategies [13]. The same model reaction was employed (Scheme 4.6), this time being guided by the crystal structure of PAL [27] with ISM being tested as the genetic strategy [13].

Knowing that lipases have “two” binding pockets, one harboring the acid and the other the alcohol part of the ester, a decision was made on where to focus saturation mutagenesis on appropriate CAST sites. Since the chiral center in substrate *rac*-6 is in the acid part of the molecule with catalytically active S82 being between the two segments, six residues were judged as being important for mutagenesis, namely M16, L17, L159, L162, L231, and V232, all surrounding the acid part of the binding pocket (Figure 4.4) [13]. The distance values of the α -C-atom of all these six residues to the α -C-atom of the acid moiety (C4) in *rac*-6 were estimated to be M16 (5.3 Å), L17 (8.4 Å), L159 (8.5 Å), L162 (6.5 Å), L231 (7.0 Å), and V232 (7.3 Å).

With this decision in hand, a choice had to be made on how to group the chosen six CAST residues into multi-residue randomization sites, if at all, and to decide whether to use a reduced amino acid alphabet so that screening is reduced (see Chapters 2 and 3). These strategic questions are addressed in a general way in Section 4.3. In the PAL study, three 2-residue randomization sites were defined, A (M16/L17), B (L159/L162), and C (L231/V232) (Figure 4.4). Each library would require the screening of about 3000 transformants for 95% coverage, which adds up to 9000. This is much more than the screening effort when generating and screening six separate single-residue libraries, but this conclusion reverses in the subsequent iterative steps if the complete ISM scheme is fully investigated (Table 4.1) [13]. Moreover, randomization at sites comprising two or three residues has an additional feature because it can lead to double or triple mutants, respectively, in which the respective point mutations interact with one another cooperatively (more than additively) [14].

If NDT codon degeneracy encoding 12 amino acids (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, and Gly) were to be chosen as the reduced alphabet, then only 430 transformants would have to be screened (total of ≈ 1300 for all three libraries). In fact, libraries A, B, and C were produced using NNK codon degeneracy, and in each case ≈ 3000 transformants were screened. In libraries A and C, slightly improved variants were discovered, but the best hit originated in library B: variant 1 F8 characterized by a single point mutation L162N with $E = 8$ in favor of (*S*)-7. At this point, a decision had to be made on how to proceed in the upward climb. A complete 3-site ISM scheme involves 6 evolutionary pathways and a total of 15 saturation mutagenesis libraries (Scheme 3.9b). Experience has shown that it is not necessary to explore all upward pathways in an ISM study [31].

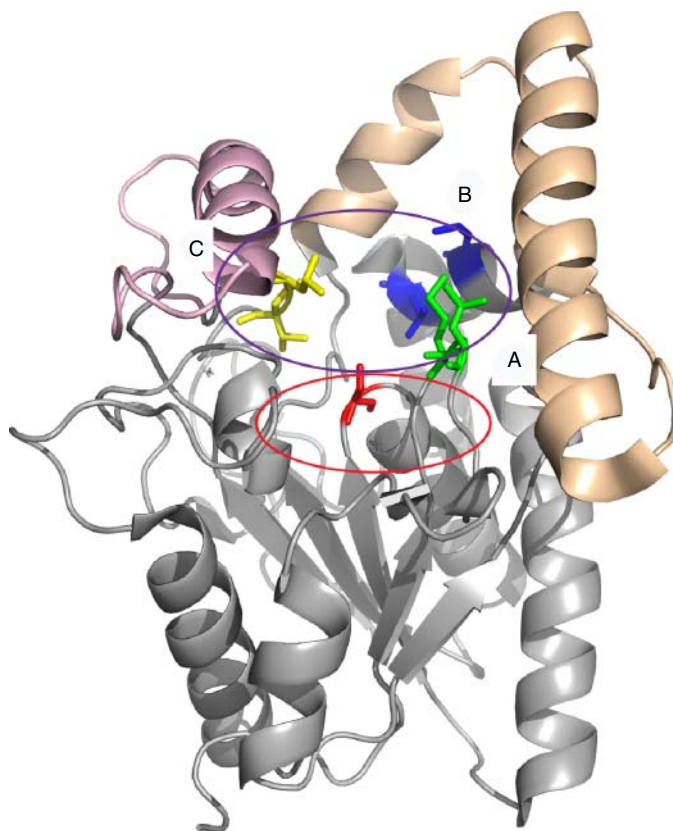


Figure 4.4 Schematic representation of amino acid residues considered for saturation mutagenesis [13], based on the X-ray structure of WT-PAL [27]: sites A (Met16/Leu17, green), B (Leu159/Leu162, blue), and C (Leu231/Val232, yellow) around the active site Ser82 (stick representation in red) in the acid-binding pocket (purple circle). The red

circle marks the alcohol-binding pocket, in the case at hand harboring the *p*-nitrophenyl moiety of *rac*-6. At the top of picture, helix and loop in wheat (right, Asp113-Leu156) and light pink (left, Pro203-Asn228) represent lid 1 and lid 2, respectively. (Reetz *et al.* [13]. Reproduced with permission of American Chemical Society.)

In the present case, mutant 1F8 was used as a template for saturation mutagenesis at sites A and C employing DNT codon degeneracy encoding 11 amino acids (Ala, Asn, Asp, Cys, Gly, Ile, Phe, Ser, Thr, Tyr, and Val). The reason for choosing DNT rather than NDT for second generation randomization at sites A and C is simple: DNT does not encode leucine, and both sites already harbor leucine, thereby reducing the amount of template mutant (“parent,” in this case 1F8) appearing in the library. This increases the library quality. Upon screening the two second generation libraries, a highly active and stereoselective variant was identified in which two new point mutations accumulated, M16A/L17F, leading to the final triple mutant M16A/L17F/L162N (enzyme variant XI) characterized by

Table 4.1 Statistical consequences as a function of grouping single CAST residues into randomization sites [13].

Type of grouping	No. of pathways in complete ISM scheme	No. of libraries in complete ISM scheme	Total number of transformants screened in complete ISM scheme for 95% coverage	
			Using NNK codon	Using NDT or DNT codon
Six single-residue sites	720	1956	183 864 (94 per library)	66 504 (34 per library)
Three double-residue sites	6	15	45 990 (3 066 per library)	6 450 (430 per library)
Two triple-residue sites	2	4	392 652 (98 163 per library)	20 700 (5 175 per library)
One six-residue site	1	1	3.21×10^9	8.95×10^6

Source: Reetz *et al.* [13]. Reproduced with permission of American Chemical Society.

a selectivity factor of $E = 594$ (Scheme 4.9) [13]. Due to the dramatically improved catalytic profile, further ISM at site C was not necessary. The triple mutant is also a good catalyst for the hydrolytic kinetic resolution of other chiral esters. The study includes kinetics, molecular dynamics (MD) computations and deconvolution experiments which uncover unusually large cooperative effects existing between the second mutational set (M16A/L17F) and the initial point mutation (L162N) (see Chapter 8).

In this study, less than 10 000 transformants were screened using a simple on-plate pre-test for activity followed by automated GC analysis of the active hits. This is much less than the invested effort in the older approach based on epPCR and DNA shuffling (50 000 transformants). The final mutant in the ISM approach is dramatically more stereoselective ($E = 594$ vs $E = 51$). At the time of the ISM study [13], the full significance of using reduced amino acid alphabets in the quest to reduce the screening effort was not as clear as it currently is. Today, a researcher would choose for the initial libraries not NNK codon degeneracy, but NDT or DNT codon degeneracy, which encode a balanced sets of polar/non-polar, charged/non-charged, and hydrophobic/hydrophilic amino acids. In such a scenario, less than 1900 transformants would have to be screened for optimizing PAL. As delineated in Section 4.3, even smaller amino acid alphabets should be considered, provided structural data is available when designing the mutant libraries. All of the 12-year efforts, which reflect the different mutagenesis strategies in optimizing PAL are summarized in Scheme 4.9.

As a general conclusion of these extensive comparative studies, epPCR and DNA shuffling can certainly be successful when attempting to manipulate stereoselectivity, rate, and substrate scope of enzymes. However, the above analyses as well as other comparative studies, although not as comprehensive [1, 18], indicate the superiority of saturation mutagenesis [32]. These are the most important catalytic parameters when applying enzymes as catalysts in organic chemistry and/or biotechnology [32, 33]. Nevertheless, researchers opting for this form of protein engineering in future studies should focus on using the optimal strategies when applying saturation mutagenesis, because this saves resources and speeds up the whole evolutionary process (Section 4.3). Finally, upon completing saturation mutagenesis cycles, it may be beneficial to add one final round of epPCR in the quest to further enhance the catalytic profile, be it activity, stereoselectivity, or stability.

4.3

Choosing the Best Strategy when Applying Saturation Mutagenesis

4.3.1

General Guidelines

Once the choice has been made in favor of saturation mutagenesis in the initiation of protein engineering of stereoselectivity, activity, or substrate scope, quite

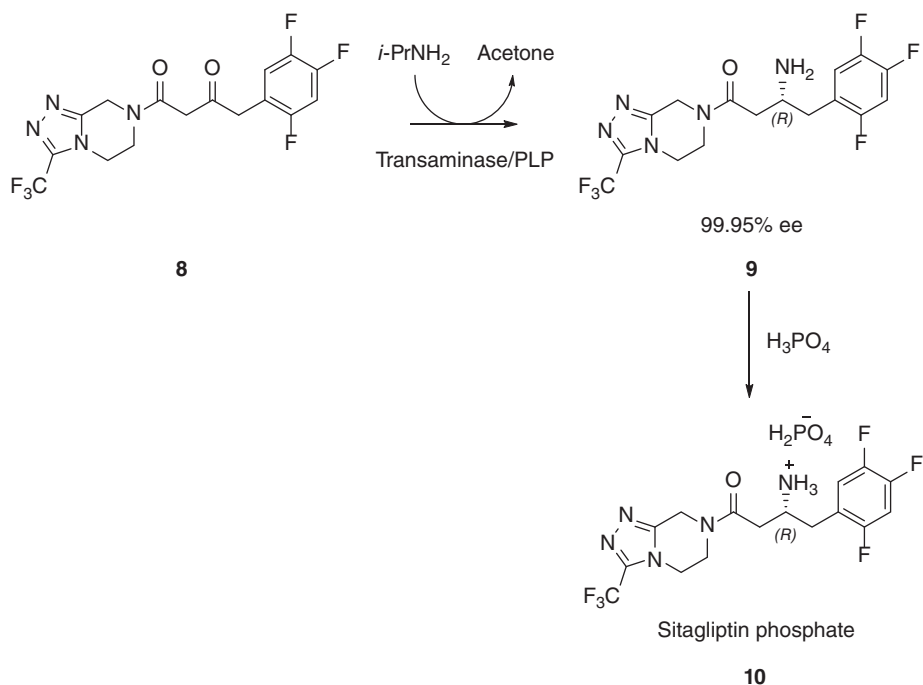
different strategies for applying this form of directed evolution are possible. Based on recent studies, the optimal approaches are now emerging. Several guidelines on how to apply saturation mutagenesis optimally have already been outlined in Sections 3.3 and 4.2, including the use of statistical metrics for estimating oversampling in relation to %-library coverage. These tips revolve around the challenge of maximizing library quality while minimizing the screening effort (bottleneck of directed evolution in general). If an enzyme under study has not been characterized by an X-ray structure needed to define CAST sites, then a homology model must be resorted to, which usually works quite well in spite of introducing some uncertainty. If no structural data of any kind is available, which is rare, epPCR, DNA shuffling, or a mutator strain should be chosen, possibly followed by saturation mutagenesis at the identified “hot spots.”

Traditionally, NNK (or NNS) codon degeneracy encoding all 20 canonical amino acids has been used in saturation mutagenesis, especially when targeting stereoselectivity [1, 8]. A crucial development for increasing the efficacy of this protein engineering technique was the utilization of reduced amino acid alphabets [10, 32] which lower structural diversity, but which constitute a simple technique for reducing the degree of oversampling in the screening step [9, 10, 12]. Saturation mutagenesis at a multi-residue site can be performed in three fundamentally different ways, each resulting in vastly different screening efforts, by utilizing:

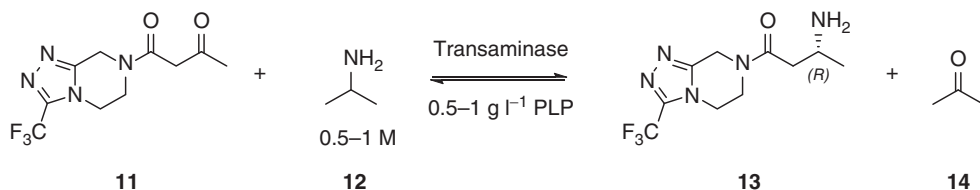
- The traditional NNK (or NNS) codon degeneracy encoding all 20 canonical amino acids at all individual positions of a multi-residue randomization site in separate saturation mutagenesis experiments [1, 8].
- One and the same codon degeneracy encoding a reduced amino acid alphabet at all positions of a multi-residue randomization site in a single saturation mutagenesis experiment [1, 8, 10].
- A different codon degeneracy at each position of a multi-residue site in a single saturation mutagenesis experiment [34].

Two representative examples of the traditional approach based on NNK codon degeneracy are analyzed here. In a study concerning the directed evolution of a stereoselective transaminase, ISM [8, 9] played a crucial role, but DNA shuffling and site-specific mutagenesis were also employed [16]. The goal was to increase the activity of the transaminase ATA-117 as a catalyst in the key step of an industrial production of sitagliptin phosphate (**10**) (Scheme 4.10).

ATA-117 is a homolog of the structurally well-characterized transaminase from *Arthrobacter* sp., which was known to be (*R*)-selective in the reductive amination of methyl ketones and small cyclic ketones [35]. At that time the industrial researchers refrained from using the “real” substrate **8** in mutagenesis/screening experiments [16], but first resorted to *in vitro* coevolution according to Scheme 4.3 [15]. Substrate walking was performed by first focusing protein engineering on the truncated methyl ketone **11** as substrate, with isopropylamine (**12**) serving as the amine-donor (Scheme 4.11). WT ATA-117 is a poor catalyst in this transformation but, nevertheless, provided 4% conversion at 2 g l⁻¹ substrate concentration.



Scheme 4.10 Biocatalytic route to sitagliptin phosphate using a transaminase evolved by applying ISM, epPCR, and DNA shuffling [16]. (Adapted and reproduced from Savile *et al.* [16]. Reproduced with permission of The American Association of the Advancement of Science.)



Scheme 4.11 Model compound (11) used in substrate walking based on *in vitro* coevolution [16]. (Adapted and reproduced from Savile *et al.* [16]. Reproduced with permission of The American Association of the Advancement of Science.)

Using a homology model of ATA-117, docking computations were performed, which allowed reasonable choices of randomization sites lining the binding pocket (CAST sites). Although extensive details were not reported, this procedure provided variant S223P with a 11-fold increase in activity in the model reaction of ketone **11** [16]. This mutant was then used as a template for ISM experiments, this time using the “real” substrate **8**. Docking experiments indicated that the trifluoromethyl group in the substrate could interact with residues V69, F122, T283, and A284. Four saturation mutagenesis libraries were created individually at these

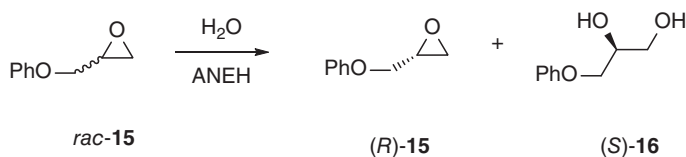
four positions, in addition to the generation of a combinatorial library using several residues simultaneously. While randomization at the single residues failed to provide improved variants, the combinatorial library led to an active variant characterized by four point mutations lining the “small” and “large” regions of the binding pocket. Double mutants F122I/V69G, F122I/A284G, F122V/V69G, F122V/A284G, F122L/V69G, and F122L/A284G proved to be hits, all containing the parent mutation S223P. Activity was still quite low, but in the absence of point mutation S223P no activity whatsoever was observed as demonstrated by a deconvolution experiment. This underscores the value of substrate walking as part of the *in vitro* coevolution approach. However, the inclusion of position 223 in an extended combinatorial library was not explored, which may have made substrate walking superfluous.

The most active variant was then used as the parent for the next round of ISM, and the beneficial mutations from the small-pocket and large-pocket saturation mutagenesis libraries were combined into a new library, which led to a variant having 12 point mutations and showing a 75-fold increase in activity. In the study nothing is mentioned regarding enantioselectivity, but it can be assumed that the researchers observed in all cases (*R*)-selectivity in favor of the desired amine **9**. Although the results at this stage are impressive, a number of practical problems still had to be solved, for example, finding reaction conditions under which the substrate is soluble, increasing substrate and amine donor concentration, solvent tuning, pH optimization as well as enhancing enzyme performance under operating conditions. Therefore, 11 further rounds of mutagenesis/screening were performed using DNA shuffling, epPCR-based random mutagenesis, rational design, and even saturation mutagenesis at second-sphere sites from the binding pocket, all of this being done parallel to process development. In some cases, mutagenesis cycles were guided by ProSAR (see Section 3.4). A total of 36 480 transformants were assayed using an LC/MS/MS screen in the early phase (low activity variants) and automated achiral HPLC system for all other transformants (≈ 2 min/sample), followed by chiral HPLC of the best hits. The best variant was reported to have 27 mutations [16]. In 50% DMSO it converts 200 g l^{-1} of the prositagliptin ketone **8** to sitagliptin (**9**) with $>99.95\%$ ee (*R*). One of the best mutants was also reported to be active and enantioselective in the reductive amination of several other structurally different ketones [16].

The catalytic performance of the best ATA-117 variant under operating conditions is clearly impressive. However, since many experimental details are lacking in this study, it is difficult to assess the actual efficiency of the reported mutagenesis strategy. Was the order of the mutagenesis cycles in the overall multi-step process planned the way it was presented? If so, why was the particular order of mutagenesis events chosen? Alternatively, were local minima in the upward climb encountered (not reported), calling for different mutagenesis techniques? With the most recent emergence of improved ISM techniques based on large randomization sites in combination with designed reduced amino acid alphabets, a saturation mutagenesis approach without the need to shuffle or invoke site-specific mutagenesis may be an attractive alternative, certainly in future studies. Nevertheless, one of

the interesting observations made in this study is the fact that saturation mutagenesis at single residue sites lining the binding pocket does not always result in the discovery of improved mutants (local minima), whereas grouping them into larger randomization sites, which are then subjected to saturation mutagenesis seems to be the better strategy [16].

The first time that ISM was applied to an enzyme in order to boost its enantioselectivity concerned the hydrolytic kinetic resolution of epoxide *rac*-**15** with formation of (*S*)-**16**, the epoxide hydrolase from *Aspergillus niger* (ANEH) serving as the biocatalyst [8]. The successes and pitfalls of this early report and subsequent studies of the model reaction are analyzed here. WT ANEH was known to be slightly (*S*)-selective ($E = 4.6$) (Scheme 4.12).



Scheme 4.12 Model hydrolytic kinetic resolution of the glycidyl ether *rac*-**15** catalyzed by ANEH [8].

Based on the crystal structure of WT ANEH [36], 15 amino acid positions lining the tunnel-like binding pocket were identified. Rather than generating 15 single residue randomization libraries at these CAST positions, they were grouped into six randomization sites comprising two or three residues: A (positions 193/195/196), B (215/217/219), C (329/330), D (349/350), E (317/318), and F (244/245/249) (Figure 4.5). All six initial libraries A–F were generated using NNK codon degeneracy and screened for enhanced enantioselectivity in the model reaction, the best variant (L215F/A217D/R219S) originating from library B. This mutant was then used as the template in the otherwise arbitrarily chosen ISM pathway WT $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D $\rightarrow$ F $\rightarrow$ E leading to the best variant LW202 characterized by nine point mutations and showing notably improved enantioselectivity ($E = 115$) [8]. Due to the already high stereoselectivity, the last site A in the series was not visited. A total of about 20 000 transformants were screened using the Mülheim multiplexing MS-based high-throughput ee-screening system [37] (see Chapter 2). Limited alternative ISM exploration was performed, pathway WT $\rightarrow$ B $\rightarrow$ C $\rightarrow$ D $\rightarrow$ E leading to a variant with a selectivity factor of $E = 49$ [8b], but continuing to site A or F failed to produce improved variants (local minima in the fitness pathway landscape). It took several years of research to figure out a way to escape from such dead ends (see following text), and to improve the ISM-based strategy so that typically only 2000–4000 transformants need to be screened.

Although this study introduced ISM for manipulating the catalytic profiles of enzymes, several methodological aspects were neglected [8]. Firstly, the mandatory statistical analysis regarding screening/oversampling, which today is routine

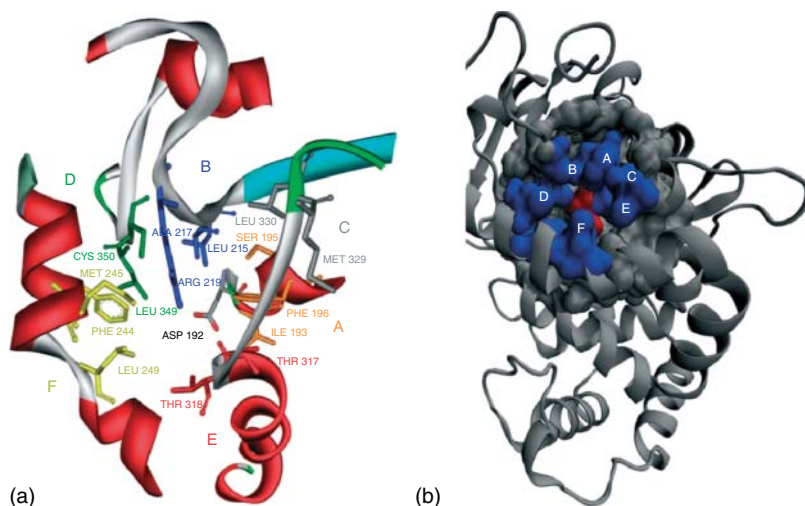


Figure 4.5 CAST sites A–E [8] of the epoxide hydrolase from *Aspergillus niger* (ANEH) chosen on the basis of the X-ray crystal structure of the WT [36]. (a) Defined randomization sites A (orange), B (blue), C (gray), D (green), and E (yellow). (b) Top view of tunnel-like ANEH binding pocket showing sites A–E (blue) and the catalytically active Asp192 (red). (Zou *et al.* [36]. Reproduced with permission of Elsevier.)

[1c, 10], was not performed. Later calculations showed that 95% library coverage was not ensured, again showing that less screening can still be successful in finding improved variants. Secondly, no attempt was made to escape from a local minimum on the fitness landscape. Thirdly, further ISM pathways were not considered. Fourthly, library quality was not checked by the Quick Quality Control (QQC) [3] nor by the more quantitative Q-values [38], which were introduced a few years later. Therefore, one may consider the evolution of the best variant LW202 to be a result of fortune. Subsequent and likewise successful ISM studies [1c, 13], which also rely on arbitrarily chosen pathways, speak against this hypothesis. Dozens of other successful ISM-based studies have appeared as summarized in reviews [32, 38]; see also Chapter 5. Even in the case of B-FIT based thermostabilization of a lipase, highly improved variants were evolved by choosing arbitrary ISM pathways [9]. Nevertheless, some uncertainty remains. Even with today's best theoretical QM/MM techniques, it is impossible to predict the optimal upward pathway in an ISM scheme.

4.3.2

Choosing Optimal Pathways in Iterative Saturation Mutagenesis (ISM)

It has become clear that some ISM pathways are more productive than others are, and some may be characterized by local minima, meaning the absence of any improved variants in a given library. Such “dead ends” are not restricted to ISM,

they are a universal phenomenon when applying other gene mutagenesis techniques such as epPCR, DNA shuffling, or mutator strains [1]. In order to address this fundamental problem, a complete ISM landscape was explored, specifically all 24 pathways of a 4-site system in which all 64 relevant mutant libraries were generated and screened for enantioselectivity [31]. It involves the epoxide hydrolase ANEH as the catalyst in the same model reaction (Scheme 4.12). However, rather than constructing the respective complete 5-site ISM scheme [8], which would require excessive exploration of 120 pathways, a truncated version based on only four CAST sites in ANEH was designed [31]. It requires considerably less laboratory work because only 24 pathways have to be constructed, involving a total of 64 mutant libraries. In order to reduce the experimental work further, the original three-residue sites B (215/217/219) and F (244/245/249) were truncated to two-residue sites B* (215/219) and F* (244/249). Moreover, NDT codon degeneracy was chosen encoding only 12 amino acids (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, and Gly), which likewise requires less screening for 95% library coverage (only about 430 transformants) [31]. These practical measures reduce structural diversity quite a bit, but this approach provided variants that are even better than the previous best mutant LW202. Moreover, the complete fitness pathway landscape reveals several phenomena that are important when applying ISM.

The final results of the 12 most productive ISM pathways are shown in Figure 4.6a, the respective terminal variants showing *E*-values in the range of 78–159 favoring (*S*)-**16**. The results of the other set of 12 trajectories are pictured in Figure 4.6b, which features variants displaying selectivity factors in the range of *E* = 28–78, which are also considerably better than the performance of WT ANEH (*E* = 4.6). Thus, all 24 pathways provide respectable results, the best one WT → F* → B* → E → D leading to variant GUY-228 (*E* = 159) [31], which is superior to the originally best mutant LW202 (*E* = 115) [8]. These results help to explain why arbitrarily chosen pathways in numerous ISM studies have been successful, although better options may have been missed [32a].

The results allow for the construction of a fitness-pathway landscape by considering the experimental data at all stages of a given pathway in a stacking mode that links WT ANEH with the respective final mutant in each case of all 24 trajectories (Figures 4.7 and 4.8).

Figure 4.6a,b reveals several local minima, meaning that in the respective libraries no improved variants were found. The evolutionary pathways also point to an important result in this study, namely a simple way to escape from such dead ends. Rather than abandoning a pathway characterized by a local minimum, or to switch to epPCR or DNA shuffling, the use of an inferior mutant showing lower enantioselectivity, as the template in the subsequent ISM step leads to a notably improved variant. One of several such pathways is WT (*E* = 4.6) → E (*E* = 21) → B* (*E* = 36) → D (*E* = 32) → F* (*E* = 97). The use of inferior mutants is reminiscent, but not identical to the concept of neutral drift [4a–c], and of the Eigen/Schuster concept of quasi-species in natural evolution [5] which has been invoked in other directed evolution studies [6, 39]. As the most general

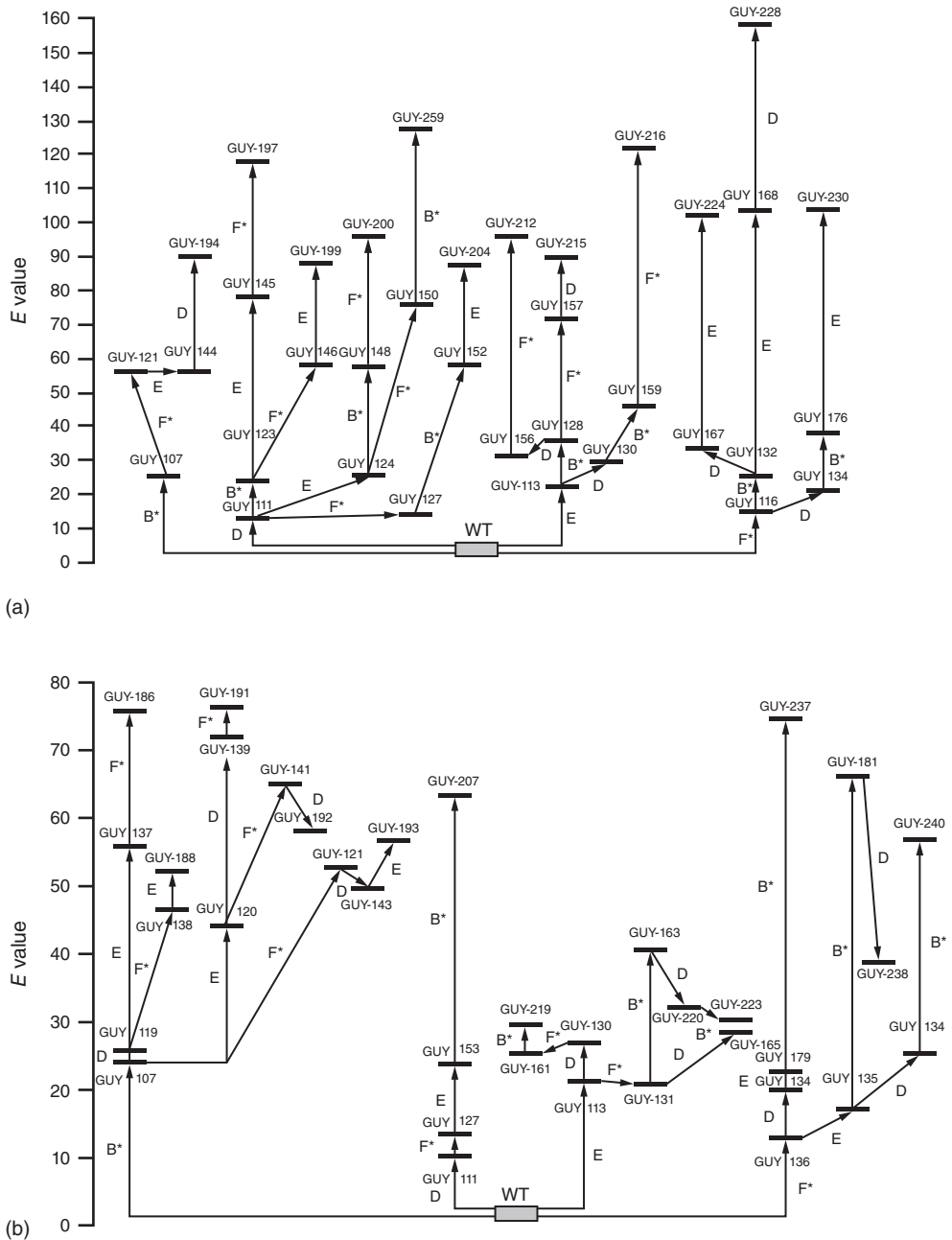


Figure 4.6 Complete experimental exploration of a 24-pathway ISM system involving the ANEH-catalyzed hydrolytic kinetic resolution of *rac*-15 (Scheme 4.12). (a) Portion of the 24-pathway ISM scheme showing the 12 best pathways leading to ANEH variants

displaying $E > 78$ (S) and (b) portion of the 24-pathway ISM scheme showing the 12 least productive pathways leading to ANEH variants with $E = 28-78$ (S) [31]. (Gumulya *et al.* [31]. Reproduced with permission of John Wiley & Sons.)

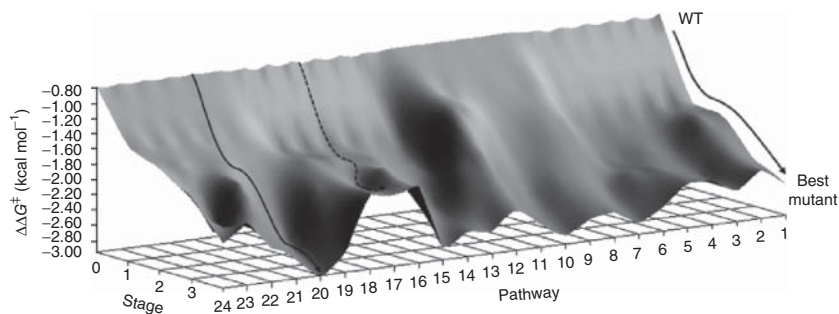


Figure 4.7 Fitness pathway landscape featuring the 24 trajectories leading from WT ANEH to the respective final variants with enhanced enantioselectivity at the end of each pathway as specified by the respective $\Delta\Delta G^\ddagger$ values [31]. Solid line: typical pathway in which each mutant library contains at least one variant displaying enhanced

enantioselectivity; dotted line: typical pathway in which at least one library is devoid of an improved variant, in which case an inferior mutant was employed in the subsequent ISM step, thereby escaping from the local minimum. (Gumulya *et al.* [31]. Reproduced with permission of John Wiley & Sons.)

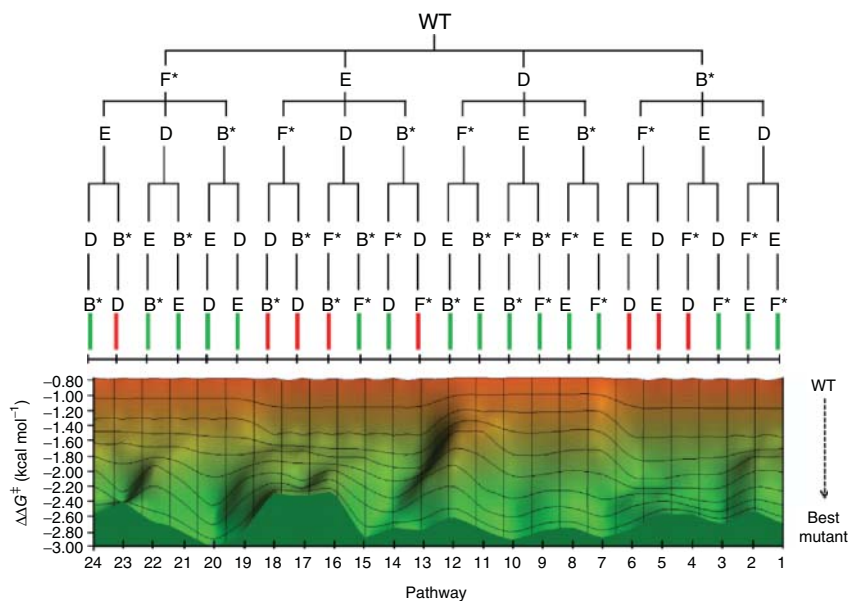


Figure 4.8 Free energy profiles of the 24 ISM pathways in the directed evolution of ANEH as pictured in a front view of the fitness-pathway landscape [31]. In the green pathways all relevant saturation mutagenesis libraries contain improved variants (enhanced enantioselectivity) in the model

reaction (Scheme 4.12); the eight red pathways denote those in which at least one library in the step evolutionary process is devoid of any improved variants (local minimum). (Gumulya *et al.* [31]. Reproduced with permission of John Wiley & Sons.)

conclusion of the study, the use of inferior mutants as templates in subsequent mutagenesis cycles is recommended whenever local minima are encountered in any directed evolution study, including those driven by epPCR or DNA shuffling. Thus far, this strategy has not been used very often in directed evolution [40] because researchers usually prefer to abandon the respective experimental platform and to switch to another mutagenesis method [1]. When analyzing the experimental data, the first derivative of $\Delta\Delta G^\ddagger$ at every stage of each of the 24 pathways is also instructive (Figure 4.9). The green areas below the XY -plane indicate those ISM pathways in which each and every library along the four-step trajectory harbors an improved hit displaying higher enantioselectivity in the model reaction $rac\text{-}\mathbf{15} \rightarrow (R)\text{-}\mathbf{15} + (S)\text{-}\mathbf{16}$ (Scheme 4.12), which was then used as a template in the subsequent saturation mutagenesis experiment. Red areas above the XY -plane indicate positive slopes, meaning pathways that are characterized by local minima, that is, only inferior mutants showing lower enantioselectivity. The best of the inferior mutants was then used in the subsequent ISM step.

As Table 5.1 in Chapter 5 reveals, ISM is being increasingly used to manipulate stereo- and regioselectivity as well as activity and substrate acceptance. The areas of application generally concern the biocatalytic synthesis of enantiopure intermediates needed in the preparation of chiral pharmaceuticals, plant protecting agents, or fragrances [32]. ISM has also been applied in completely different areas, for example, in the directed evolution of N -oligosaccharyltransferases of relevance in vaccine production [41].

Yet another area of application is the potential production of universal blood by enzyme-catalyzed selective cleavage of the antigenic components [42]. Decades ago a vision was put forward calling for selective and active glycoside hydrolases [43]. Unfortunately, all presently known glycoside hydrolases that catalyze such challenging reactions are so slow that the original vision could not be implemented experimentally. The situation changed when ISM-based

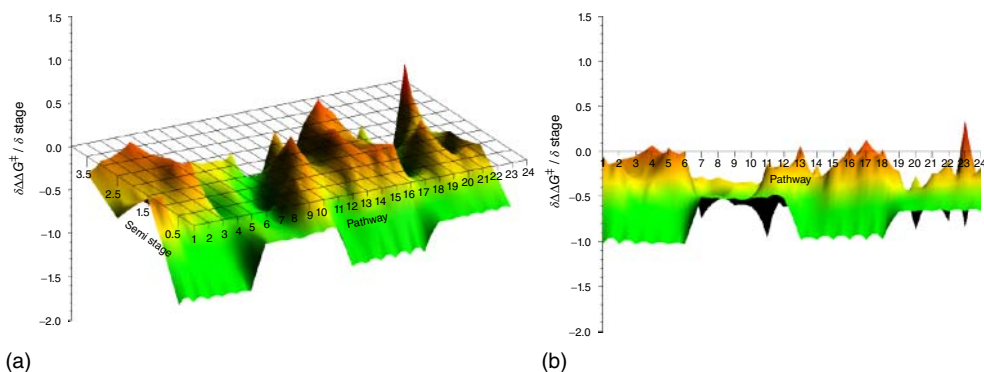


Figure 4.9 First derivative of $\Delta\Delta G^\ddagger$ at every stage of each of the 24 ISM pathways in the directed evolution of ANEH (see Figures 4.7 and 4.8). (a) View from top of

fitness pathway landscape and (b) view from the side [31]. (Gumulya *et al.* [31]. Reproduced with permission of John Wiley & Sons.)

directed evolution was applied to the glycoside hydrolase from *Streptococcus pneumoniae* SP3-BS, which cleaves selectively the entire terminal trisaccharide antigenic determinants of both A- and B-antigens from some of the linkages on the surface glycans of red blood cells [42]. In order to enhance enzyme activity for the cleavage of the Gal β -1,3-GlcNAc linkage of type 1A antigens (Figure 4.10a–d), a high-throughput microtiter-based assay was first established in which a fluorogenic substrate was devised comprising the type 1A blood group pentasaccharide linked by a β -glycosidic bond to a methylumbelliferone.

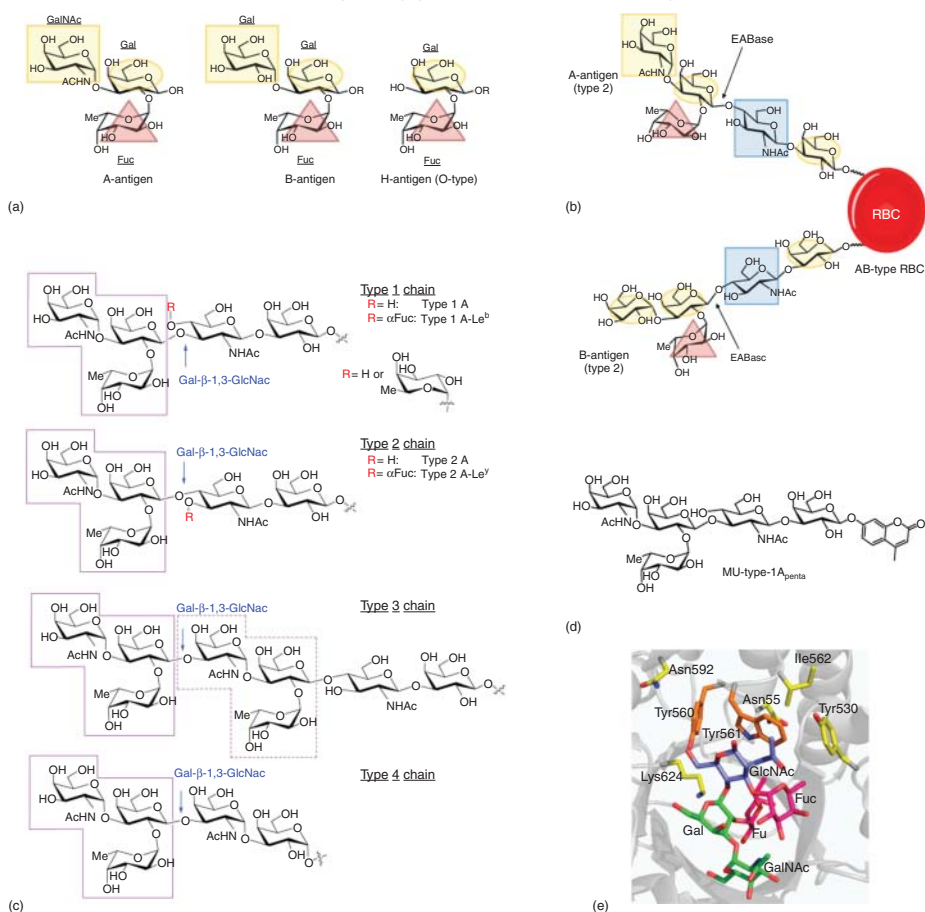
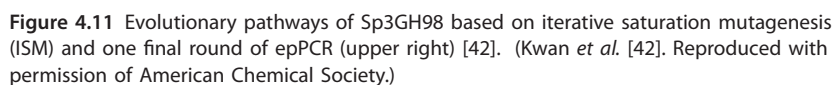


Figure 4.10 Toward universal blood [42]. (a) Carbohydrate antigenic determinants of A-, B-, and H-antigens. The H-antigen is present on glycans of the O blood-group, and typically nonantigenic except in rare cases. (b) Site of cleavage of A- and B-antigens by GH98 EABase enzymes from type 2 chains of erythrocytes. (c) Various chain types to which A-antigens are present on erythrocytes and

other cell types. (d) Structure of the fluorogenic substrate MUType1Apenta. (e) First- and second-sphere randomization sites chosen for iterative saturation mutagenesis (ISM), guided by the X-ray structure of Sp3GH98. First sphere: Tyr 560 and Trp561; second sphere: Tyr 530, Asn559, Ile 562, Asn592, and Lys624. (Kwan *et al.* [42]. Reproduced with permission of American Chemical Society.)

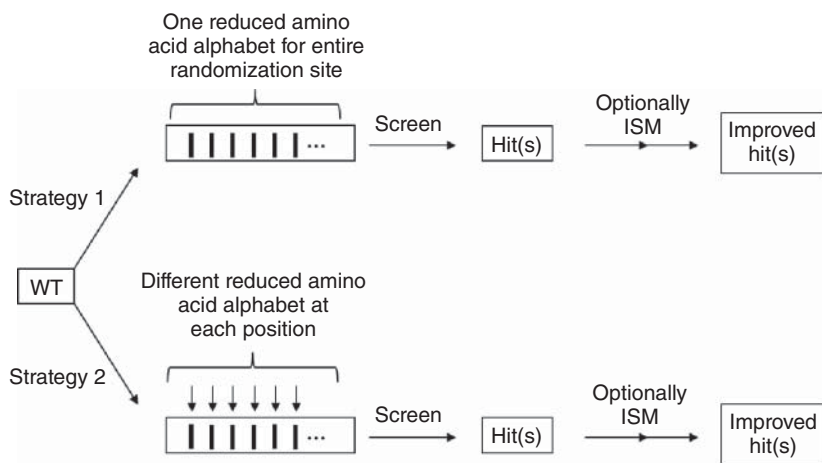
As analyzed in a highlight featuring this research, the overall strategy has opened a new door in this exciting research field [45]. It was suggested that further ISM experimentation is likely to provide even better results. This could involve the grouping of the individual amino acid positions into multi-residue sites followed by ISM on the basis of appropriately chosen reduced amino acid alphabets.



4.3.3

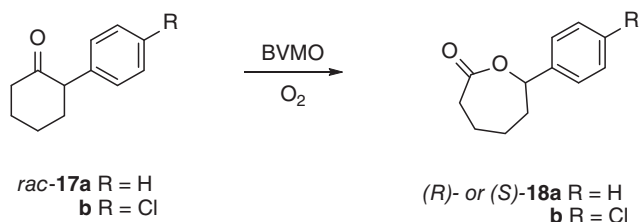
Systematization of Saturation Mutagenesis

Following the initial introduction of reduced amino acid alphabets in saturation mutagenesis employing one and the same reduced amino acid alphabet for simultaneous randomization of a multi-residue site [10], an alternative strategy was developed for saturation mutagenesis employed in the directed evolution of stereoselective enzymes: A different codon degeneracy is used at each position of a multi-residue site in a single saturation mutagenesis experiment [34a]. The two distinctly different strategies are illustrated in Scheme 4.13. Both approaches make use of X-ray structures, consensus sequence data, and/or computational guidance. Relative to strategy 1, strategy 2 requires a greater number of specific decisions regarding the optimal choice of amino acids as building blocks, which may in some cases be difficult. However, when reliable guesses can be made on the basis of sound experimental or computational data, strategy 2 may constitute the preferred option. In both approaches a large randomization site, for example, one comprising eight or more residues, can be split into several smaller ones in order to reduce the screening effort.



Scheme 4.13 Two different approaches to the use of reduced amino acid alphabets in saturation mutagenesis, if necessary followed ISM [1b, 10, 12, 34].

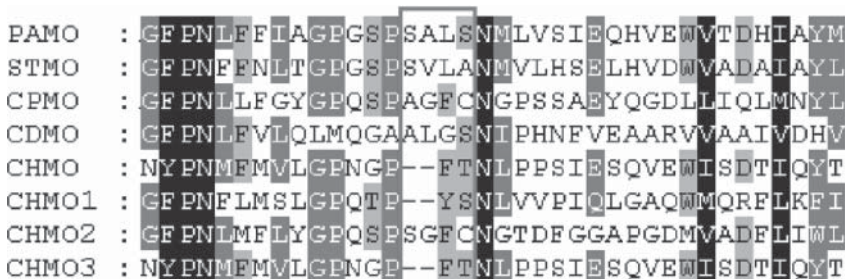
In some cases the consensus approach based on multiple sequence alignment (MSA), originally developed for protein thermostabilization [46], constitutes a useful guide when focusing on activity and/or stereoselectivity. The first example of strategy 2 for enhancing stereoselectivity was the directed evolution of phenyl acetone monooxygenase (PAMO) [34a]. This robust Baeyer–Villiger monooxygenase readily accepts phenyl acetone, but not such synthetically interesting compounds as cyclohexanone or its derivatives [47]. For example, ketones **17a–b** are essentially inert to PAMO-catalyzed oxidation (Scheme 4.14) [34a]. Thus, the goal was to evolve substrate acceptance (rate) as well as



Scheme 4.14 Oxidative kinetic resolution catalyzed by PAMO mutants [34a]. (Reetz and Wu [34a]. Reproduced with permission of Royal Society of Chemistry.)

enantioselectivity in the oxidative kinetic resolution of these compounds using saturation mutagenesis guided by MSA at residue of a randomization site at the binding pocket.

First, on the basis of the PAMO crystal structure [48] four residues in loop 441–444 next to the binding pocket were identified as possible CAST sites. Previous application of rational design using site-specific mutagenesis at positions in the loop was only partially successful [49]. NNK-based randomization of a four-residue CAST site would require the screening of 3.1 million transformants for 95% library coverage, and even NDT codon degeneracy would still call for $\approx 62\,000$ screened clones (Table 3.3). Although considerably less library coverage may still provide improved variants [20], this was not tested in the study. Rather, eight Baeyer–Villiger monooxygenases were aligned with focus on the loop region (Scheme 4.15). As can be seen, only a limited number of amino acids are conserved at the four positions: Ser and Ala (position 441), Ala, Val, Gly, and Leu (position 442), Leu, Phe, Gly, and Tyr (position 443) and Ser, Ala, Cys, and Thr (position 444). The plan was to use these amino acids as building blocks at the respective positions of the four-residue randomization site, the degree of oversampling being highly reduced by this strategy [34a].



Scheme 4.15 Sequence alignment of BVMOs (441–444 loop in gray box) [34a]. (Reetz and Wu [34a]. Reproduced with permission of Royal Society of Chemistry.)

Appropriate codon degeneracies were designed in order to match the amino acids occurring at these four positions as extensively as possible, while also introducing a limited number of additional amino acids as building blocks for

Table 4.2 Choice of codon degeneracies at each position in the 441–444 loop of PAMO.

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

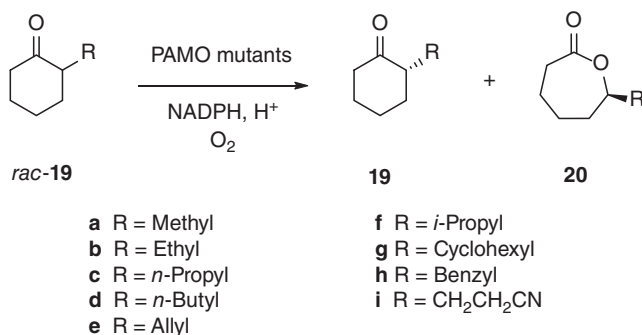
Degenerate codons: A (adenine); B (cytosine/guanine/thymine); C (cytosine); G (guanine); S (cytosine/guanine); K (guanine/thymine); N (adenine/cytosine/guanine/thymine) [34a]. In the column listing the encoded amino acids, the WT amino acids are shown in parentheses.

Source: Reetz and Wu [34a]. Reproduced with permission of Royal Society of Chemistry.

slightly enhanced diversity (Table 4.2). At all positions the WT amino acid is maintained as defined by the chosen codon degeneracy. At position 441, KCA codon degeneracy means the introduction of only one new amino acid (in this case Ala), while structural diversity at the other positions is higher.

Upon screening only 1700 transformants (2587 would be necessary for 95% library coverage), several active hits were identified, PAMO variant Ser441Ala/Ala442Trp/Leu443Tyr/Ser444Thr showing the highest activity and enantioselectivity ($E = 70$ in favor of *R*-**18a**). This variant is an even better catalyst for the reaction of substrate **17b** ($E > 200$), which is also not accepted by wild-type phenyl acetone monooxygenase WT PAMO. Two conclusions result from this proof-of-principle study: (i) a different reduced amino acid alphabet can be used effectively at each position within a multi-residue randomization site in a single saturation mutagenesis experiment for stereoselectivity enhancement and (ii) MSA-based bioinformatics constitutes a practical guide in this endeavor [34]. MSA-guided saturation mutagenesis using a defined amino acid alphabet for the entire randomization site has since been applied and extended to other stereoselectivity-relevant directed evolution studies [50].

Even more challenging is protein engineering of PAMO as catalysts in the oxidative kinetic resolution of inert 2-alkyl substituted cyclohexanone derivatives. Variant Ser441Ala/Ala442Trp/Leu443Tyr/Ser444Thr is not active in these cases. Rather than resorting to *in vitro* coevolution by applying substrate walking (which could well be successful) [15], a completely different strategy was tested [51]. Further analysis of the sequence alignment (Scheme 4.16) showed that proline at positions 437 and 440 is highly conserved, cyclododecanone monooxygenase (CDMO) being the only exception. This suggests that proline, known to impart some degree of rigidity to proteins [52], is necessary at these positions for maintaining stability and function. By standard logic, it should not be exchanged by another amino acid. Nevertheless, in hope of evolving a new catalytic profile, the opposite strategy was implemented by performing NNK-based saturation mutagenesis at positions 437 and 440 (a second sphere CAST residue) [51]. This unorthodox MSA-guided approach proved to be successful, the randomization library at position 440 containing several highly active and stereoselective single



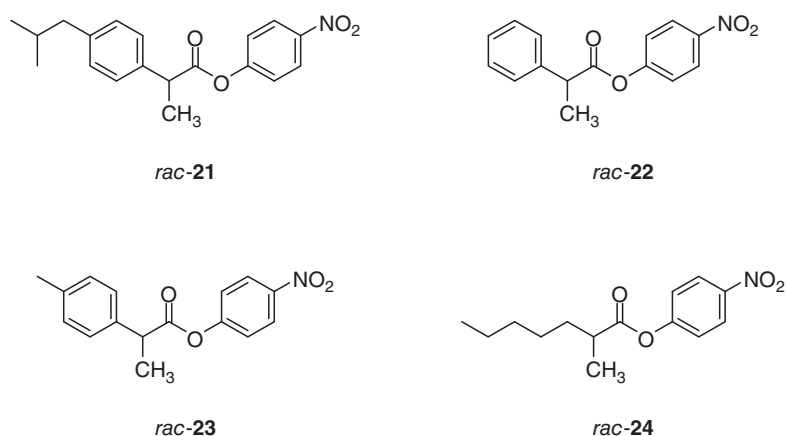
Scheme 4.16 PAMO-catalyzed oxidative kinetic resolution of 2-alkyl substituted cyclohexanone derivatives [51]. (Reetz and Wu [34a]. Reproduced with permission of American Chemical Society.)

mutants with essentially no trade-off in terms of stability. Variants Pro440Leu, Pro440Ile, Pro440Asn, Pro440His, Pro440Tyr, and Pro440Trp are active and highly stereoselective in oxidative kinetic resolution of the previous ketones **17a–b**, but also in reactions of a wide variety of structurally different 2-alkyl derivatives **19a–i** (Scheme 4.16), selectivity factors of $E = 150–200$ being typical. Even the bulky substrate 2-cyclohexylcyclohexanone (**19g**) is readily accepted with high enantioselectivity and activity ($E > 200$).

Increased flexibility of the extended PAMO loop may be the reason for the dramatic effects, but a thorough theoretical analysis of the experimental results has yet to be performed. It is also uncertain how general the MSA-based strategy is, that is, whether amino acid exchange events at highly conserved residues remote or near the binding pocket are successful when evolving other enzymes. The opposite strategy using combinatorial proline scanning with introduction of one or more prolines at CAST sites (or remote loops) also needs to be explored.

A practical and useful application of the approach based on the use of a different amino acid alphabet at each position of a multi-residue site (strategy 2 in Scheme 4.13) was reported in the directed evolution of *Candida antarctica* A (CALA) as a biocatalyst in the hydrolytic kinetic resolution of α -substituted carboxylic acid esters [34b]. An earlier study by the same group had shown that CAST/ISM is successful in evolving a number of variants for asymmetric transformations of this kind [53], but extension to more bulky ibuprofen-type esters was not achieved. The new study concentrated on substrates **21–24** (Scheme 4.17) by first applying substrate walking [15]. A triple mutant (F149Y/I150N/F233G) that had proven to be highly active and enantioselective for substrate **18** using CAST/ISM, but which was a poor catalyst for the larger substrate **21** (slow and stereorandom reaction), was used as the starting template for further saturation mutagenesis experiments [34b].

The bulky substrate **21** was docked inside the CALA binding pocket in the oxyanion form (tetrahedral intermediate at Ser184), leading to the conclusion that nine residues at the acyl binding region should be mutable (Figure 4.12).



Scheme 4.17 Substrates investigated in the saturation mutagenesis based directed evolution of CALA using mutant F149Y/I150N/F233G as template [34b]. (Adapted from Sandström *et al.* [34b]. Reproduced with permission of National Academy of Sciences.)

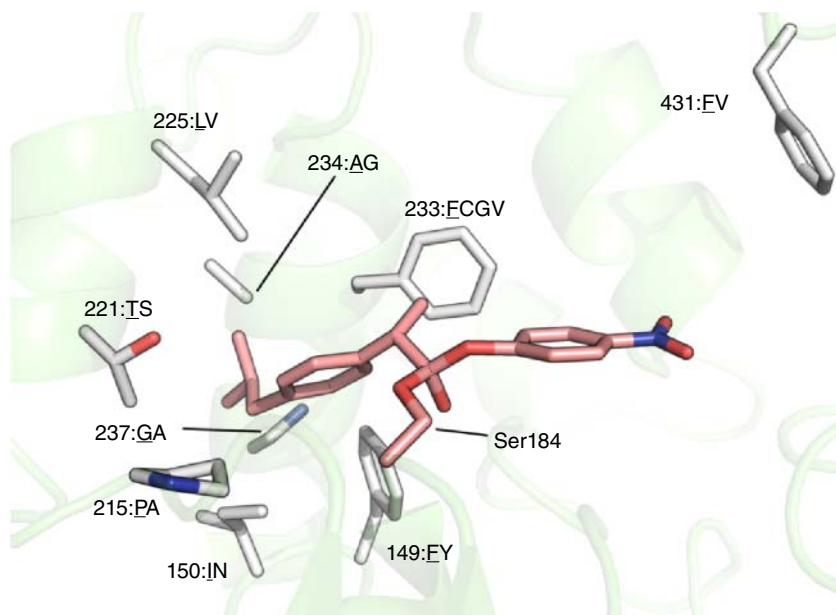


Figure 4.12 Binding pocket of CALA showing tetrahedral intermediate with substrate 21 and nine residues for potential saturation mutagenesis [34b]. The original WT residues

are underlined. (Adapted from Sandström *et al.* [34b]. Reproduced with permission of National Academy of Sciences.)

Other residues at this large CAST site were eliminated from further consideration because they proved to be highly conserved as shown by an MSA (PSI-BLAST) analysis. In view of the PAMO study (Scheme 4.16) [51], this may not have been necessary, but nine otherwise well-chosen positions sufficed for reshaping the binding pocket. Since substrate **21** is too bulky to be readily accepted, small amino acids as building blocks in reduced amino acid alphabets were chosen for saturation mutagenesis. Phe149Tyr and Ile150Asn were included because in the earlier study this pair had been shown to be essential in the evolution of high enantioselectivity toward a fairly broad range of similar substrates [53].

Sequence proximity in the nine chosen residues allowed ready clustering into four primer pairs. Together with primers for the ends of the gene, they were used in generating five partly overlapping fragments. These were then simultaneously assembled in a modified overlap extension polymerase chain reaction (OE-PCR) (see Chapter 3). The final set of amino acids as building blocks for simultaneous randomization at the nine-residue CAST site was chosen on the basis of structural considerations and perhaps some intuition (Table 4.3) [34b].

About 2400 transformants were screened corresponding to $\approx 90\%$ coverage of a highly condensed library, but only a few variants proved to be active toward substrate **21**. The best hit was a penta-substituted variant Thr221Ser/Leu225Val/Phe233Cys/Gly237Ala/Phe431Val in which four different amino acids were introduced at five different positions, leading to high stereoselectivity ($E = 100$). The best hit as well as other CALA variants in the hydrolytic kinetic resolution of the other substrates in Scheme 4.17 likewise ensured acceptable levels of enantioselectivity. Using NNK codon degeneracy, encoding all 20 canonical amino acids for 95% library coverage would have required the screening of 10^{14} potentially enantioselective clones, an impossible mission. Deconvolution experiments revealed cooperative (more than additive) effects [14], and suggested that the particular penta-substituted variant would

Table 4.3 Combinatorial use of amino acids as building blocks employed in saturation mutagenesis at the nine-residue randomization site of CALA (Figure 4.12) [34b].

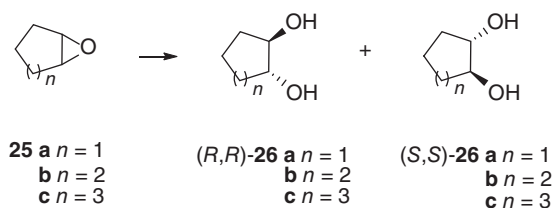
Position	WT residue	Alternative residue(s)
149	Phe	Tyr
150	Ile	Asn
215	Pro	Ala
221	Thr	Ser
225	Leu	Val
233	Phe	Cys/Gly/Val
234	Ala	Gly
237	Gly	Ala
431	Phe	Val

Source: Adapted from Sandström *et al.* [34b]. Reproduced with permission of National Academy of Sciences.

not be accessible by ISM [34b]. Extensive ISM was not applied, but, in principle, such exploration could provide many more mutants of different sequences that also feature highly improved catalytic profiles.

The most interesting aspect in the CALA study [34b] was the use of codon degeneracies, which led to the combinatorial introduction of a single defined amino acid at eight positions (rather than just one position as in the PAMO study [34a]) and three amino acids at the ninth position (in addition to the respective WT amino acid, respectively) [34b]. A small and smart library resulted in one step, as in the PAMO study [34a], thereby avoiding the necessity to probe several cycles of mutagenesis/expression/screening (step economy). The option defined by strategy 2 (Scheme 4.13) was also chosen in an investigation of CALA as catalyst in the acylating kinetic resolution of secondary alcohols [34c]; in a study of the evolution of P450-BM3 mutants, which accept small alkanes such as propane [34d]; and in other applications [34e–i].

The molecular biology in the two structure-guided utilizations of reduced amino acid alphabets in saturation mutagenesis is straightforward, but it is not clear whether strategy 1 [9, 10, 13, 38] or strategy 2 [34] is superior (Scheme 4.13). They need to be studied more thoroughly before final assessments are made. When comparing NNK with NDT codon degeneracy as part of strategy 1, the benefits of resorting to the reduced amino acid alphabet have been demonstrated [10a], but in that study alternative reduced amino acid alphabets were not tested. In contrast, strategy 1 has been tested fairly systematically by studying the consequences of utilizing different reduced amino acid alphabets, limonene epoxide hydrolase (LEH) from *Rhodococcus erythropolis* DCL14 serving as the catalyst in the hydrolytic desymmetrization of meso-type epoxides (Scheme 4.18) [54]. LEH is mechanistically different from other epoxide hydrolases because epoxide substrates undergo S_N2 reactions with water acting as the nucleophile (not an aspartate-residue), which has been modeled in a QM/MM study [55a,b] on the basis of the LEH crystal structure [55c]. In the initial report, the conventional CAST/ISM approach was successfully applied using NDT codon degeneracy in the randomization of four two-residue sites Met32/Leu35, Leu74/Ile80, Leu114/Ile116, and Met78/Val83 lining the LEH binding pocket (CASTing) [54]. Employing epoxide **25a** as the model substrate and arbitrarily chosen ISM pathways, both (*R,R*)- and (*S,S*)-selective mutants were evolved,



Scheme 4.18 Hydrolytic desymmetrization of meso-epoxides catalyzed by LEH and mutants thereof [54, 55a, 56a,b].

enantiomeric ratios reaching $er = 90:10$ and $er = 96:4$, respectively. A total of 5000 transformants were screened.

4.3.4

Single Code Saturation Mutagenesis (SCSM): Use of a Single Amino Acid as Building Block

The initial LEH study [54] was followed by a second investigation, this time posing the curiosity-driven question whether the smallest amino acid alphabet, namely a single amino acid as building block for randomization, can be used effectively in saturation mutagenesis at a large multi-residue site lining the binding pocket [56a]. It constitutes the extreme case of minimal building blocks in reshaping enzyme binding pockets according to strategy 1, and is therefore fundamentally different from the CALA-study (strategy 2 in Scheme 4.13), in which up to four different amino acids were introduced in the final mutant (Table 4.3) [34b]. The hydrolytic desymmetrization of cyclohexene oxide (**25b**) with formation of (*R,R*)-**26b** and (*S,S*)-**26b** served as the model reaction. Ten CAST residues were identified for saturation mutagenesis (Leu74, Phe75, Met78, Ile80, Leu103, Leu114, Ile116, Phe134, Phe139, and Leu147), which were grouped into a single randomization site (Figure 4.13) [56a,b]. WT LEH is only slightly (*S,S*)-selective ($er = 52:48$; $ee = 4\%$). If such a 10-residue site were to be randomized using NNK

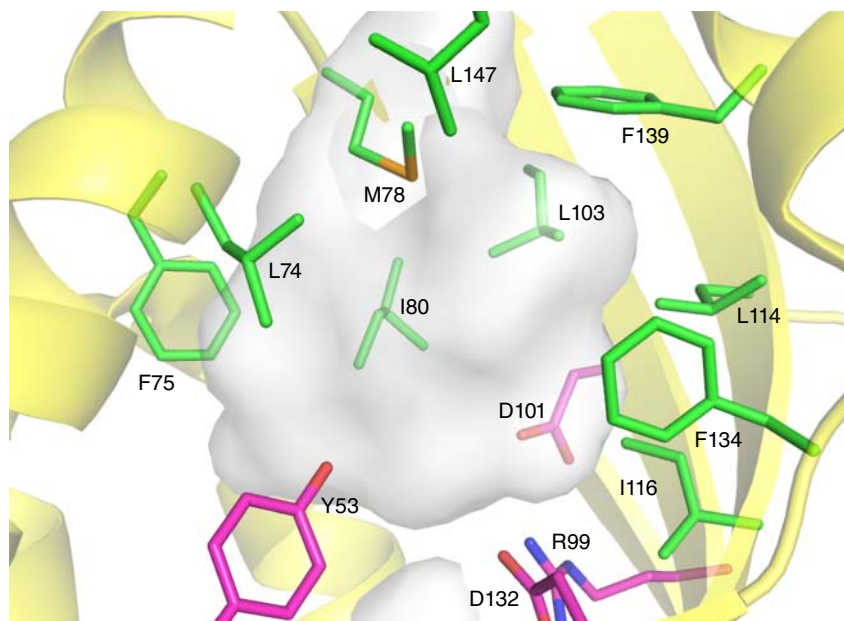
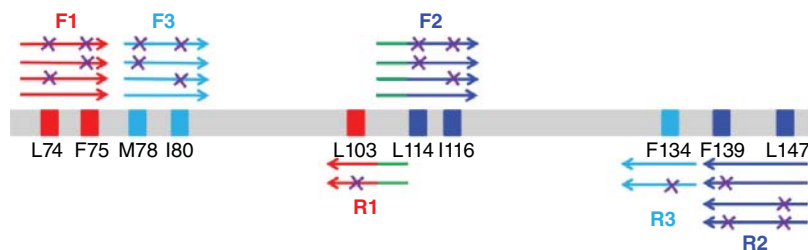


Figure 4.13 Large randomization site defined by 10 amino acid positions (green) chosen on the basis of the crystal structure of LEH [55c] with the catalytic residues being shown in pink [56a,b].

codon degeneracy (20 amino acid alphabet) or NDT codon degeneracy (12 amino acid alphabet), then one would have to screen for 95% library coverage about 10^{15} or 10^{11} transformants, respectively. In contrast, when employing the smallest amino acid alphabet, a single amino acid, only about 3000 transformants, would have to be screened for essentially complete library coverage. However, structural diversity would be dramatically reduced. It was speculated that such a strategy could be successful if the right decision were to be made regarding the choice of the amino acid in single codon saturation mutagenesis (SCSM). The crystal structure of LEH reveals that most of the amino acids surrounding the binding pocket are hydrophobic [55c].

Therefore, valine was chosen as the sole building block in saturation mutagenesis at the 10-residue randomization site. This procedure is reminiscent of combinatorial alanine scanning used in changing the binding properties of proteins [57]. In one earlier case, combinatorial alanine scanning was applied in order to increase the activity of a P450 monooxygenase (although activity was measured using lysates and not isolated mutants, which causes some uncertainty in the assessment of activity) [57c]. The reason for choosing valine in the LEH study as opposed to alanine has to do with the expectation that the sterically more demanding side-chain of this amino acid would compensate for lowest structural diversity. Primer design is shown in Scheme 4.19 [56a].



Scheme 4.19 Primer design and library construction using valine as the sole building block and the 10 randomization positions in LEH according to Figure 4.13 [56a]. (Sun *et al.* [56a]. Reproduced with permission of John Wiley & Sons.)

Using the adrenaline pre-test for epoxide activity [58], 35 microtiter plates of 96-well format were screened ($96 - 2$ positive controls $- 2$ negative controls = 92 corresponding to 3220 transformants) [56a]. A large number of active hits (533) were discovered, which is an indication that this strategy is productive. These were then assessed for enantioselectivity by automated chiral GC. The result is shown in Figure 4.14, which features seven of the best hits discovered in this single mutant library, four of them favoring (*S,S*)-**26b** and three being (*R,R*)-selective, with enantiomeric ratios up to $er = 97:3$ and $er = 12:88$, respectively. It can be seen that three to five valines were introduced at different positions of the 10-residue site [56a].

The use of phenylalanine as the sole amino acid as building block in an analogous manner also worked well, although the degree of enantioselectivity was generally

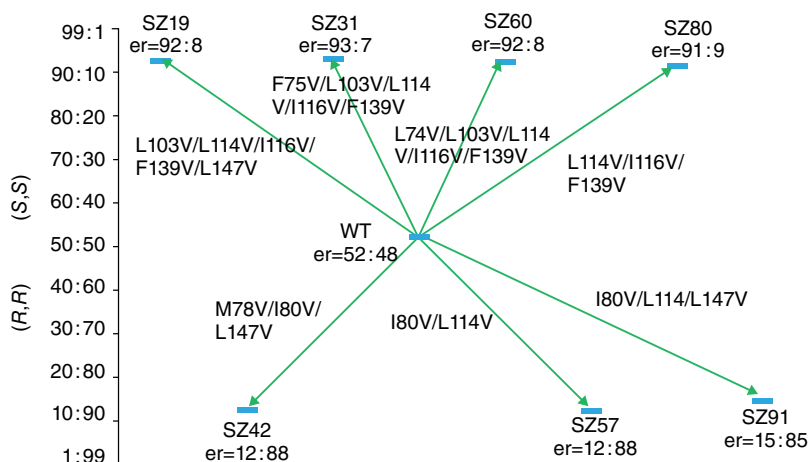


Figure 4.14 Best hits discovered in a mutant library created by a single saturation mutagenesis experiment using valine as the sole building block at a 10-residue randomization site in LEH serving as the catalyst in

the hydrolytic desymmetrization of epoxide **25b** (Scheme 4.18) [56a]. (Sun *et al.* [56a]. Reproduced with permission of John Wiley & Sons.)

somewhat lower. ISM employing valine and phenylalanine in successive randomization steps as part of ISM led to the highest enantioselectivities (enantiomeric ratios up to 98 : 2). In further control experiments, other amino acids such as serine or proline as the sole building blocks failed to provide improved mutants, which is of mechanistic significance because it demonstrates that the rationale behind choosing valine was correct. (*R,R*)- and (*S,S*)-selective mutants were characterized by X-ray structural analyses which revealed the distinct changes in the shape of the respective binding pockets [56a].

Although the apparent success when using the smallest reduced amino acid alphabet as part of strategy 1 (Scheme 4.13) is indisputable, it cannot be expected to be general. The choice of the best single amino acid in addition to WT was fairly straightforward in the epoxide hydrolase case because the binding pocket of WT is surrounded by hydrophobic amino acids. However, in the case of many (if not most) other enzymes, the structural situation is more complex, making the correct choice of a single amino acid as the sole building block at a large randomization site more difficult. This problem also applies to strategy 2 when using a single amino acid in addition to WT amino acid (Scheme 4.13).

4.3.5

Triple Code Saturation Mutagenesis (TCSM): A Viable Compromise when Choosing the Optimal Reduced Amino Acid Alphabet

An alternative way to explore strategy 1 more systematically is to test amino acid alphabets slightly larger than one at extended randomization sites (e.g.,

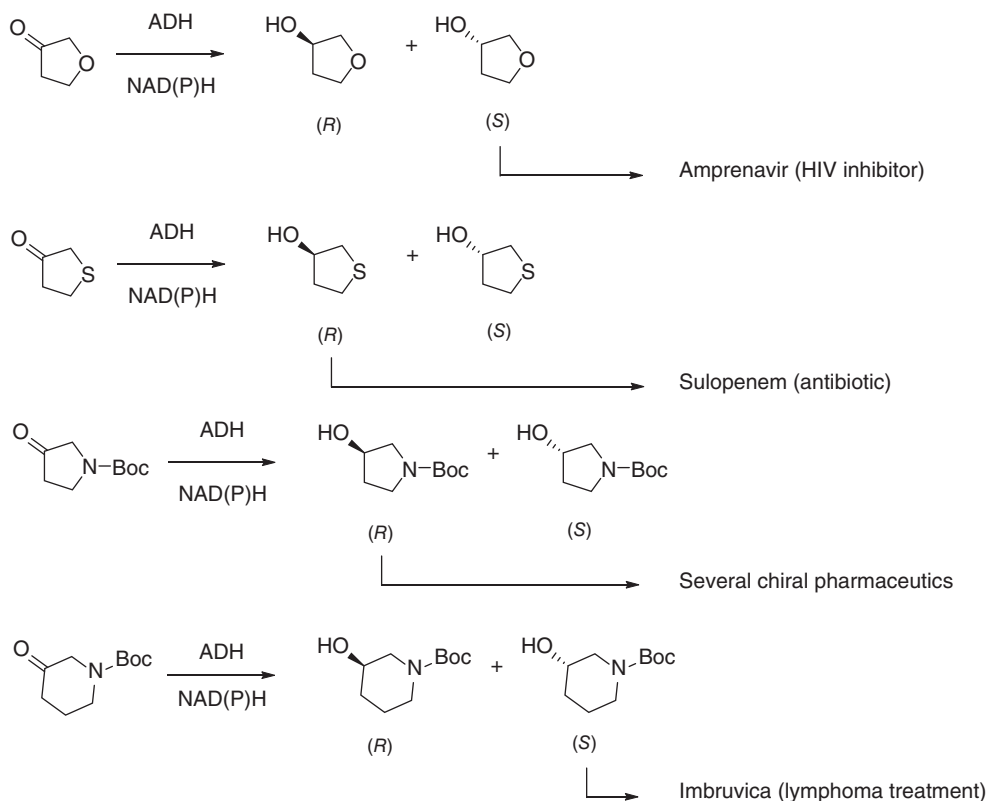
5–10-residues), keeping the statistical aspects of oversampling in mind. One option is to use two amino acids as building blocks instead of just one (in addition to WT), which can be called “double codon saturation mutagenesis” (DCSM). The total structural diversity can be increased further by using three amino acids as building blocks in a process that has been dubbed “triple code saturation mutagenesis” (TCSM) [56b]. Three options are then possible while still keeping screening to a minimum (2000–3000 transformants):

- Screen less than 95% library coverage.
- Split the multi-residue randomization site into two smaller sites A and B, and then apply ISM along pathway $A \rightarrow B$ or $B \rightarrow A$ (or both if one of them is not as unproductive as desired).
- Split the multi-residue randomization site into three smaller ones A, B, and C, and proceed with ISM exploration.

When choosing strategy 1 according to Scheme 4.13, one and the same three-membered reduced amino acid alphabet can be employed at all (smaller) sites derived from splitting a large multi-residue site, or a different triple code can be chosen at each of the split randomization sites. Initial projects using TCSM proved to be surprisingly successful, especially when guided by X-ray structural data, consensus sequence alignment, and computational aids [56b,c]. It is a practical compromise between the degree of structural diversity and extent of screening. In one study focusing on LEH as the catalyst in the model hydrolytic desymmetrization of cyclohexene oxide (**25b**) (Scheme 4.18), a triple code comprising Val-Phe-Tyr was chosen on structural and mechanistic grounds, which led to high (*S,S*)-selectivity (99% ee) and (*R,R*)-selectivity (89% ee) without resorting to ISM; (*R,R*)-selectivity was boosted to 97% by a single ISM step [56b]. Therefore, when comparing SCSM with TCSM in the same model reaction, the latter using three amino acids as building blocks is clearly superior to the former based on a single amino acid.

TCSM was also applied to the thermally robust ADH from *Thermoethanolicus brockii* (TbSADH) as catalyst in the asymmetric reduction of difficult-to-reduce ketones such as tetrahydrofuran-3-one [56c]. WT TbSADH shows low (*R*)-selectivity (23% ee). Docking computations revealed five residues to be in contact with the substrate: A85, I86, W110, L294, and C295 [56c]. In order to make a rational choice regarding the triple code, NNK-based saturation mutagenesis at all five positions was first performed, the observed amino acid substitutions that ensure higher (*R*)- or (*S*)-selectivity then being used as building blocks in subsequent TCSM. Based on the data from the NNK-experiments and X-ray structural information, two randomization sites were designed: A (A85/I86/L294/C295) and B (A85/I86/W110/L294). At site A, triple code Val-Asn-Leu was chosen, and based on all of the available data, mainly (*R*)-selective variants were expected. At site B, triple code Val-Gln-Leu was suggested to be optimal, leading to the expectation that mainly (*S*)-selective variants should appear in the subsequent saturation mutagenesis library. Indeed, the respective TCSM libraries harbored several highly (*R*)-selective variants (97–99% ee) and (*S*)-selective counterparts

(94–95% ee), respectively [56c]. The best variants were used as catalysts in the asymmetric reduction of other challenging substrates, their products being synthons for the preparation of a number of important therapeutic drugs (Scheme 4.20). Finally, TCSM was also successful in the P450-BM3 catalyzed regio- and enantioselective oxidative hydroxylation of cyclohexanone. Therefore, based on three initial studies, structure-based TCSM is emerging as the method of choice when evolving stereo- and regioselectivity, substrate scope and/or activity.



Scheme 4.20 Application of best variants of alcohol dehydrogenase TbSADH as catalysts in the asymmetric reduction of difficult-to-reduce ketones, evolved by application of triple code saturation mutagenesis (TCSM) [56c]. (Sun *et al.* [56a]. Reproduced with permission of John Wiley & Sons.)

In principle, DCSM or TCSM can also be applied when opting for strategy 2 (Scheme 4.13) [34b]. This means that decisions regarding the choice of a defined degenerate codon need to be made at each randomization site.

When choosing any one of the various approaches to saturation mutagenesis, such factors as structural data, bioinformatics information, and computational aids in designing mutant libraries as well as quality control of libraries need to

be considered. For this purpose, a flow sheet has been published that constitutes a user-friendly guide when applying saturation mutagenesis or its extension to ISM (Scheme 4.21) [59].

4.4

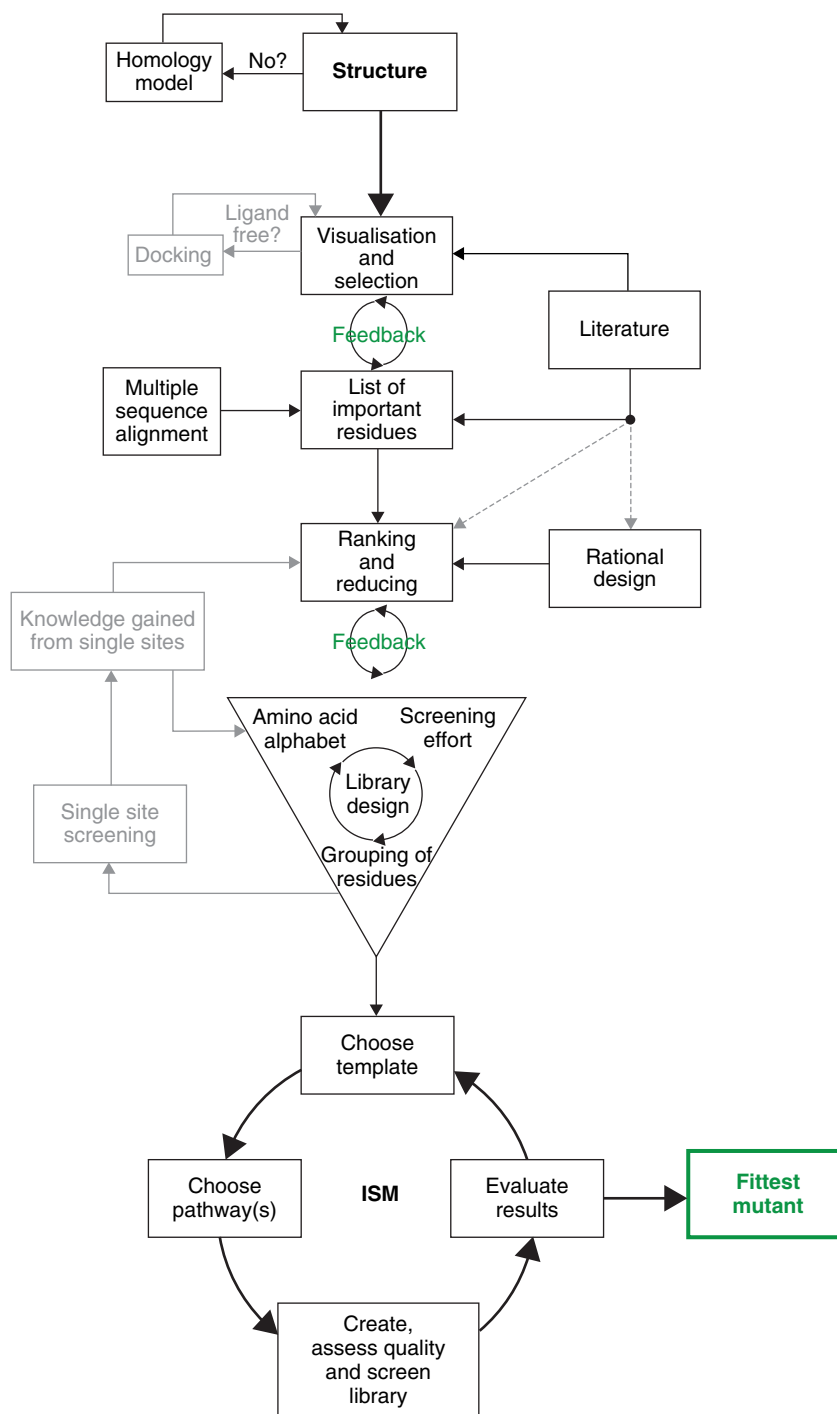
Techno-Economical Analyses of Saturation Mutagenesis Strategies

The above discussions focus on the use of saturation mutagenesis in protein engineering of enzymes in the quest to minimize laboratory work, especially the amount of screening (bottleneck of directed evolution). An issue that has been neglected in directed evolution in general is the question of the number and respective cost of primers needed when implementing a given mutagenesis strategy. Therefore, it is advisable to perform an economical analysis before initiating a directed evolution project along the following lines [60]:

- Apply a library quality control that is cheaper and faster than in the past.
- Consider the role of primer purity and costs according to supplier information with and without redundancy.
- Compare library quality, yield, randomization efficiency, and annealing bias using traditional and emergent randomization techniques based on mixtures of mutagenic primers.
- Choose the most cost-effective saturation mutagenesis scheme while also considering the screening costs and other experimental work, in a “compromise” procedure.

Only a few of the major highlights of this study are delineated here. The experimental results from which the techno-economical analysis [60] was performed refer to saturation mutagenesis of P450-BM3 as the catalyst in the regio- and stereoselective oxidative hydroxylation of steroids [59]. A total of 12 libraries were generated by different versions of saturation mutagenesis: (i) traditional use of NNK codon degeneracy; (ii) traditional use of NNS codon degeneracy; (iii) 22c-trick [61]; and (iv) Tang-approach [62]. In the comparative study, these different embodiments of saturation mutagenesis aim for the same goal regarding catalyst improvement, but require different primers and, therefore, lead to different costs. Moreover, primer quality and prices vary according to supplier. In all cases, the QuikChange protocol (Section 3.3) was used, but by employing a polymerase different from the traditional recipe [60]. The results of the QQC [3] and the respective quantitative Q -values [38] were obtained, revealing to what extent the desired diversity was actually introduced in the codon using a pooled DNA sequence electropherogram [59]. Table 4.4 summarizes the results of this type of analysis as applied to the given case. The sequence results are summarized in Table 4.5.

Upon comparing the six NNK libraries, it became clear that those using primers from supplier 3 lead to the highest mean Q_{pool} value relative to the results obtained when using primers from suppliers 1 and 2. In the techno-economical study, only primers from supplier 3 were used thereafter, specifically when



Scheme 4.21 Flow diagram of structure-based directed evolution via ISM [59]. (Acevedo-Rocha *et al.* [59]. Reproduced with permission of Springer.)

Table 4.4 Quick quality control and Q-values.

Library	Randomization scheme	Primer		QQC charts and Q-values			
		Supplier	Purity	Theoretical Q-value	Solid Q _{pool}	Liquid Q _{pool}	Experimental Q _{pool}
1	NNK	IDT Technologies	Desalted	1.000	0.470	0.515	0.798
2	NNK	IDT Technologies	HPLC	1.000	0.721	0.706	0.837
3	NNK	Life Technologies	Desalted	1.000	0.484	0.438	0.800
4	NNK	Life Technologies	HPLC	1.000	0.486	0.620	0.773
5	NNK	Metabion International	Desalted	1.000	0.770	0.657	0.863
6	NNK	Metabion International	HPLC	1.000	0.766	0.759	0.888
7	NNS	Metabion International	Desalted	1.000	0.643	0.549	0.872
8	NNS	Metabion International	HPLC	1.000	0.614	0.592	0.844
9	22c-trick	Metabion International	Desalted	1.000	0.557	0.575	0.979
10	22c-trick	Metabion International	HPLC	1.000	0.590	0.579	0.798
11	Tang	Metabion International	Desalted	1.000	0.668	0.543	0.834
12	Tang	Metabion International	HPLC	1.000	0.581	0.516	0.742

Library specifications with resulting QQC charts and Q-values [60]. The three pie charts in each column/row correspond to the three positions in a codon. Black, guanidine; green, adenosine; red, threonine; and blue, cytosine.

Source: Acevedo-Rocha, <http://www.nature.com/articles/srep10654>. Used under CC-BY 4.0 <http://creativecommons.org/licenses/by/4.0/>.

comparing the traditional codon degeneracy, Tang *et al.* [62] and 22c-trick [61] approaches [60]. Figure 4.15 reflects the total expenditure as a function of screening cost when randomizing a single position using five different saturation mutagenesis schemes [60].

The data in Figure 4.15 is based on the assumption that the primer cost equals the fixed cost. However, the economically optimal randomization scheme is

Table 4.5 Summary of P450-BM3 sequencing results obtained from 96 single colonies formed on agar plates per library [60].

Library	Successfully randomized	Yield (%)	> 1 base per position	Non-target mutations	Primer misinsertions	Suboptimal sequencing	Missed amino acids
1	68	72.3	19	6	1	2	Met, Asp
2	65	68.4	24	6	—	1	Lys, Asn, His
3	55	59.8	29	5	3	4	Met, Lys, Asn, Phe
4	54	58.1	37	2	—	3	Ile
5	50	52.1	42	4	—	—	—
6	64	66.7	27	5	—	—	—
7	39	41.9	50	4	—	3	Met, Ile, Gln, Trp
8	64	67.4	24	6	1	1	Phe, Tyr
9	57	59.4	37	1	1	—	—
10	67	69.8	24	4	1	—	Asp, Tyr
11	51	53.7	41	1	2	1	Asp, Tyr
12	65	69.1	23	6	—	2	Lys, Asp, Tyr

Source: Acevedo-Rocha, <http://www.nature.com/articles/srep10654>. Used under CC-BY 4.0 <http://creativecommons.org/licenses/by/4.0/>.

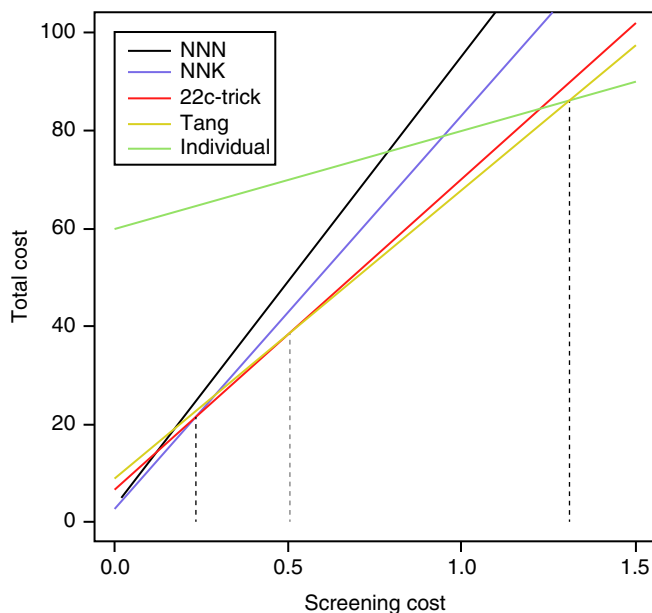


Figure 4.15 Total cost as a function of screening cost, when randomizing a single position using five randomization schemes. Primer cost is $c_{\text{primer}} = 1$ [60].

(Acevedo-Rocha, <http://www.nature.com/articles/srep10654>. Used under CC-BY 4.0 <http://creativecommons.org/licenses/by/4.0/>.)

influenced by changes in both primer cost and screening cost, relative to the fixed cost of the transformation of interest.

Figure 4.16 outlines how cost space, shown in the two-dimensional plane with axes as primer and screening costs, become partitioned into mutually exclusive regions. Thus, each of them corresponds to a different optimal choice of randomization procedure. It turns out, for example, that NNK is economically the first choice when the screening cost is lower than the primer cost by a factor of at least 4.25, as given by the slope of the line separating the NNK and the 22c-trick areas. When the screening cost increases, then the 22c-trick becomes optimal followed by the Tang approach and finally by the individual generation of all 20 variants. Generally, the 22c-trick and Tang approaches are fairly similar, sometimes the latter being better under certain conditions. In further analytical refinements, schemes such as the one depicted in Figure 4.16 were modified [60].

These and other analyses, for example, of annealing probabilities, culminated in several general conclusions and guidelines. When the screening costs are high, then approaches such as the 22c-trick [61] or the Tang-method [62] are economically superior. A faster and more economical method for performing the QQC is to use liquid cultures, also presented in this study [60]. It is highly recommended in order to save time and expenses when designing directed evolution experiments using saturation mutagenesis [59, 60]. Of significant practical importance is the

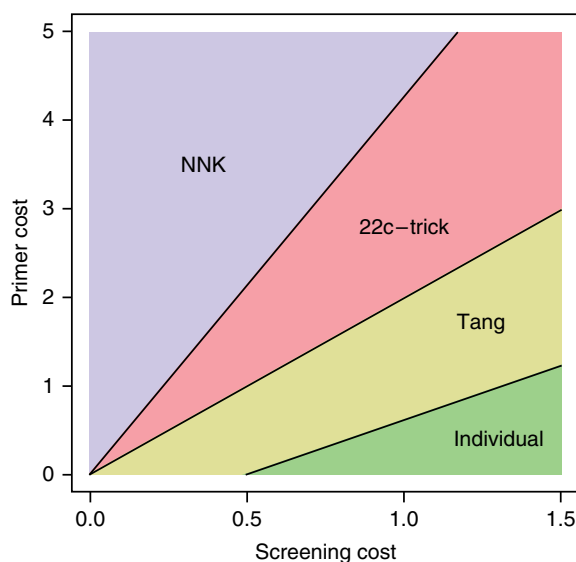


Figure 4.16 Cost space partitioned into regions according to the optimal randomization scheme (a single randomized position, assuming 100% yield, and no WT bias) [60].

(Acevedo-Rocha, <http://www.nature.com/articles/srep10654>. Used under CC-BY 4.0 <http://creativecommons.org/licenses/by/4.0/>.)

conclusion that QQC should be combined with the *Q*-values in order to assess the overall quality of libraries.

It was also shown that primer purity has a notable effect on library yield [60]. Some suppliers offer primers of higher quality than others without additional purification steps. Researchers are advised to read the details of this study, which provides guidelines for choosing a randomization scheme optimally as a function of the screening costs and other experimental parameters [60]. These guidelines are also of help when applying any PCR-based method for library creation such as epPCR or DNA shuffling, including combinatorial gene preparation [63], gene assembly [64], and overlap extension PCR [65]. Finally, yet another cost aspect that needs to be considered concerns a different approach to mutant library construction, namely combinatorial solid-phase gene synthesis (Section 4.5).

4.5

Combinatorial Solid-Phase Gene Synthesis: An Alternative for the Future?

All of the methods of mutant library generation utilizing epPCR, saturation mutagenesis, or DNA shuffling are PCR-based. A principally different approach is based on combinatorial solid-phase gene synthesis, which can be used to create mutant libraries at the DNA level [66] (see also Section 3.6). In this study, it has been demonstrated that the quality of such synthetic gene libraries is higher than

that of the respective PCR-based counterparts. P450-BM3 served as the model enzyme, catalyzing the regio- and stereoselective hydroxylation of steroids. If the costs of gene syntheses continue to go down significantly in the coming years, then the traditional PCR-based methods for library generation may be replaced by this emerging approach to smart library creation [66]. Nowadays, researchers, especially from industrial laboratories, can design their saturation mutagenesis libraries, send the genetic information to an appropriate company that offers combinatorial gene synthesis, and receive the respective libraries a few weeks later.

Gene synthesis on micro-chips appears to be even faster and ultimately cheaper [67]. When applied to mutant library creation for the purpose of directed evolution analogous to the use of the Sloning technique [66], new perspectives emerge. If prices continue to go down, these approaches to library construction may revolutionize future directed evolution.

References

1. (a) Bommarius, A.S. (2015) Biocatalysis: a status report. *Annu. Rev. Chem. Biomol. Eng.*, **6**, 319–345; (b) 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; (c) Reetz, M.T. (2012) Directed evolution of enzymes, in *Enzyme Catalysis in Organic Synthesis*, 3rd edn (eds K. Drauz, H. Gröger, and O. May), Wiley-VCH, Weinheim, pp. 119–190; (d) Siloto, R.M.P. and Weselake, R.J. (2012) Site saturation mutagenesis: methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.*, **1**, 181–189; (e) Lutz, S. and Bornscheuer, U.T. (eds) (2009) *Protein Engineering Handbook*, Wiley-VCH Verlag GmbH, Weinheim.
2. Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006) Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 5869–5874.
3. Bougioukou, D.J., Kille, S., Taglieber, A., and Reetz, M.T. (2009) Directed evolution of an enantioselective enoate-reductase: testing the utility of iterative saturation mutagenesis. *Adv. Synth. Catal.*, **351** (18), 3287–3305.
4. (a) Peisajovich, S.G. and Tawfik, D.S. (2007) Protein engineers turned evolutionists. *Nat. Methods*, **4** (12), 991–994; (b) DePristo, M.A. (2007) The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J.*, **1** (2), 94–98; (c) Rockah-Shmuel, L., Toth-Petroczy, A., and Tawfik, D.S. (2015) Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. *PLoS Comput. Biol.*, **11** (8), e1004421; (d) Bloom, J.D. and Arnold, F.H. (2009) In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **106** (Suppl. 1), 9995–10000.
5. Eigen, M., McCaskill, J., and Schuster, P. (1988) Molecular quasi-species. *J. Phys. Chem.*, **92** (24), 6881–6891.
6. Kurtovic, S. and Mannervik, B. (2009) Identification of emerging quasi-species in directed enzyme evolution. *Biochemistry*, **48** (40), 9330–9339.
7. Tracewell, C.A. and Arnold, F.H. (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.*, **13**, 3–9.
8. (a) Reetz, M.T., Wang, L.-W., and Bocla, M. (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. Ed.*, **45** (8), 1236–1241; (b) Initial study reporting four cycles of ISM for enhancing the stereoselectivity of the epoxide

- hydrolase ANEH:Reetz, M.T. (2005) Evolution im Reagenzglas: Neue Perspektiven für die Weiße Biotechnologie, in *Tätigkeitsberichte der Max-Planck-Gesellschaft*, pp. 327–331.
9. 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.
 10. (a) Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution. *ChemBioChem*, **9** (11), 1797–1804; (b) Clouthier, C.M., Kayser, M.M., and Reetz, M.T. (2006) Designing new Baeyer–Villiger monooxygenases using restricted casting. *J. Org. Chem.*, **71** (22), 8431–8437.
 11. Reetz, M.T. and Sanchis, J. (2008) Constructing and analyzing the fitness landscape of an experimental evolutionary process. *ChemBioChem*, **9** (14), 2260–2267.
 12. Reetz, M.T., Kahakeaw, D., and Sanchis, J. (2009) Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. *Mol. Biosyst.*, **5** (2), 115–122.
 13. Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y., and Bocola, M. (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.*, **132** (26), 9144–9152.
 14. Reetz, M.T. (2013) The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.*, **52** (10), 2658–2666.
 15. Chen, Z. and Zhao, H. (2005) Rapid creation of a novel protein function by in vitro coevolution. *J. Mol. Biol.*, **348** (5), 1273–1282.
 16. Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., and Hughes, G.J. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, **329** (5989), 305–309.
 17. Zhang, J.-H., Dawes, G., and Stemmer, W.P.C. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. U.S.A.*, **94** (9), 4504–4509.
 18. Parikh, M.R. and Matsumura, I. (2005) Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of β -fucosidase from β -galactosidase. *J. Mol. Biol.*, **352**, 621–628.
 19. Juers, D.H., Heightman, T.D., Vasella, A., McCarter, J.D., Mackenzie, L., Withers, S.G., and Matthews, B.W. (2001) A structural view of the action of *Escherichia coli* (*lacZ*) β -galactosidase. *Biochemistry*, **40** (49), 14781–14794.
 20. Reetz, M.T., Wilensek, S., Zha, D., and Jaeger, K.-E. (2001) Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angew. Chem. Int. Ed.*, **40** (19), 3589–3591.
 21. (a) Nov, Y. (2012) When second best is good enough: another probabilistic look at saturation mutagenesis. *Appl. Environ. Microbiol.*, **78** (1), 258–262; (b) Nov, Y. (2013) Fitness loss and library size determination in saturation mutagenesis. *PLoS One*, **8** (7), e68069; (c) Nov, Y. (2014) Probabilistic methods I directed evolution: library size, mutation rate, and diversity. *Methods Mol. Biol.*, **1179**, 261–278; (d) Nov, Y., Fulton, A., and Jaeger, K.-E. (2013) Optimal scanning of all single-point mutants of a protein. *J. Comput. Biol.*, **20**, 990–997.
 22. Reetz, M.T., Zonta, A., Schimossek, K., Jaeger, K.-E., and Liebeton, K. (1997) Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. *Angew. Chem. Int. Ed. Engl.*, **36** (24), 2830–2832.
 23. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., and Jaeger, K.-E. (2000) Directed evolution of an enantioselective lipase. *Chem. Biol.*, **7** (9), 709–718.
 24. 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.
 25. Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D., and Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site

- saturation test. *Angew. Chem. Int. Ed.*, **44** (27), 4192–4196.
26. (a) Schimossek, K. (1998) Dissertation, *Neue Lipasen für die organische Chemie*, Ruhr-Universität Bochum; (b) Wilensek, S. (2001) Dissertation, *Gerichtete Evolution als ein Mittel zur Erzeugung enantioselektiver Enzyme für die organische Synthese*, Ruhr-Universität Bochum.
 27. Nardini, M., Lang, D.A., Liebeton, K., Jaeger, K.-E., and Dijkstra, B.W. (2000) Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation: the prototype for family I.1 of bacterial lipases. *J. Biol. Chem.*, **275** (40), 31219–31225.
 28. Crameri, A. and Stemmer, W.P.C. (1995) Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wild-type sequences. *BioTechniques*, **18**, 194–196.
 29. (a) Bocola, M., Otte, N., Jaeger, K.-E., Reetz, M.T., and Thiel, W. (2004) Learning from directed evolution: theoretical investigations into cooperative mutations in lipase enantioselectivity. *ChemBioChem*, **5** (2), 214–223; (b) Reetz, M.T., Puls, M., Carballeira, J.D., Vogel, A., Jaeger, K.-E., Eggert, T., Thiel, W., Bocola, M., and Otte, N. (2007) Learning from directed evolution: further lessons from theoretical investigations into cooperative mutations in lipase enantioselectivity. *ChemBioChem*, **8** (1), 106–112.
 30. Zha, D., Wilensek, S., Hermes, M., Jaeger, K.-E., and Reetz, M.T. (2001) Complete reversal of enantioselectivity of an enzyme-catalyzed reaction by directed evolution. *Chem. Commun.*, (24), 2664–2665.
 31. Gumulya, Y., Sanchis, J., and Reetz, M.T. (2012) Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *ChemBioChem*, **13** (7), 1060–1066.
 32. (a) Review of directed evolution of stereoselective enzymes with emphasis on ISM: 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) Perspective on biocatalysts including the development of directed evolution: Reetz, M.T. (2013) Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.*, **135** (34), 12480–12496.
 33. (a) Drauz, K., Gröger, H., and May, O. (eds) (2012) *Enzyme Catalysis in Organic Synthesis*, 3rd edn, Wiley-VCH Verlag GmbH, Weinheim; (b) Faber, K. (2011) *Biotransformations in Organic Chemistry*, 6th edn, Springer, Heidelberg; (c) Liese, A., Seelbach, K., and Wandrey, C. (eds) (2006) Weinheim, in *Industrial Biotransformations*, Wiley-VCH Verlag GmbH; (d) Gotor, V., Alfonso, I., and Garcia-Urdiales, E. (eds) (2008) *Asymmetric Organic Synthesis with Enzymes*, Wiley-VCH Verlag GmbH, Weinheim; (e) Tao, J., Lin, G.-Q., and Liese, A. (2009) *Biocatalysis for the Pharmaceutical Industry*, John Wiley & Sons Asia (Pte) Ltd, Singapore.
 34. (a) Reetz, M.T. and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chem. Commun.*, (43), 5499–5501; (b) Sandström, A.G., Wikmark, Y., Engström, K., Nyhlén, J., and Bäckvall, J.-E. (2012) Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (1), 78–83; (c) Wikmark, Y., Svedendahl Humble, M., and Bäckvall, J.E. (2015) Combinatorial library based engineering of *Candida antarctica* lipase A for enantioselective transacylation of secondary alcohols in organic solvent. *Angew. Chem. Int. Ed.*, **54**, 4284–4288; (d) Chen, M.M.Y., Snow, C.D., Vizarra, C.L., Mayo, S.L., and Arnold, F.H. (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng. Des. Sel.*, **25**, 171–178; (e) Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011) Directed evolution of the nonribosomal peptide synthetase AdmK

- generates new andrimid derivatives in vivo. *Chem. Biol.*, **18** (5), 601–607;
- (f) Swe, P.M., Copp, J.N., Green, L.K., Guise, C.P., Mowday, A.M., Smaill, J.B., Patterson, A.V., and Ackerley, D.F. (2012) Targeted mutagenesis of the *Vibrio fischeri* flavin reductase FRase I to improve activation of the anticancer pro-drug CB1954. *Biochem. Pharmacol.*, **84** (6), 775–783; (g) Dudek, H., Fink, M., Shivange, A., Dennig, A., Mihovilovic, M., Schwaneberg, U., and Fraaije, M. (2014) Extending the substrate scope of a Baeyer–Villiger monooxygenase by multiple-site mutagenesis. *Appl. Microbiol.*, **98** (9), 4009–4020; (h) Zhang, L., Lu, L., Fan, S., Jin, L., Gu, G., Xu, L., and Xiao, M. (2015) One-step synthesis of α -Gal epitope and globotriose derivatives by an engineered α -galactosidase. *RSC Adv.*, **5** (29), 22361–22364; (i) Chuang, H.-Y., Suen, C.-S., Hwang, M.-J., and Roffler, S.R. (2015) Toward reducing immunogenicity of enzyme replacement therapy: altering the specificity of human β -glucuronidase to compensate for α -iduronidase deficiency. *Protein Eng. Des. Sel.*, **28** (11), 519–530.
35. (a) Truppo, M.D., Turner, N.J., and Rozzell, J.D. (2009) Efficient kinetic resolution of racemic amines using a transaminase in combination with an amino acid oxidase. *Chem. Commun.*, (16), 2127–2129; (b) Koszelewski, D., Clay, D., Rozzell, D., and Kroutil, W. (2009) Deracemisation of α -chiral primary amines by a one-pot, two-step cascade reaction catalysed by ω -transaminases. *Eur. J. Org. Chem.*, **2009** (14), 2289–2292.
36. Zou, J., Hallberg, B.M., Bergfors, T., Oesch, F., Arand, M., Mowbray, S.L., and Jones, T.A. (2000) Structure of *Aspergillus niger* epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure*, **8** (2), 111–122.
37. 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.
38. Sullivan, B., Walton, A.Z., and Stewart, J.D. (2013) Library construction and evaluation for site saturation mutagenesis. *Enzyme Microbiol. Technol.*, **53** (1), 70–77.
39. Emrén, L.O., Kurtovic, S., Runarsdottir, A., Larsson, A.-K., and Mannervik, B. (2006) Functionally diverging molecular quasi-species evolve by crossing two enzymes. *Proc. Natl. Acad. Sci. U.S.A.*, **103** (29), 10866–10870.
40. Gumulya, Y. and Reetz, M.T. (2011) Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *ChemBioChem*, **12** (16), 2502–2510.
41. Ihssen, J., Haas, J., Kowarik, M., Wiesli, L., Wacker, M., Schwede, T., and Thöny-Meyer, L. (2015) Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.*, **5** (4), 140227.
42. 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** (17), 5695–5705.
43. Goldstein, J., Sivilgia, G., Hurst, R., Lenny, L., and Reich, L. (1982) Group B erythrocytes enzymatically converted to group O survive normally in A, B, and O individuals. *Science*, **215** (4529), 168–170.
44. Higgins, M.A., Whitworth, G.E., El Warry, N., Randrianisoa, M., Samain, E., Burke, R.D., Vocadlo, D.J., and Boraston, A.B. (2009) Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. *J. Biol. Chem.*, **284** (38), 26161–26173.
45. Sun, Z., Ilie, A., and Reetz, M.T. (2015) Towards the production of universal blood by structure-guided directed evolution of glycoside hydrolases. *Angew. Chem. Int. Ed.*, **54** (32), 9158–9160.

46. (a) Steipe, B., Schiller, B., Plückthun, A., and Steinbacher, S. (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.*, **240** (3), 188–192; (b) Steipe, B. (2004) Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. *Methods Enzymol.*, **388**, 176–186; (c) Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng. Des. Sel.*, **15** (5), 403–411; (d) Polizzi, K.M., Chaparro-Riggers, J.F., Vazquez-Figueroa, E., and Bommarius, A.S. (2006) Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Biotechnol. J.*, **1** (5), 531–536; (e) Amin, N., Liu, A.D., Ramer, S., Aehle, W., Meijer, D., Metin, M., Wong, S., Gualfetti, P., and Schellenberger, V. (2004) Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.*, **17**, 787–793.
47. (a) Fraaije, M., Wu, J., Heuts, D.H.M., van Hellemond, E., Spelberg, J.L., and Janssen, D. (2005) Discovery of a thermostable Baeyer–Villiger monooxygenase by genome mining. *Appl. Microbiol. Biotechnol.*, **66** (4), 393–400; (b) de Gonzalo, G., Pazmiño, D.E.T., Ottolina, G., Fraaije, M.W., and Carrea, G. (2005) Oxidations catalyzed by phenylacetone monooxygenase from *Thermobifida fusca*. *Tetrahedron: Asymmetry*, **16** (18), 3077–3083.
48. Malito, E., Alfieri, A., Fraaije, M.W., and Mattevi, A. (2004) Crystal structure of a Baeyer–Villiger monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.*, **101** (36), 13157–13162.
49. Bocola, M., Schulz, F., Leca, F., Vogel, A., Fraaije, M.W., and Reetz, M.T. (2005) Converting phenylacetone monooxygenase into phenylcyclohexanone monooxygenase by rational design: towards practical Baeyer–Villiger monooxygenases. *Adv. Synth. Catal.*, **347** (7–8), 979–986.
50. (a) Jochens, H. and Bornscheuer, U.T. (2010) Natural diversity to guide focused directed evolution. *ChemBioChem*, **11** (13), 1861–1866; (b) Höhne, M., Schätzle, S., Jochens, H., Robins, K., and Bornscheuer, U.T. (2010) Rational assignment of key motifs for function guides in silico enzyme identification. *Nat. Chem. Biol.*, **6** (11), 807–813; (c) Godinho, L.F., Reis, C.R., Rozeboom, H.J., Dekker, F.J., Dijkstra, B.W., Poelarends, G.J., and Quax, W.J. (2012) Enhancement of the enantioselectivity of carboxylesterase A by structure-based mutagenesis. *J. Biotechnol.*, **158** (1–2), 36–43.
51. Reetz, M.T. and Wu, S. (2009) Laboratory evolution of robust and enantioselective Baeyer–Villiger monooxygenases for asymmetric catalysis. *J. Am. Chem. Soc.*, **131** (42), 15424–15432.
52. (a) Fersht, A. (2000) *Structure and Mechanism in Protein Science*, W.H. Freeman and Co., New York; (b) Shen, B., Bai, J., and Vihinen, M. (2008) Physicochemical feature-based classification of amino acid mutations. *Protein Eng. Des. Sel.*, **21**, 37–44.
53. Engström, K., Nyhlén, J., Sandström, A.G., and Bäckvall, J.-E. (2010) Directed evolution of an enantioselective lipase with broad substrate scope for hydrolysis of α -substituted esters. *J. Am. Chem. Soc.*, **132** (20), 7038–7042.
54. Zheng, H. and Reetz, M.T. (2010) Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis. *J. Am. Chem. Soc.*, **132** (44), 15744–15751.
55. (a) Hopmann, K.H., Hallberg, B.M., and Himo, F. (2005) Catalytic mechanism of limonene epoxide hydrolase, a theoretical study. *J. Am. Chem. Soc.*, **127** (41), 14339–14347; (b) Lind, M.E.S. and Himo, F. (2013) Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angew. Chem. Int. Ed.*, **52**, 4563–4567; (c) Arand, M., Hallberg, B., Zou, J., Bergfors, T., Oesch, F., van der Werf, M.J., de Bont, J.A.M., Jones, T., and Mowbray, S.L. (2003) Structure of

- Rhodococcus erythropolis* limonene-1,2-epoxide hydrolase reveals a novel active site. *EMBO J.*, **22** (11), 2583–2592.
56. (a) Sun, Z., Lonsdale, R., Kong, X.-D., Xu, J.-H., Zhou, J., and Reetz, M.T. (2015) Reshaping an enzyme binding pocket for enhanced and inverted stereoselectivity: use of smallest amino acid alphabets in directed evolution. *Angew. Chem. Int. Ed.*, **54**, 12410–12415; (b) Sun, Z., Lonsdale, R., Wu, L., Li, G., Li, A., Wang, J., Zhou, J., and Reetz, M.T. (2016) Structure-guided triple code saturation mutagenesis: efficient tuning of the stereoselectivity of an epoxide hydrolase. *ACS Catal.*, **6**, 1590–1597; (c) Sun, Z., Lonsdale, R., Ilie, A., Li, G., Zhou, J., and Reetz, M.T. (2016) Catalytic asymmetric reduction of difficult-to-reduce ketones: triple code saturation mutagenesis of an alcohol dehydrogenase. *ACS Catal.*, **6**, 1598–1605; (d) Li, G., Zhang, H., Sun, Z., Liu, X., and Reetz, M.T. (2016) Multiparameter Optimization in Directed Evolution: Engineering Thermostability, Enantioselectivity, and Activity of an Epoxide Hydrolase. *ACS Catal.*, **6**, 3679–3687.
 57. (a) Morrison, K.L. and Weiss, G.A. (2001) Combinatorial alanine-scanning. *Curr. Opin. Chem. Biol.*, **5** (3), 302–307; (b) Massova, I. and Kollman, P.A. (1999) Computational alanine scanning to probe protein–protein interactions: A novel approach to evaluate binding free energies. *J. Am. Chem. Soc.*, **121** (36), 8133–8143; (c) Lewis, J.C., Mantovani, S.M., Fu, Y., Snow, C.D., Komor, R.S., Wong, C.-H., and Arnold, F.H. (2010) Combinatorial alanine substitution enables rapid optimization of cytochrome P450BM3 for selective hydroxylation of large substrates. *ChemBioChem*, **11** (18), 2502–2505.
 58. Wahler, D. and Reymond, J.-L. (2002) The adrenaline test for enzymes. *Angew. Chem. Int. Ed.*, **41** (7), 1229–1232.
 59. Acevedo-Rocha, C., Hoebenreich, S., and Reetz, M.T. (2014) Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol. Biol.*, **1179**, 103–128.
 60. Acevedo-Rocha, C.G., Reetz, M.T., and Nov, Y. (2015) Economical analysis of saturation mutagenesis experiments. *Sci. Rep.*, **5**, 10654.
 61. Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.-G., Opperman, D.J., Reetz, M.T., and Acevedo, J.P. (2013) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.*, **2** (2), 83–92.
 62. Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M., and Jiang, R. (2012) Construction of “small-intelligent” focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques*, **52** (3), 149–158.
 63. Currin, A., Swainston, N., Day, P.J., and Kell, D.B. (2014) Speedygenes: an improved gene synthesis method for the efficient production of error-corrected, synthetic protein libraries for directed evolution. *Protein Eng. Des. Sel.*, **27** (9), 273–280.
 64. Acevedo-Rocha, C.G. and Reetz, M.T. (2014) Assembly of designed oligonucleotides: a useful tool in synthetic biology for creating high quality combinatorial DNA Libraries. *Methods Mol. Biol.*, **1179**, 189–206.
 65. Williams, E.M., Copp, J.N., and Ackerley, D.F. (2014) Site-saturation mutagenesis by overlap extension PCR. *Methods Mol. Biol.*, **1179**, 83–101.
 66. Hoebenreich, S., Zilly, F.E., Acevedo-Rocha, C.G., Zilly, M., and Reetz, M.T. (2015) Speeding up directed evolution: combining the advantages of solid-phase combinatorial gene synthesis with statistically guided reduction of screening effort. *ACS Synth. Biol.*, **4** (3), 317–331.
 67. (a) Schmidt, T.L., Beliveau, B.J., Uca, Y.O., Theimann, M., Da Cruz, F., Wu, C.-T., and Shi, W.M. (2015) Scalable amplification of strand subsets from chip-synthesized oligonucleotide libraries. *Nat. Commun.*, **6**, 8634; (b) For a company specializing in micro-chip combinatorial gene synthesis, see: LabGeni www.labgeni.us (accessed 10 May 2016).

5

Selected Examples of Directed Evolution of Enzymes with Emphasis on Stereo- and Regioselectivity, Substrate Scope, and/or Activity

5.1

Explanatory Remarks

As noted in the previous chapters, the often observed insufficient activity, narrow substrate scope, and/or poor or wrong stereo- and regioselectivity of enzymes constitute the major reasons why biocatalysts have not been employed more often in organic chemistry and biotechnology. Directed evolution has addressed all of these issues, which are the subject of this chapter. The first study reporting the directed evolution of a stereoselective enzyme involved a lipase (Chapter 1) [1]. Since then most of the major enzyme types have been subjected to directed evolution in successful attempts to enhance and/or invert enantioselectivity: hydrolases (e.g., lipases, esterases, nitrilases, epoxide hydrolases, glycosidases), acylases (e.g., penicillin G), monooxygenases (e.g., P450 and Baeyer–Villiger enzymes, monoamine oxidases), reductases (e.g., alcohol dehydrogenases, enoate reductases), and ligases (e.g., aldolases, oxynitrilases, thiamine diphosphate-dependent decarboxylases), lyases (e.g., aryl malonate decarboxylases). Some of these advances have been summarized in general reviews of directed evolution [2] and in protein engineering accounts of specific enzyme types, for example, lipases and esterases [3], aldolases [4a,b], hydroxynitrile lyases [4c], P450 monooxygenases [5], glycosyltransferases [6a], ligninolytic enzymes [6b], laccases [6c], cyclodextrin glycosyltransferases [6d], cellulases [6e], nucleases (including zinc finger proteins) [7], Baeyer–Villiger monooxygenases [8], and phosphotriesterases [9].

Table 5.1 features typical studies from the recent literature, arranged according to chemical transformations and enzyme type. The interested researcher intending to perform directed evolution of a given enzyme can readily find selected studies focusing on the respective enzyme type, thereby enabling fast comparison of the different approaches. All major gene mutagenesis methods are considered, including error-prone polymerase chain reaction (epPCR), saturation mutagenesis (SM), iterative saturation mutagenesis (ISM), and DNA shuffling as well as other approaches such as circular permutation. Although most of these methods continue to be applied with the aim of influencing activity, substrate scope, stereo-, and regioselectivity, SM at sites lining the binding pocket has emerged as the most

reliable approach, often in a recursive manner (ISM) (Chapters 3 and 4). Indeed, a recent literature search by the author has revealed that SM-based gene mutagenesis methods are used most often when manipulating these catalytic parameters, generally guided by structural, mechanistic, computational, and/or sequence data.

Table 5.1 is meant to be representative, not comprehensive. Further examples are presented in Chapter 7, which treats the directed evolution of promiscuous enzymes. In the vast majority of studies based on SM, randomization was focused on sites lining or near the binding pocket (first and second sphere residues), which means that the Combinatorial Active-Site Saturation Test (CAST) [2e] was employed. In many cases, the authors of the cited papers in Table 5.1 used this convenient acronym to distinguish the process from SM at remote sites for other purposes; in other studies the term CAST was not. For this reason, CAST was not used as a descriptor in Table 5.1, although in essentially all listed cases it was actually involved. In some studies several mutagenesis methods were applied, in which cases this information is indicated in Table 5.1.

Traditional rational design leading to the introduction of a defined mutation at a predetermined position using site-specific mutagenesis is not featured to any great extent in this monograph. However, some recent studies employing site-specific mutagenesis report more than one amino acid substitution at a given position followed by screening of the small generated mutant library, a process that begins to mimic directed evolution. In some of these cases, the best mutant was used as a template for another round of site-specific mutagenesis at a different position at the binding pocket, which is reminiscent of ISM. For illustrative purposes, a few of these studies have been included in Table 5.1. The results of structure-based SM and ISM continue to provide data that is useful in designing new improved enzyme variants by site-specific mutagenesis, and vice versa. Thus, the two approaches to protein engineering are beginning to merge.

Table 5.1 Typical directed evolution studies of enzymes for enhanced stereo- and/or regioselectivity, activity, shifted substrate scope, selected from the literature 2010 up to 2016.

Enzyme	Evolved property	Mutagenesis method	Comment	References
Lipase (<i>Rhizomucor miehei</i>)	Activity	epPCR	Four epPCR cycles	[10]
Lipase (<i>Pseudomonas aeruginosa</i>)	Activity, enantioselectivity	ISM	Reduced amino acid alphabet	[11]
Lipase (CALA)	Enantioselectivity	ISM	Reduced amino acid alphabet	[12]
Lipase (CALA)	Enantioselectivity	SM	Different codon degeneracy at each position; immobilization on Ni-coated 96-well microtiter plates	[13]
Lipase (CALA)	Enantioselectivity	SM	Different codon degeneracy at each position	[14]
Lipase (CALA)	<i>Trans/cis</i> substrate selectivity	SM	Reduced amino acid alphabet	[15]
Lipase (CALB)	Enantioselectivity	ISM	Kinetic resolution, reduced amino acid alphabet, profen-esters	[16]
Lipase (<i>Candida rugosa</i>)	Activity, substrate-selectivity	SM	NNK codon degeneracy	[17]
Lipase (pancreatic)	Activity, substrate-selectivity	epPCR	One cycle only	[18]
Lipase (CALB)	Enantioselectivity	ISM	Reduced amino acid alphabet	[19]
Lipase (<i>Pseudomonas</i> sp. Lip I.3)	Substrate scope, activity	SM	Entrance channel and lid mutations	[20]
Lipase (CALA)	Acyl transfer activity	Site-specific mutagenesis; 28 variants generated	Out of 28 rationally designed variants a single hit	[21]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Lipase (<i>Penicillium expansum</i>)	Activity; stability	SM	B-FIT approach led to higher activity; lid hinge region randomized	[22]
Lipase (<i>Thermomyces lanuginosus</i>)	Stereoselectivity	epPCR, SM	Diastereoselective “kinetic resolution”	[23]
Lipase (<i>Bacillus pumilus</i>)	Activity; stability	Family shuffling	Palmitate as substrate	[24]
Lipase (<i>Bacillus subtilis</i> A)	Activity	SM	Loop randomization	[25]
Lipase	Stability, substrate scope	Domain recombination	Lipase/esterase/protease (sub)domains	[26]
Esterase (<i>Rhodobacter sphaeroides</i>)	Enantioselectivity	epPCR	Three cycles	[27]
Esterase (YbF from <i>Escherichia coli</i>)	Activity, enantioselectivity	SM	NDT codon; challenging substrate	[28]
Esterase (<i>B. subtilis</i>)	Enantioselectivity	SM, site-specific mutagenesis, ISM, epPCR	Challenging substrate	[29]
Esterase (<i>Pseudomonas fluorescens</i>)	Substrate scope, enantioselectivity	Site-specific mutagenesis, SM, combining mutations	<i>In silico</i> guidance	[30]
Esterase (<i>Pseudomonas fluorescens</i>)	Activity, enantioselectivity	SM	Reduced amino acid alphabets	[31]
Esterase (<i>Rhodococcus</i> sp. ECU1013)	Activity	epPCR, SM	Substrate channel evolution, cilastatin synthesis	[32]
PTE-like lactonase (<i>Deinococcus radiodurans</i>)	Substrate scope	epPCR, SteP shuffling, ISM	Promiscuous mutant accepts paraoxon	[33]

Phytase (<i>Yersinia mollaretii</i>)	Activity, stability	Random mutagenesis by SeSaM	epPCR not as effective	[34]
Phytase (<i>B. subtilis</i> 168)	Activity	epPCR, SM	Activity increase in acidic and neutral media	[35]
Phytase (<i>Aspergillus niger</i> N25)	Activity	epPCR	Close and remote mutations	[36]
Penicillin G acylase	Diastereoselectivity Substrate scope	SM	Mutants for both diastereomers evolved	[37]
Acylaminoacyl peptidase				[38]
(<i>Aeropyrum pernix</i> K1)	Substrate scope Activity	SM	Phylogenetic analysis Biosensor	[39]
Cephalosporin C acylase				[40]
L-Asparaginase (<i>Erwinia carotovora</i>)	Activity	Random mutagenesis (SeSaM), combining mutations	Thermostability also evolved	[41]
Protease (<i>Bacillus gibsonii</i>)				
Metallo- β -lactamase (MBL9)	Substrate scope	SM	Different codon at different sites; resistance to seven antibiotics	[42]
Sortase A transpeptidase (<i>Staphylococcus aureus</i>)	Activity	Random mutagenesis using triphosphate derivatives of nucleoside analogs		[43]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Nonribosomal peptide synthetase (<i>Pantoea agglomerans</i>)	Substrate scope	SM	Different codon at different positions	[44]
Carbonic anhydrase (<i>Desulfovibrio vulgaris</i>)	Activity, stability	ISM, DNA shuffling	CO ₂ sequestration in <i>N</i> -methyl-diethanolamine	[45]
Phosphotriesterase (<i>Pseudomonas diminuta</i>)	Esterase promiscuity	epPCR; DNA shuffling	18 rounds; 10 ⁴ clones	[46]
Phosphotriesterase (<i>Agrobacterium radiobacter</i>)	Substrate scope	SM	Malathion acceptance	[47]
Phosphotriesterase (<i>P. diminuta</i>)	Activity, enantioselectivity	epPCR, SM, ISM	Degradation of warfare agent VX	[48]
Phosphotriesterase (<i>Brevundimonas diminuta</i>)	Substrate scope, enantioselectivity	Site-specific mutagenesis, SM	Rosetta computationally designed focused libraries	[49]
Epoxide hydrolase (metagenomic)	Enantioselectivity; activity	ISM	Several substrates	[50]
Limonene epoxide hydrolase	Enantioselectivity	ISM	Smallest amino acid alphabets	[51]
Epoxide hydrolase (<i>A. niger</i>)	Enantioselectivity	ISM	All 24 ISM pathways explored	[52]
Epoxide hydrolase (<i>A. niger</i>)	Enantioselectivity and expression efficiency	ISM	Reduced amino acid alphabets	[53]
Limonene epoxide hydrolase	Enantioselectivity	ISM	Reduced amino acid alphabet	[54]

Epoxide hydrolase (<i>A. niger</i>)	Enantio- convergence	ISM	Reduced amino acid alphabet	[55]
Epoxide hydrolase (<i>Solanum tuberosum</i>)	Activity, enantioselectivity	ISM	Kinetic resolution	[56]
Epoxide hydrolase (<i>S. tuberosum</i>)	Stereoselectivity	ISM	Enantioconvergence	[57]
Epoxide hydrolase (<i>A. niger</i> M200)	Stereoselectivity	ISM	Enantioconvergence	[58]
Nitrilase (2A6 nitA)	Enantioselectivity	ISM	Dynamic kinetic resolution	[59a]
Nitrilase (<i>Burkholderia</i> <i>cenocepacia</i> J2315)	Activity, stereoselectivity	epPCR, SM	(<i>R</i>)- <i>o</i> -chloromandelic acid production	[60]
Nitrilase (<i>Alcaligenes</i> <i>faecalis</i>)	Activity, enantioselectivity	SM/ISM	Upscale (<i>R</i>)-mandelic acid production	[59b]
Arginine deiminase (<i>Lactococcus lactis</i>)	Activity	epPCR	Production of L-citrulline	[61]
Amylosucrase	Substrate scope in glycosylation reactions	ISM	Sucrose as donor	[62]
Glycosidase (β -fructofuranosidase/ <i>Schwanniomyces</i> <i>occidentalis</i>)	Transglycosylation activity, substrate selectivity	epPCR, DNA shuffling	6- Kestose synthesis	[63]
Glycosidase (<i>N</i> -oligosaccharyl transferase; <i>Campylobacter jejuni</i>)	Activity	ISM	Application in vaccine	[64]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
β -Glycosidase (<i>Thermus thermophilus</i>)	Activity	Site-specific mutagenesis, epPCR	Promiscuous reaction	[65]
Glycosidase (Arabinofuranosyl hydrolase; <i>Clostridium thermocellum</i>)	Substrate scope in transglycosylation	epPCR	Aliphatic alcohols as acceptors	[66]
Glycosidase (cyclodextrin glycosyltransferase; <i>Paenibacillus macerans</i>)	Substrate acceptance	ISM	2-O-D-glucopyranosyl-L-ascorbic acid production	[67]
β -Fructofuranosidase (<i>Aspergillus japonicus</i>)	Prevention of inhibition by glucose	epPCR	IR-detection assay	[68]
β -Glycosidase Zm-p601 (maize)	Activity	SM, site-specific mutagenesis	NNM, NNK codon degeneracies	[69]
β -Glycosidase (<i>Agrobacterium</i> sp.)	Substrate scope	SM	C3-modified donor sugars	[70]
α -Galactosidase (<i>Penicillium janczewskii</i> zalek)	Activity	epPCR, DNA shuffling	Expression rate also enhanced	[71]
<i>N</i> -oligosaccharyltransferase (<i>C. jejuni</i>)	Activity	SM, epPCR	Structural insights	[72]
α -L-Arabinofuranosidase (<i>Thermobacillus xylanilyticus</i>)	Chemoselectivity	epPCR	Transglycosidase activity and less hydrolysis	[73]
Glycosyltransferase (OleD)	Substrate scope	SM	Sugar nucleotide synthesis	[74]

Endo-1,4- β -xylanase (<i>Geobacillus stearothermophilus</i>)	Activity	SM, epPCR	9000 Colonies screened	[75]
Amylosucrase (<i>Neisseria polysaccharea</i>)	Activity	SM	Thermostability also improved	[76]
Human β -glucuronidase	Substrate scope	SM	Different codon at different positions; α -iduroniside activity; reduced immunogenicity	[77]
Polysialyltransferase (<i>Nmb</i>)	Size distribution of polymeric products	epPCR, SM	Potential medical applications	[78]
Exoglucanase (<i>Cellulomonas fimi</i>)	Activity	epPCR	Whole engineered cellulase system	[79]
Xylanase (<i>T. xylanilyticus</i>)	Activity	epPCR, SM, SteP shuffling	Biomass degradation (wheat straw)	[80]
α -2,6-Sialyltransferase (<i>Photobacterium damsela</i>)	Activity	SM, alanine scanning	Sialyllactose production	[81]
α -Cyclodextrin glycosyltransferase	α/β -selectivity	SM, epPCR	α -Dextrin formation from starch favored	[82]
Xylose isomerase (<i>Piromyces</i> sp.)	Activity	epPCR	Enhanced xylose catabolism in yeast	[83]
Tetralose phosphorylase (<i>Caldanaerobacter subterraneus</i>)	Chemo- and regioselectivity	ISM	β -Galactose-1-phosphate production	[84]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Tagatose epimerase (<i>Pseudomonas cichorii</i>)	Substrate scope	ISM	C3-epimerization of D-fructose and L-sorbose	[85]
Sugar invertase (TnINV; <i>Thermotoga neapolitana</i>)	Activity	ISM	Production of D-psicose from sucrose in a cascade reaction	[86]
P450-BM3	Substrate scope	SM	Combinatorial alanine scanning	[87]
P450-BM3	Substrate scope	SM	Comparative study of three saturation mutagenesis approaches	[88]
P450-BM3	Activity, regioselectivity	ISM	Bioorthogonal deprotection of caged compounds	[89]
P450-BM3	Activity, regioselectivity	ISM	Ortho-selective hydroxylation of aromatics	[90]
P450-BM3	Regio- and enantioselectivity	SM, combining mutations	Chiral acyloins	[91]
P450-BM3	Regio- and enantioselectivity	SM, synthetic libraries	Solid-phase combinatorial gene synthesis compared to PCR	[92]
P450-BM3	Regio- and enantioselectivity	SM	Four-selectivity in tetralones	[93]
P450-BM3	Regio- and enantioselectivity	ISM	Both (<i>R</i>)- and (<i>S</i>)-selective hydroxylation of small molecules	[94]
P450-BM3	Regio- and stereoselectivity	ISM	Hydroxylation of steroids	[95]
P450-BM3	Regio-, diastereo-, and enantioselectivity	SM	Six-membered reduced amino acid alphabet; simultaneous creation of two chirality centers	[96]
P450-BM3	Regio- and diastereoselectivity	Successive site-specific mutagenesis	Hydroxylation of β -cembrenediol	[97]

P450-pyr (<i>Sphingomonas</i> sp. HXN-200)	Regio- and enantioselectivity	ISM	Colorimetric ee-assay	[98]
P450-pyr (<i>Sphingomonas</i> sp. HXN-200)	Regio- and enantioselectivity	ISM	Both enantiomers evolved	[99]
P450-pyr (<i>Sphingomonas</i> sp. HXN-200)	Regioselectivity	ISM	<i>n</i> -Butanol to 1,4-butanediol reaction	[100]
P450-BM3	Regioselectivity	SM	Indirubin synthesis	[101]
P450 monooxygenase (human CYP11B1)	Activity, expression rate	epPCR, site-specific mutagenesis	11 β -Hydroxylation in cortisol production	[102]
P450-BM3	Regio- and diastereoselectivity	ISM	Artemisinin as substrate; fingerprinting	[103]
P450 (<i>Streptomyces carbophilus</i>)	Activity	ISM	Mevastatin to pravastatin by regio- and stereoselective hydroxylation	[104]
P450-BM3	Regio- and diastereoselectivity	epPCR, site-specific mutagenesis; combining mutations	15 β - and 16 β -selectivity in hydroxylation of norethisterone	[105]
P450-BM3	Regioselectivity	epPCR, DNA shuffling, SM	Terminal hydroxylation of palmitic acid	[106]
P450-pyr (<i>Sphingomonas</i> sp. HXN-200)	Enantioselectivity	ISM	Asymmetric sulfoxidation	[107]
P450 (steroidal 15 β -hydroxylase)	Regioselectivity	epPCR, SM, combining mutations	Progesterone 11 α -selectivity	[108]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
P450 (13-hydroxyperoxide lyase)	Activity	DNA shuffling, epPCR	Thermostability also improved	[109]
P450-BM3	Substrate scope	epPCR, DNA shuffling	Chrysene and pyrene as substrates	[110]
P450-pyr (<i>Sphingomonas</i> sp. HXN-200)	Regio- and enantioselectivity	ISM	Subterminal hydroxylation of alkanes	[111]
Alkane monooxygenase (<i>Geobacillus thermodenitrificans</i>)	Regioselectivity	epPCR, SM	Hexadecane terminal hydroxylation	[112]
Galactase oxidase (<i>Fusarium</i> sp.)	Substrate scope	SM, ISM	Useful for labeling glycoproteins	[113]
Laccase (<i>CotA/B. subtilis</i>)	Substrate selectivity	epPCR of 19 amino acid region, Step shuffling	Turning generalist into specialist	[114a]
Laccase (<i>Pycnoporus cinnabarinus</i>)	Activity	ISM	Oxidation of sinapic acid	[114b]
Laccase (CECT 2971)	Activity in blood	epPCR, Step shuffling, DNA shuffling, SM	High redox potential maintained	[115]
Metallo-oxidase McoA (<i>Aquifex aeolicus</i>)	Evolving laccase catalysis	epPCR	Four epPCR rounds; 94 000 transformants	[116]
Laccase (metagenomic)	Activity	epPCR	Textile dye decoloration	[117]
Peroxygenase (<i>Agroclybe aegerita</i>)	Substrate acceptance	epPCR, DNA shuffling, site-specific mutagenesis	Hydroxylation, epoxidation	[118]

Glucose oxidase (<i>A. niger</i>)	Activity	SM	16-residue randomization site; ultra-high-throughput screen	[119]
Bayer – Villiger monooxygenase (PAMO)	Activity, diastereoselectivity	ISM	Five-membered reduced amino acid alphabet	[120]
Bayer – Villiger monooxygenase (PAMO)	Activity, <i>cis/trans</i> -diastereoselectivity	ISM	Two five-residue SM sites; four and five-membered reduced amino acid alphabets at different residues	[121]
Bayer – Villiger monooxygenase (PAMO)	Enantioselectivity	ISM	Asymmetric sulfoxidation	[122]
Bayer – Villiger monooxygenase (PAMO)	Substrate scope, enantioselectivity	SM	Remote “allosteric site” randomization site	[123]
Alcohol dehydrogenase (LIA _{dhA})	NADPH to NADH switch	epPCR, ISM	Next-generation biofuel issue	[124]
Alcohol dehydrogenase (LIA _{dhA})	Activity in isobutanol formation	epPCR	Next-generation biofuel issue	[125]
Alcohol dehydrogenase (<i>Parvibaculum lavamentivorans</i>)	Activity, enantioselectivity	SM	Combining mutations from two libraries	[126]
Alcohol dehydrogenase (<i>Thermoethanolicus brockii</i>)	Enantioselectivity	SM	Induced axial chirality; eight-membered reduced amino acid alphabet	[127]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Succinic semialdehyde reductase (AKR7A5)	Activity	Six variants at position 13 by site-specific mutagenesis	Levulinic acid into 4-hydroxyvaleric acid	[128]
ADH (<i>Candida parapsilosis</i>)	Activity	Site-specific mutagenesis, SM	Stability and solvent resistance also evolved; enantioselectivity not reported	[129]
Alcohol dehydrogenase (<i>Thermoanaerobacter ethanolicus</i>)	Enantioselectivity	SM	Three phenyl-substituted ketones as substrates	[130a]
Carbonyl reductase (<i>Sporobolomyces salmonicolor</i>)	Enantioselectivity	SM	Substituted acetophenones	[130b]
Keto-ester reductase (<i>Candida glabrata</i>)	Substrate scope, enantioselectivity	SM	Reduced amino acid alphabet	[131]
Carbonyl reductase (<i>Leifsonia xyli</i> HS0904)	Enantioselectivity	epPCR, site-specific mutagenesis	Ketone reduction of medicinal interest	[132]
Aldehyde reductase (<i>Saccharomyces cerevisiae</i>)	Activity	epPCR	5-Hydroxymethylfurfural substrate; NADH to NADPH switch	[133]
Aldehyde dehydrogenase (<i>Thermoplasma acidophilum</i>)	Activity; solubility	SM, epPCR	Alternating cycles of mutagenesis methods	[134]

Aminoalcohol dehydrogenase (<i>Rhodococcus erythropolis</i>)	Activity	epPCR	Elimination of product inhibition; D-pseudoephedrine production	[135]
L-Aspartate- β -semialdehyde dehydrogenase (<i>E. coli</i>)	Cofactor NADH acceptance	ISM	L-Homoserine production	[136]
Propanediol oxidoreductase	Activity, furfural tolerance	SM	14 Residues individually targeted; NNK	[137]
Toluene-4-monoxygenase	Activity, regio-stereoselectivity	SM	Hydroxytyrosol production	[138]
Flavoprotein moxygenase (<i>Stenotrophomonas maltophilia</i>)	Enantioselectivity; cofactor promiscuity	SM	Sulfoxidation	[139]
Toluene- <i>o</i> -xylene monoxygenase (<i>Pseudomonas</i> sp. OX1)	Substrate scope, regioselectivity	SM	Oxidative hydroxylation	[140]
Toluene monoxygenase (<i>Bradyrhizobium</i> sp. BTA1)	Activity, regioselectivity	SM, site-specific mutagenesis	Regioselective aromatic hydroxylation	[141]
Leucine dehydrogenase (<i>Bacillus steareothermophilus</i>)	Substrate scope, enantioselectivity	ISM	Reductive amination, single enantiomer evolved	[142]
Phenylalanine dehydrogenase (<i>Rhodococcus</i> sp. M4)	Substrate scope, enantioselectivity	ISM	Reductive amination, single enantiomer evolved	[143]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Phenylalanine dehydrogenase (<i>Bacillus badius</i>)	Substrate scope, enantioselectivity	ISM	Reduced amino acid alphabet, single enantiomer evolved	[144]
D-Amino acid oxidase (from porcine kidney)	Enantioselectivity	ISM	Reversed enantioselectivity favoring (<i>R</i>)-product	[145]
Alanine dehydrogenase (<i>Mycobacterium tuberculosis</i>)	Substrate scope	Site-specific mutagenesis, SM	Evolutionary history of an enzyme can be misleading	[146]
Glycine oxidase (<i>Bacillus cereus</i>)	Activity	epPCR, DNA shuffling	Glyphosate degradation	[147]
Glutamine dehydrogenase (<i>E. coli</i>)	Substrate scope	SM	NNK at four-residue site; 6 million clones screened; L-homophenylalanine production	[148]
D-Lactase dehydrogenase (<i>Lactobacillus helveticus</i>)	Activity	SM	Cofactor switch to nicotinamide flucytosine dinucleotide for bioorthogonality	[149]
Tyrosinase (<i>Ralstonia solanacearum</i>)	Activity	epPCR	Cooperative mutational effects	[150]
Nitrobenzene dioxygenase	Activity	epPCR, SM, DNA shuffling, ISOR	Hydroxytyrosol production	[151]
Dioxygenase (hyoscyamine 6 β -hydroxylase; <i>Anisodus acutangulus</i>)	Activity, regio- and diastereoselectivity	epPCR, SM	Scopolamine synthesis	[152]

Monoamine oxidase (<i>A. niger</i>)	Substrate scope, enantioselectivity	ISM	(<i>R</i>)-Mexiletine synthesis	[153]
Monoamine oxidase (<i>A. niger</i>)	Enantioselectivity	Mutator strain	Deracemization of alkaloids by mutant obtained earlier via mutator strain	[154]
Fructosyl amine oxidase (<i>A. niger</i>)	Substrate scope	SM; combining mutations	Cooperative mutational effects	[155]
Ene reductase (OYE; <i>S. pastorianus</i>)	Enantioselectivity	SM	Z- β -aryl- β -cyanoacrylates as substrates; β -lactams as final products	[156]
Ene reductase (<i>Pichia stipites</i>)	Enantioselectivity	ISM	Baylis – Hillman substrates	[157]
Ene-reductase (<i>Candida macedoniensis</i>)	Stereoselectivity	Site-specific mutagenesis, SM	Loop optimization	[158]
Cutinase (<i>Fusarium solani</i>) reductase	Activity	SM	NNK at seven-residue site; FACS/ <i>in vitro</i> compartmentalization	[159]
Flavin reductase (<i>Vibrio fischeri</i>)	Activity	SM	Different codon at different positions; prodrug activation	[160]
[FeFe] Hydrogenase (<i>Clostridium pasteurianum</i>)	Activity, decreased O ₂ sensitivity	epPCR, SM	Mechanism of deactivation	[161]
[FeFe] hydrogenase (<i>Clostridium acetobutylicum</i>)	Activity	SM	NNK; C298 highly conserved	[162]
N-Acetylaminic acid racemase (<i>Amycolatopsis</i> sp.)	Activity	SM, epPCR, mutator strain	Amino acid dynamic kinetic resolution	[163]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Transaminase (ATA-117)	Activity, enantioselectivity	epPCR, DNA shuffling, ISM	50 kg upscaling of pharma-product sitagliptin; ProSar guidance	[164]
Transaminase (CV 2025)	Activity	SM	Phylogenetic analysis	[165]
Haloalkane dehalogenase (DhaH31)	Stereoselectivity	ISM	Stereocomplementary mutants	[166]
Halohydrin dehalogenase (<i>Arthrobacterium radiobacter</i> HheC)	Enantioselectivity, activity	ISM	DC-analyzer applied	[167]
Halohydrin dehalogenase (<i>Arthrobacter</i> sp. AD2)	Enantioselectivity	ISM	Both enantiomers evolved	[168]
Halogenase (tryptophan 7-halogenase RebH)	Activity, regioselectivity	epPCR	Chlorination; substrate walking	[169]
Arylmalonate decarboxylase (<i>Bordatella bronchoseptica</i>)	Racemase activity	Mutations at different positions by site-directed mutagenesis	Promiscuous reaction mode evolved	[170]
Arylmalonate decarboxylase (<i>B. bronchoseptica</i>)	Racemase activity	Site-directed mutagenesis, SM	Promiscuous reaction mode evolved	[171]
Dialkylglycine decarboxylase	Substrate scope	epPCR, DNA shuffling, SM	Close but mostly remote mutations	[172]
Lysine decarboxylase (<i>Hafnia alvei</i> ASI.1009)	Substrate scope	epPCR, DNA shuffling	Cadaverine production	[173]

Aldolase (<i>N</i> -acetylneuraminic acid lyase; <i>E. coli</i>)	Activity; diastereoselectivity	SM	All 20 variants at position 192	[174]
Aldolase (2-keto-3-deoxy- 6-phosphogluconate, KDPG; <i>E. coli</i>)	Substrate scope, stereoselectivity	SM	Additive mutational effects	[175]
Aldolase (<i>T. maritima</i>)	Substrate selectivity	epPCR	Many distal mutations	[176]
Aldolase (L-KDO)	Substrate scope	epPCR, site-specific mutagenesis ISM	X-ray structures of mutants	[177]
Transketolase (<i>E. coli</i>)	Substrate scope, stereoselectivity	ISM	Single diastereomer evolved	[178]
Transketolase (<i>G. stearothermophilus</i>)	Substrate scope, enantioselelectivity	SM	Different aldehydes; (S)-enantiomers formed	[179]
Transketolase (<i>E. coli</i>)	Substrate acceptance, enantioselelectivity	SM	Different aldehydes; (R)- and (S)-enantiomers	[180]
Transketolase (<i>E. coli</i>)	Substrate scope	ISM	Hydroxy- and carboxy-substituted aromatic aldehydes as substrates	[181]
Transaldolase B (<i>E. coli</i>)	Substrate acceptance	SM, ISM	Nonphosphorylated substrates; synthesis of D-fructose, D-xylulose and L-sorbose	[182]
Transaldolase (<i>Pichia stipitis</i>)	Activity	epPCR	Xylose to ethanol transformation	[183]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Thiamine diphosphate enzyme (<i>E. coli</i>)	Enantioselectivity, substrate acceptance	Site-specific mutagenesis, six variants generated	(<i>R</i>)- and (<i>S</i>)-Enantiomers	[184]
Hydroxynitrile lyase (<i>Baliospermum montanum</i>)	Enantioselectivity, activity	epPCR, SM	(<i>S</i>)-Mandelonitrile produced	[185a]
Hydroxynitrilase (<i>Granulicella tundricola</i>)	Activity, enantioselectivity	SM	Four substrates for pharmaceuticals production	[185b]
Methylaspartate ammonia lyase	Substrate scope, regio- and stereoselectivity	SM	Different amines instead of ammonia	[186]
Phenylalanine aminomutase (<i>Taxus chinensis</i>)	Regioselectivity	SM	Reduced amino acid alphabet; β -selective ammonia addition to cinnamates	[187]
Aspartase (<i>Bacillus</i> sp. YM55-1)	Promiscuity	SM	β -Amino acid lyase activity; cluster screening of 3×10^5 clones	[188]
Pyrrolysyl-tRNA synthetase (<i>Methanosarcina mazei</i>)	Substrate scope	ISM	Modifying proteins by unnatural amino acids for click chemistry	[189]
Human glutathione transferase (A2-2)	Substrate scope	SM	Different codon at each position	[190]
Biotin protein ligase	Substrate scope	epPCR, SM	Desthiobiotin as new substrate	[191]

Human estrogen sulfotransferase	Substrate selectivity	Shuffling (ISOR)	Back-to-consensus library	[192]
Aminoglycoside aminotransferase (APH(3')IIIa)	Substrate scope	epPCR	Amikarin accepted; fitness costs	[193]
Nattokinase (<i>Bacillus natto</i>)	Activity	Family DNA shuffling	Three homologous genes	[194]
RNA polymerase (T7)	Substrate scope	SM	2'-O-Methyl-modified RNA synthesis	[195]
Aminoacyl-tRNA synthetase	Substrate scope	SM	Eight-residue randomization site; fluorotyrosine incorporation	[196]
Aminoacyl-tRNA (<i>Methanosarcina</i>)	Substrate scope	epPCR	Pyrrolysine incorporation	[197]
Pyrrolysyl-tRNA synthetase	Substrate acceptance	SM	Unnatural amino acid acceptance	[198]
DNA polymerase (<i>Thermusaquaticus</i>)	Translesion synthesis activity	epPCR, combining mutations	Forensic application; ancient DNA	[199]
<i>Taq</i> polymerase	Substrate scope	epPCR	Additional nucleotide base pairs with different H-bond patterns accepted	[200]
Deacetoxycephalosporin C synthase (<i>Streptomyces clavuligerus</i>)	Activity	ISM	Ring-expansion of penicillin G to cephalosporin	[201]
Deacetoxy/deacetylcephalosporin C synthase (<i>Acremonium chrysogenum</i>)	Substrate scope	SM	Ring-expansion of penicillins to cephalosporins	[202]

(continued overleaf)

Table 5.1 (Continued)

Enzyme	Evolved property	Mutagenesis method	Comment	References
Carotenoid synthase (<i>Straphylococcus aureus</i> C30)	Substrate scope	epPCR	C30 and C40 carotenoid synthesis	[203]
Polyketide synthase (<i>Saccharopolyspora erythraea</i>)	Substrate scope	SM	Nonnative functional group into erythromycin	[204]
Malonyl-CoA synthetase (<i>Rn trifolii</i>)	Substrate scope	SM	Extended polyketide chemistry	[205]
Cytokine (MIF)	Dehalogenase activity	SM	Promiscuous reaction	[206]
Cytolysin A (<i>Salmonella typhi</i>)	Hemolytic activity	epPCR	Electrical properties in lipid bilayers; nanopore control	[207]
Methyltransferase	Regioselectivity, activity	ISM	Modulating lignin biosynthesis for paper production and biofuels	[208]
Catalytic antibody (6D9)	Activity	SM	Six-amino acid alphabet; phage display analysis	[209]
Antibody against snake venom	Binding selectivity	epPCR	Phage display detection	[210]

epPCR, error-prone polymerase chain reaction; SM, saturation mutagenesis; ISM, iterative saturation mutagenesis; SteP, staggered extension process; SeSaM, sequence saturation mutagenesis; and ISOR, incorporating synthetic oligonucleotides via gene reassembly.

In some cases different enzyme labels are used for the same reaction type, for example, alcohol dehydrogenase and carbonyl reductase, both as catalysts in enantioselective reduction of ketones.

5.2

Collection of Selected Examples from the Literature 2010 up to 2016

References

1. Reetz, M.T., Zonta, A., Schimossek, K., Jaeger, K.-E., and Liebeton, K. (1997) Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. *Angew. Chem. Int. Ed. Engl.*, **36** (24), 2830–2832.
2. Recent reviews of directed evolution of enzymes: (a) 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; (b) Bommarius, A.S. (2015) Biocatalysis: a status report. *Annu. Rev. Chem. Biomol. Eng.*, **6**, 319–345; (c) Jäckel, C. and Hilvert, D. (2010) Biocatalysts by evolution. *Curr. Opin. Biotechnol.*, **21** (6), 753–759; (d) Goldsmith, M. and Tawfik, D.S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.*, **22** (4), 406–412; (e) 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; (f) Widersten, M. (2014) Protein engineering for development of new hydrolytic biocatalysts. *Curr. Opin. Chem. Biol.*, **21**, 42–47; (g) Denard, C.A., Ren, H., and Zhao, H. (2015) Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.*, **25**, 55–64; (h) Lutz, S. and Bornscheuer, U.T. (eds) (2009) *Protein Engineering Handbook*, Wiley-VCH Verlag GmbH, Weinheim.
3. (a) Bassegoda, A., Cesarini, S., and Diaz, P. (2012) Lipase improvement: goals and strategies. *Comput. Struct. Biotechnol. J.*, **2**, e201209005; (b) Schmidt, M., Böttcher, D., and Bornscheuer, U.T. (2009) Protein engineering of carboxyl esterases by rational design and directed evolution. *Protein Pept. Lett.*, **16**, 1162–1171; (c) Kourist, R., Brundiek, H., and Bornscheuer, U.T. (2010) Protein engineering and discovery of lipases. *Eur. J. Lipid Sci. Technol.*, **112**, 64–74.
4. (a) Bolt, A., Berry, A., and Nelson, A. (2008) Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Arch. Biochem. Biophys.*, **474**, 318–330; (b) Clapes, P. and Garrabou, X. (2011) Current trends in asymmetric synthesis with aldolases. *Adv. Synth. Catal.*, **353**, 2263–2283; (c) Dadashpour, M. and Asano, Y. (2011) Hydroxynitrile lyases: insights into biochemistry, discovery and engineering. *ACS Catal.*, **1**, 1121–1149.
5. (a) Whitehouse, C.J., Bell, S.G., and Wong, L.-L. (2012) P450 (BM3) (CYP102A1): connecting the dots. *Chem. Soc. Rev.*, **41** (3), 1218–1260; (b) Fasan, R. (2012) Tuning P450 enzymes as oxidation catalysts. *ACS Catal.*, **2** (4), 647–666; (c) Bernhardt, R. and Urlacher, V.B. (2014) Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl. Microbiol. Biotechnol.*, **98** (14), 6185–6203; (d) Holtmann, D., Fraaije, M.W., Arends, I.W., Opperman, D.J., and Hollmann, F. (2014) The taming of oxygen: biocatalytic oxyfunctionalisations. *Chem. Commun.*, **50** (87), 13180–13200; (e) Roiban, G.-D. and Reetz, M.T. (2015) Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.*, **51**, 2208–2224.
6. (a) Homann, A. and Seibel, J. (2009) Towards tailor-made oligosaccharides-chemo-enzymatic approaches by enzyme and substrate engineering. *Appl. Microbiol. Biotechnol.*, **83** (2), 209–216; (b) Alcalde, M. (2015) Engineering the ligninolytic consortium. *Trends Biotechnol.*, **33**, 155–162; (c) Pardo, I. and Camarero, S. (2015) Laccase engineering by rational and

- evolutionary design. *Cell. Mol. Life Sci.*, **72**, 897–910; (d) Han, R., Li, J., Shin, H.D., Chen, R.R., Du, G., Liu, L., and Chen, J. (2014) Recent advances in discovery, heterologous expression, and molecular engineering of cyclodextrin glycosyltransferase for versatile applications. *Biotechnol. Adv.*, **32** (2), 415–428; (e) Bommarius, A.S., Sohn, M., Kang, Y., Lee, J.H., and Realff, M.J. (2014) Protein engineering of cellulases. *Curr. Opin. Biotechnol.*, **29**, 139–145.
7. Pingoud, A. and Wende, W. (2011) Generation of novel nucleases with extended specificity by rational design and combinatorial strategies. *Chem-BioChem*, **12**, 1495–1500.
 8. Zhang, Z.G., Parra, L.P., and Reetz, M.T. (2012) Protein engineering of stereoselective Baeyer-Villiger monooxygenases. *Chem. Eur. J.*, **18** (33), 10160–10172.
 9. (a) Tsai, P.C., Fox, N., Bigley, A.N., Harvey, S.P., Barondeau, D.P., and Raushel, F.M. (2012) Enzymes for the homeland defense: optimizing phosphotriesterase for the hydrolysis of organophosphate nerve agents. *Biochemistry*, **51** (32), 6463–6475; (b) Khersonsky, O. and Tawfik, D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.*, **79**, 471–505; (c) Toth-Petroczy, A. and Tawfik, D.S. (2014) The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.*, **26**, 131–138.
 10. Wang, J., Wang, D., Wang, B., Mei, Z.-H., Liu, J., and Yu, H.-W. (2012) Enhanced activity of *Rhizomucor miehei* lipase by directed evolution with simultaneous evolution of the propeptide. *Appl. Microbiol. Biotechnol.*, **96** (2), 443–450.
 11. Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y., and Bocola, M. (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.*, **132** (26), 9144–9152.
 12. Engström, K., Nyhlen, J., Sandström, A.G., and Bäckvall, J.E. (2010) Directed evolution of an enantioselective lipase with broad substrate scope for hydrolysis of alpha-substituted esters. *J. Am. Chem. Soc.*, **132** (20), 7038–7042.
 13. Wikmark, Y., Svedendahl Humble, M., and Bäckvall, J.-E. (2015) Combinatorial library based engineering of *Candida antarctica* lipase A for enantioselective transacylation of sec-alcohols in organic solvent. *Angew. Chem. Int. Ed.*, **54** (14), 4284–4288.
 14. Sandström, A.G., Wikmark, Y., Engström, K., Nyhlen, J., and Bäckvall, J.E. (2012) Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (1), 78–83.
 15. Brundiek, H.B., Evitt, A.S., Kourist, R., and Bornscheuer, U.T. (2012) Creation of a lipase highly selective for trans fatty acids by protein engineering. *Angew. Chem. Int. Ed.*, **51** (2), 412–414.
 16. Qin, B., Liang, P., Jia, X., Zhang, X., Mu, M., Wang, X.-Y., Ma, G.-Z., Jin, D.-N., and You, S. (2013) Directed evolution of *Candida antarctica* lipase B for kinetic resolution of profen esters. *Catal. Commun.*, **38**, 1–5.
 17. Yen, C.-C., Malmis, C.C., Lee, G.-C., Lee, L.-C., and Shaw, J.-F. (2010) Site-specific saturation mutagenesis on residues 132 and 450 of *Candida rugosa* lip2 enhances catalytic efficiency and alters substrate specificity in various chain lengths of triglycerides and esters. *J. Agric. Food Chem.*, **58** (20), 10899–10905.
 18. Colin, D.Y., Deprez-Beauclair, P., Silva, N., Infantes, L., and Kerfelec, B. (2010) Modification of pancreatic lipase properties by directed molecular evolution. *Protein Eng. Des. Sel.*, **23** (5), 365–373.
 19. Wu, Q., Soni, P., and Reetz, M.T. (2013) Laboratory evolution of enantiocomplementary *Candida antarctica* lipase B mutants with broad substrate scope. *J. Am. Chem. Soc.*, **135** (5), 1872–1881.
 20. Panizza, P., Cesarini, S., Diaz, P., and Rodriguez Giordano, S. (2015) Saturation mutagenesis in selected amino acids to shift *Pseudomonas* sp. acidic lipase lip I.3 substrate specificity and

- activity. *Chem. Commun.*, **51** (7), 1330–1333.
21. 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.
 22. Tang, L., Su, M., Yan, J., Xie, S., and Zhang, W. (2015) Lid hinge region of *Penicillium expansum* lipase affects enzyme activity and interfacial activation. *Process Biochem.*, **50**, 1218–1223.
 23. Li, X.-J., Zheng, R.-C., Ma, H.-Y., Huang, J.-F., and Zheng, Y.-G. (2014) Key residues responsible for enhancement of catalytic efficiency of *Thermomyces lanuginosus* lipase Lip revealed by complementary protein engineering strategy. *J. Biotechnol.*, **188**, 29–35.
 24. Akbulut, N., Tuzlakoglu Öztürk, M., Pijning, T., İşsever Öztürk, S., and Gümüsel, F. (2013) Improved activity and thermostability of *Bacillus pumilus* lipase by directed evolution. *J. Biotechnol.*, **164** (1), 123–129.
 25. Kumar, V., Yedavalli, P., Gupta, V., and Rao, N.M. (2014) Engineering lipase a from mesophilic *Bacillus subtilis* for activity at low temperatures. *Protein Eng. Des. Sel.*, **27** (3), 73–82.
 26. Zhou, X., Gao, L., Yang, G., Liu, D., Bai, A., Li, B., Deng, Z., and Feng, Y. (2015) Design of hyperthermophilic lipase chimeras by key motif-directed recombination. *ChemBioChem*, **16**, 455–462.
 27. Ma, J., Wu, L., Guo, F., Gu, J., Tang, X., Jiang, L., Liu, J., Zhou, J., and Yu, H. (2013) Enhanced enantioselectivity of a carboxyl esterase from *Rhodobacter sphaeroides* by directed evolution. *Appl. Microbiol. Biotechnol.*, **97** (11), 4897–4906.
 28. Godinho, L.F., Reis, C.R., van Merkerk, R., Poelarends, G.J., and Quax, W.J. (2012) An esterase with superior activity and enantioselectivity towards 1,2-O-isopropylidenglycerol esters obtained by protein design. *Adv. Synth. Catal.*, **354** (16), 3009–3015.
 29. (a) Godinho, L.F., Reis, C.R., Rozeboom, H.J., Dekker, F.J., Dijkstra, B.W., Poelarends, G.J., and Quax, W.J. (2012) Enhancement of the enantioselectivity of carboxylesterase a by structure-based mutagenesis. *J. Biotechnol.*, **158** (1–2), 36–43; (b) Gong, Y., Xu, G.-C., Chen, Q., Yin, J.-G., Li, C.-X., and Xu, J.-H. (2016) Iterative multitarget evolution dramatically enhances the enantioselectivity and catalytic efficiency of *Bacillus subtilis* esterase towards bulky benzoate esters of D,L-menthol. *Catal. Sci. Technol.* doi: 10.1039/c5cy01723h.
 30. Nobili, A., Tao, Y., Pavlidis, I.V., van den Bergh, T., Joosten, H.-J., Tan, T., and Bornscheuer, U.T. (2015) Simultaneous use of in silico design and a correlated mutation network as a tool to efficiently guide enzyme engineering. *ChemBioChem*, **16** (5), 805–810.
 31. Jochens, H. and Bornscheuer, U.T. (2010) Natural diversity to guide focused directed evolution. *ChemBioChem*, **11** (13), 1861–1866.
 32. Luan, Z.-J., Li, F.-L., Dou, S., Chen, Q., Kong, X.-D., Zhou, J., Yu, H.-L., and Xu, J.-H. (2015) Substrate channel evolution of an esterase for the synthesis of cilastatin. *Catal. Sci. Technol.*, **5** (5), 2622–2629.
 33. Meier, M.M., Rajendran, C., Malisi, C., Fox, N.G., Xu, C., Schlee, S., Barondeau, D.P., Höcker, B., Sterner, R., and Rauschel, F.M. (2013) Molecular engineering of organophosphate hydrolysis activity from a weak promiscuous lactonase template. *J. Am. Chem. Soc.*, **135** (31), 11670–11677.
 34. Shivange, A., Serwe, A., Dennig, A., Roccatano, D., Haefner, S., and Schwaneberg, U. (2012) Directed evolution of a highly active *Yersinia mollaretii* phytase. *Appl. Microbiol. Biotechnol.*, **95** (2), 405–418.
 35. Chen, W., Ye, L., Guo, F., Lv, Y., and Yu, H. (2015) Enhanced activity of an alkaline phytase from *Bacillus subtilis* 168 in acidic and neutral environments by directed evolution. *Biochem. Eng. J.*, **98**, 137–143.
 36. Liao, Y., Zeng, M., Wu, Z.-F., Chen, H., Wang, H.-N., Wu, Q., Shan, Z., and Han, X.-Y. (2012) Improving

- phytase enzyme activity in a recombinant mutant phytase from *Aspergillus niger* N25 by error-prone PCR. *Appl. Biochem. Biotechnol.*, **166** (3), 549–562.
37. Deaguero, A.L., Blum, J.K., and Bommarius, A.S. (2012) Improving the diastereoselectivity of penicillin G acylase for ampicillin synthesis from racemic substrates. *Protein Eng. Des. Sel.*, **25** (3), 135–144.
 38. Liu, C., Yang, G., Wu, L., Tian, G., Zhang, Z., and Feng, Y. (2011) Switch of substrate specificity of hyperthermophilic acylaminoacyl peptidase by combination of protein and solvent engineering. *Protein Cell*, **2** (6), 497–506.
 39. Conti, G., Pollegioni, L., Molla, G., and Rosini, E. (2014) Strategic manipulation of an industrial biocatalyst – evolution of a cephalosporin C acylase. *FEBS J.*, **281** (10), 2443–2455.
 40. Kotzia, G.A. and Labrou, N.E. (2011) Engineering substrate specificity of *E. carotovora* L-asparaginase for the development of biosensor. *J. Mol. Catal. B: Enzym.*, **72** (3–4), 95–101.
 41. Martinez, R., Jakob, F., Tu, R., Siegert, P., Maurer, K.-H., and Schwaneberg, U. (2013) Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol. Bioeng.*, **110** (3), 711–720.
 42. Sun, S., Zhang, W., Mannervik, B., and Andersson, D.I. (2013) Evolution of broad spectrum β -lactam resistance in an engineered metallo- β -lactamase. *J. Biol. Chem.*, **288** (4), 2314–2324.
 43. Chen, I., Dorr, B.M., and Liu, D.R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.*, **108** (28), 11399–11404.
 44. Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011) Directed evolution of the nonribosomal peptide synthetase ADMK generates new andrimid derivatives in vivo. *Chem. Biol.*, **18** (5), 601–607.
 45. Alvizo, O., Nguyen, L.J., Savile, C.K., Bresson, J.A., Lakhapatri, S.L., Solis, E.O.P., Fox, R.J., Broering, J.M., Benoit, M.R., Zimmerman, S.A., Novick, S.J., Liang, J., and Lalonde, J.J. (2014) Directed evolution of an ultrastable carbonic anhydrase for highly efficient carbon capture from flue gas. *Proc. Natl. Acad. Sci. U.S.A.*, **111** (46), 16436–16441.
 46. Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R., and Tawfik, D.S. (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.*, **3**, 1257.
 47. Naqvi, T., Warden, A.C., French, N., Sugrue, E., Carr, P.D., Jackson, C.J., and Scott, C. (2014) A 5000-fold increase in the specificity of a bacterial phosphotriesterase for malathion through combinatorial active site mutagenesis. *PLoS One*, **9** (4), e94177.
 48. Bigley, A.N., Xu, C., Henderson, T.J., Harvey, S.P., and Rauschel, F.M. (2013) Enzymatic neutralization of the chemical warfare agent VX: evolution of phosphotriesterase for phosphorothiolate hydrolysis. *J. Am. Chem. Soc.*, **135** (28), 10426–10432.
 49. Cherny, I., Greisen, P. Jr., Ashani, Y., Khare, S.D., Oberdorfer, G., Leader, H., Baker, D., and Tawfik, D.S. (2013) Engineering V-type nerve agents detoxifying enzymes using computationally focused libraries. *ACS Chem. Biol.*, **8** (11), 2394–2403.
 50. Kotik, M., Zhao, W., Iacazio, G., and Archelas, A. (2013) Directed evolution of metagenome-derived epoxide hydrolase for improved enantioselectivity and enantioconvergence. *J. Mol. Catal. B: Enzym.*, **91**, 44–51.
 51. Sun, Z., Lonsdale, R., Kong, X.-D., Xu, J.-H., Zhou, J., and Reetz, M.T. (2015) Reshaping an enzyme binding pocket for enhanced and inverted stereoselectivity: use of smallest amino acid alphabet in directed evolution. *Angew. Chem. Int. Ed.*, **54**, 12410–12415.
 52. Gumulya, Y., Sanchis, J., and Reetz, M.T. (2012) Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *ChemBioChem*, **13** (7), 1060–1066.
 53. Reetz, M.T. and Zheng, H.B. (2011) Manipulating the expression rate

- and enantioselectivity of an epoxide hydrolase by using directed evolution. *ChemBioChem*, **12** (10), 1529–1535.
54. Zheng, H.B. and Reetz, M.T. (2010) Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis. *J. Am. Chem. Soc.*, **132** (44), 15744–15751.
 55. Zheng, H.B., Kahakeaw, D., Acevedo, J.P., and Reetz, M.T. (2010) Directed evolution of enantioconvergence: the case of an epoxide hydrolase-catalyzed reaction of a racemic epoxide. *Chem-CatChem*, **2** (8), 958–961.
 56. Gurell, A. and Widersten, M. (2010) Modification of substrate specificity resulting in an epoxide hydrolase with shifted enantiopreference for (2,3-epoxypropyl)benzene. *ChemBioChem*, **11** (10), 1422–1429.
 57. Carlsson, Å.J., Bauer, P., Ma, H., and Widersten, M. (2012) Obtaining optical purity for product diols in enzyme-catalyzed epoxide hydrolysis: contributions from changes in both enantio- and regioselectivity. *Biochemistry*, **51** (38), 7627–7637.
 58. Kotik, M., Archelas, A., Faměrová, V., Oubrechtová, P., and Křen, V. (2011) Laboratory evolution of an epoxide hydrolase – towards an enantioconvergent biocatalyst. *J. Biotechnol.*, **156** (1), 1–10.
 59. (a) Xue, Y.-P., Shi, C.-C., Xu, Z., Jiao, B., Liu, Z.-Q., Huang, J.-F., Zheng, Y.-G., and Shen, Y.-C. (2015) Design of nitrilases with superior activity and enantioselectivity towards sterically hindered nitrile by protein engineering. *Adv. Synth. Catal.*, **357** (8), 1741–1750; (b) Liu, Z.-Q., Zhang, X.-H., Xue, Y.-P., Xu, M., and Zheng, Y.-G. (2014) Improvement of *Alcaligenes faecalis* Nitrilase by gene site saturation mutagenesis and its application in stereospecific biosynthesis of (R)-(-)-mandelic acid. *J. Agric. Food Chem.*, **62** (20), 4685–4694.
 60. Wang, H., Gao, W., Sun, H., Chen, L., Zhang, L., Wang, X., and Wie, D. (2015) Protein engineering of a nitrilase from *Burgholderia cenocepacia* J 2315 for efficient and enantioselective production of (R)-o-chloromandelic acid. *Appl. Environ. Microbiol.*, **81**, 8469–8477.
 61. Song, W., Sun, X., Chen, X., Liu, D., and Liu, L. (2015) Enzymatic production of L-citrulline by hydrolysis of the guanidinium group of L-arginine with recombinant arginine deiminase. *J. Biotechnol.*, **208**, 37–43.
 62. Champion, E., Guérin, F., Moulis, C., Barbe, S., Tran, T.H., Morel, S., Descroix, K., Monsan, P., Mourey, L., Mulard, L.A., Tranier, S., Remaud-Siméon, M., and André, I. (2012) Applying pairwise combinations of amino acid mutations for sorting out highly efficient glucosylation tools for chemo-enzymatic synthesis of bacterial oligosaccharides. *J. Am. Chem. Soc.*, **134** (45), 18677–18688.
 63. de Abreu, M., Alvaro-Benito, M., Sanz-Aparicio, J., Plou, F.J., Fernandez-Lobato, M., and Alcalde, M. (2013) Synthesis of 6-kestose using an efficient β -fructofuranosidase engineered by directed evolution. *Adv. Synth. Catal.*, **355** (9), 1698–1702.
 64. Ihssen, J., Haas, J., Kowarik, M., Wiesli, L., Wacker, M., Schwede, T., and Thöny-Meyer, L. (2015) Increased efficiency of *Campylobacter jejuni* N-oligosaccharyltransferase PglB by structure-guided engineering. *Open Biol.*, **5** (4), 140227.
 65. André-Miral, C., Koné, F.M., Solleux, C., Grandjean, C., Dion, M., Tran, V., and Tellier, C. (2015) De novo design of a trans- β -N-acetylglucosaminidase activity from a GH1 β -glycosidase by mechanism engineering. *Glycobiology*, **25** (4), 394–402.
 66. Pennec, A., Daniellou, R., Loyer, P., Nugier-Chauvin, C., and Ferrières, V. (2015) Araf51 with improved transglycosylation activities: one engineered biocatalyst for one specific acceptor. *Carbohydr. Res.*, **402**, 50–55.
 67. Han, R., Liu, L., Shin, H.-d., Chen, R.R., Li, J., Du, G., and Chen, J. (2013) Iterative saturation mutagenesis of -6 subsite residues in cyclodextrin glycosyltransferase from *Paenibacillus macerans* to improve maltodextrin

- specificity for 2-O-d-glucopyranosyl-L-ascorbic acid synthesis. *Appl. Environ. Microbiol.*, **79** (24), 7562–7568.
68. Trollope, K.M., Nieuwoudt, H.H., Görgens, J.F., and Volschenk, H. (2014) Screening a random mutagenesis library of a fungal β -fructofuranosidase using FT-MIR ATR spectroscopy and multivariate analysis. *Appl. Microbiol. Biotechnol.*, **98** (9), 4063–4073.
 69. Turek, D., Klimeš, P., Mazura, P., and Brzobohatý, B. (2014) Combining rational and random strategies in β -glucosidase Zm-p60.1 protein library construction. *PLoS One*, **9** (9), e108292.
 70. Shim, J.-H., Chen, H.-M., Rich, J.R., Goddard-Borger, E.D., and Withers, S.G. (2012) Directed evolution of a β -glycosidase from *Agrobacterium* sp. to enhance its glycosynthase activity toward C3-modified donor sugars. *Protein Eng. Des. Sel.*, **25** (9), 465–472.
 71. Chen, Y., Zhang, B., Pei, H., Lv, J., Yang, W., Cao, Y., and Dong, B. (2012) Directed evolution of *Penicillium janczewskii* zalesk α -galactosidase toward enhanced activity and expression in *Pichia pastoris*. *Appl. Biochem. Biotechnol.*, **168** (3), 638–650.
 72. Ihssen, J., Kowarik, M., Wiesli, L., Reiss, R., Wacker, M., and Thony-Meyer, L. (2012) Structural insights from random mutagenesis of *Campylobacter jejuni* oligosaccharyltransferase PglB. *BMC Biotech.*, **12** (1), 67.
 73. Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., Fauré, R., and O'Donohue, M.J. (2013) Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.*, **30** (5), 536–544.
 74. Gantt, R.W., Peltier-Pain, P., Singh, S., Zhou, M., and Thorson, J.S. (2013) Broadening the scope of glycosyl transferase-catalyzed sugar nucleotide synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (19), 7648–7653.
 75. Wang, Y., Feng, S., Zhan, T., Huang, Z., Wu, G., and Liu, Z. (2013) Improving catalytic efficiency of endo- β -1,4-xylanase from *Geobacillus stearothermophilus* by directed evolution and H179 saturation mutagenesis. *J. Biotechnol.*, **168** (4), 341–347.
 76. Daudé, D., Topham, C.M., Remaud-Siméon, M., and André, I. (2013) Probing impact of active site residue mutations on stability and activity of *Neisseria polysaccharea* amylsucrase. *Protein Sci.*, **22** (12), 1754–1765.
 77. Chuang, H.-Y., Suen, C.-S., Hwang, M.-J., and Roffler, S.R. (2015) Toward reducing immunogenicity of enzyme replacement therapy: altering the specificity of human β -glucuronidase to compensate for α -iduronidase deficiency. *Protein Eng. Des. Sel.* doi: 10.1093/protein/gzv041
 78. Keys, T.G., Fuchs, H.L.S., Ehrhit, J., Alves, J., Freiberger, F., and Gerardy-Schahn, R. (2014) Engineering the product profile of a polysialyltransferase. *Nat. Chem. Biol.*, **10** (6), 437–442.
 79. (a) Liu, M., Xie, W., Xu, H., Gu, J., Lv, X., Yu, H., and Ye, L. (2014) Directed evolution of an exoglucanase facilitated by a co-expressed β -glucosidase and construction of a whole engineered cellulase system in *Escherichia coli*. *Biotechnol. Lett.*, **36** (9), 1801–1807; see also: (b) Liu, M., Gu, J., Xie, W., and Yu, H. (2013) Directed co-evolution of an endoglucanase and a β -glucosidase in *Escherichia coli* by a novel high-throughput screening method. *Chem. Commun.*, **49** (65), 7219–7221.
 80. Song, L., Siguier, B., Dumon, C., Bozonnet, S., and O'Donohue, M. (2012) Engineering better biomass-degrading ability into a GH11 xylanase using a directed evolution strategy. *Biotechnol. Biofuels*, **5** (1), 3.
 81. Choi, Y.H., Kim, J.H., Park, J.H., Lee, N., Kim, D.-H., Jang, K.-S., Park, I.-H., and Kim, B.-G. (2014) Protein engineering of α 2,3/2,6-sialyltransferase to improve the yield and productivity of in vitro sialyllactose synthesis. *Glycobiology*, **24** (2), 159–169.
 82. Song, B., Yue, Y., Xie, T., Qian, S., and Chao, Y. (2014) Mutation of tyrosine167histidine at remote substrate binding subsite –6 in α -cyclodextrin

- glycosyltransferase enhancing α -cyclodextrin specificity by directed evolution. *Mol. Biotechnol.*, **56** (3), 232–239.
83. Lee, S.-M., Jellison, T., and Alper, H.S. (2012) Directed evolution of xylose isomerase for improved xylose catabolism and fermentation in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **78** (16), 5708–5716.
 84. Van der Borght, J., Chen, C., Hoflack, L., Van Renterghem, L., Desmet, T., and Soetaert, W. (2011) Enzymatic properties and substrate specificity of the tetralose phosphorylase from *Caldanaerobacter subterraneus*. *Appl. Environ. Microbiol.*, **77**, 6939–6944.
 85. Bosshart, A., Hee, C.S., Bechtold, M., Schirmer, T., and Panke, S. (2015) Directed divergent evolution of a thermostable D-tagatose epimerase towards improved activity for two hexose substrates. *ChemBioChem*, **16** (4), 592–601.
 86. Wagner, N., Bosshart, A., Failmezger, J., Bechtold, M., and Panke, S. (2015) A separation-integrated cascade reaction to overcome thermodynamic limitations in rare-sugar synthesis. *Angew. Chem. Int. Ed.*, **54** (14), 4182–4186.
 87. Lewis, J.C., Mantovani, S.M., Fu, Y., Snow, C.D., Komor, R.S., Wong, C.-H., and Arnold, F.H. (2010) Combinatorial alanine substitution enables rapid optimization of cytochrome P450 BM3 for selective hydroxylation of large substrates. *ChemBioChem*, **11** (18), 2502–2505.
 88. Chen, M.M.Y., Snow, C.D., Vizcarra, C.L., Mayo, S.L., and Arnold, F.H. (2012) Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng. Des. Sel.*, **25** (4), 171–178.
 89. Ritter, C., Nett, N., Acevedo-Rocha, C.G., Lonsdale, R., Kräling, K., Dempwolff, F., Hoebeinreich, S., Graumann, P.L., Reetz, M.T., and Meggers, E. (2015) Bioorthogonal enzymatic activation of caged compounds. *Angew. Chem. Int. Ed.*, **54**, 13440–13443.
 90. Dennig, A., Lulsdorf, N., Liu, H.F., and Schwaneberg, U. (2013) Regioselective o-hydroxylation of monosubstituted benzenes by P450 BM3. *Angew. Chem. Int. Ed.*, **52** (32), 8459–8462.
 91. Agudo, R., Roiban, G.-D., Lonsdale, R., Ilie, A., and Reetz, M.T. (2015) Biocatalytic route to chiral acyloins: P450-catalyzed regio- and enantioselective α -hydroxylation of ketones. *J. Org. Chem.*, **80**, 950–956.
 92. Hoebeinreich, S., Zilly, F.E., Acevedo-Rocha, C.G., Zilly, M., and Reetz, M.T. (2015) Speeding up directed evolution: combining the advantages of solid-phase combinatorial gene synthesis with statistically guided reduction of screening effort. *ACS Synth. Biol.*, **4** (3), 317–331.
 93. Roiban, G.D., Agudo, R., Ilie, A., Lonsdale, R., and Reetz, M.T. (2014) CH-activating oxidative hydroxylation of 1-tetralones and related compounds with high regio- and stereoselectivity. *Chem. Commun.*, **50** (92), 14310–14313.
 94. Agudo, R., Roiban, G.-D., and Reetz, M.T. (2012) Achieving regio- and enantioselectivity of P450-catalyzed oxidative CH activation of small functionalized molecules by structure-guided directed evolution. *Chem-BioChem*, **13** (10), 1465–1473.
 95. Kille, S., Zilly, F.E., Acevedo, J.P., and Reetz, M.T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.*, **3** (9), 738–743.
 96. Roiban, G.-D., Agudo, R., and Reetz, M.T. (2014) Cytochrome P450 catalyzed oxidative hydroxylation of achiral organic compounds with simultaneous creation of two chirality centers in a single C-H activation step. *Angew. Chem. Int. Ed.*, **53** (33), 8659–8663.
 97. Le-Huu, P., Heidt, T., Claasen, B., Laschat, S., and Urlacher, V.B. (2015) Chemo-, regio-, and stereoselective oxidation of the monocyclic diterpenoid β -cembrene-10,11-diol by P450 BM3. *ACS Catal.*, **5** (3), 1772–1780.
 98. Tang, W.L., Li, Z., and Zhao, H. (2010) Inverting the enantioselectivity of

- P450pyr monooxygenase by directed evolution. *Chem. Commun.*, **46** (30), 5461–5463.
99. (a) Pham, S.Q., Pompidor, G., Liu, J., Li, X.-D., and Li, Z. (2012) Evolving P450pyr hydroxylase for highly enantioselective hydroxylation at non-activated carbon atom. *Chem. Commun.*, **48** (38), 4618–4620; (b) Pham, S.Q., Gao, P., and Li, Z. (2013) Engineering of recombinant *E. coli* cells co-expressing P450pyrTM monooxygenase and glucose dehydrogenase for highly regio- and stereoselective hydroxylation of alicycles with cofactor recycling. *Biotechnol. Bioeng.*, **110** (2), 363–373; Review of directed evolution of P450-pyr: (c) Yang, Y. and Li, Z. (2015) Evolving P450pyr monooxygenase for regio- and stereoselective hydroxylations. *Chimia*, **69**, 136–141.
 100. Yang, Y., Chi, Y.T., Toh, H.H., and Li, Z. (2015) Evolving P450pyr monooxygenase for highly regioselective terminal hydroxylation of *n*-butanol to 1,4-butanediol. *Chem. Commun.*, **51** (5), 914–917.
 101. Hu, S., Huang, J., Mei, L., Yu, Q., Yao, S., and Jin, Z. (2010) Altering the regioselectivity of cytochrome P450 BM-3 by saturation mutagenesis for the biosynthesis of indirubin. *J. Mol. Catal. B: Enzym.*, **67** (1–2), 29–35.
 102. Schiffer, L., Anderko, S., Hobler, A., Hannemann, F., Kagawa, N., and Bernhardt, R. (2015) A recombinant CYP11B1 dependent *Escherichia coli* biocatalyst for selective cortisol production and optimization towards a preparative scale. *Microb. Cell Fact.*, **14** (1), 25.
 103. Zhang, K., Shafer, B.M., Demars, M.D., Stern, H.A., and Fasan, R. (2012) Controlled oxidation of remote sp³ C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.*, **134** (45), 18695–18704.
 104. Ba, L., Li, P., Zhang, H., Duan, Y., and Lin, Z. (2013) Semi-rational engineering of cytochrome P450sca-2 in a hybrid system for enhanced catalytic activity: insights into the important role of electron transfer. *Biotechnol. Bioeng.*, **110** (11), 2815–2825.
 105. Reinen, J., Vredenburg, G., Klaering, K., Vermeulen, N.P.E., Commandeur, J.N.M., Honing, M., and Vos, J.C. (2015) Selective whole-cell biosynthesis of the designer drug metabolites 15- or 16- β -hydroxynorethisterone by engineered cytochrome P450 BM3 mutants. *J. Mol. Catal. B: Enzym.*, **121**, 64–74.
 106. Brühlmann, F., Fourage, L., Ullmann, C., Haefliger, O.P., Jeckelmann, N., Dubois, C., and Wahler, D. (2014) Engineering cytochrome P450 BM3 of *Bacillus megaterium* for terminal oxidation of palmitic acid. *J. Biotechnol.*, **184**, 17–26.
 107. Gao, P., Li, A., Lee, H.H., Wang, D.I.C., and Li, Z. (2014) Enhancing enantioselectivity and productivity of P450-catalyzed asymmetric sulfoxidation with an aqueous/ionic liquid biphasic system. *ACS Catal.*, **4** (10), 3763–3771.
 108. Nguyen, K.T., Virus, C., Günnewich, N., Hannemann, F., and Bernhardt, R. (2012) Changing the regioselectivity of a P450 from C15 to C11 hydroxylation of progesterone. *ChemBioChem*, **13** (8), 1161–1166.
 109. Brühlmann, F., Bosijokovic, B., Ullmann, C., Auffray, P., Fourage, L., and Wahler, D. (2013) Directed evolution of a 13-hydroperoxide lyase (CYP74B) for improved process performance. *J. Biotechnol.*, **163** (3), 339–345.
 110. Sideri, A., Goyal, A., Di Nardo, G., Tsotsou, G.E., and Gilardi, G. (2013) Hydroxylation of non-substituted polycyclic aromatic hydrocarbons by cytochrome P450 BM3 engineered by directed evolution. *J. Inorg. Biochem.*, **120**, 1–7.
 111. Yang, Y., Liu, J., and Li, Z. (2014) Engineering of P450pyr hydroxylase for the highly regio- and enantioselective subterminal hydroxylation of alkanes. *Angew. Chem. Int. Ed.*, **53** (12), 3120–3124.
 112. Dong, Y., Yan, J., Du, H., Chen, M., Ma, T., and Feng, L. (2012) Engineering of LadA for enhanced hexadecane

- oxidation using random- and site-directed mutagenesis. *Appl. Microbiol. Biotechnol.*, **94** (4), 1019–1029.
113. Rannes, J.B., Ioannou, A., Willies, S.C., Grogan, G., Behrens, C., Flitsch, S.L., and Turner, N.J. (2011) Glycoprotein labeling using engineered variants of galactose oxidase obtained by directed evolution. *J. Am. Chem. Soc.*, **133** (22), 8436–8439.
 114. (a) Gupta, N., Lee, F.S., and Farinas, E.T. (2010) Laboratory evolution of laccase for substrate specificity. *J. Mol. Catal. B: Enzym.*, **62** (3–4), 230–234; (b) Pardo, I., Santiago, G., Gentili, P., Lucas, F., Monza, E., Medrano, F.J., Galli, C., Martinez, A.T., Guallar, V., and Camerero, S. (2016) Re-designing the substrate binding pocket of laccase for enhanced oxidation of sinapic acid. *Catal. Sci. Technol.* doi: 10.1039/x0xx00000x
 115. Mate, D.M., Gonzalez-Perez, D., Falk, M., Kittl, R., Pita, M., Lacey, D., Antonio, L., Ludwig, R., Shleev, S., and Alcalde, M. (2013) Blood tolerant laccase by directed evolution. *Chem. Biol.*, **20** (2), 223–231.
 116. Brissos, V., Ferreira, M., Grass, G., and Martins, L.O. (2015) Turning a hyperthermostable metallo-oxidase into a laccase by directed evolution. *ACS Catal.*, **5** (8), 4932–4941.
 117. Liu, Y., Ye, M., Lu, Y., Zhang, X., and Li, G. (2011) Improving the decolorization for textile dyes of a metagenome-derived alkaline laccase by directed evolution. *Appl. Microbiol. Biotechnol.*, **91** (3), 667–675.
 118. Molina-Espeja, P., Garcia-Ruiz, E., Gonzalez-Perez, D., Ullrich, R., Hofrichter, M., and Alcalde, M. (2014) Directed evolution of unspecific peroxxygenase from *Agrocyste aegerita*. *Appl. Environ. Microbiol.*, **80** (11), 3496–3507.
 119. Ostafe, R., Prodanovic, R., Nazor, J., and Fischer, R. (2014) Ultra-high-throughput screening method for the directed evolution of glucose oxidase. *Chem. Biol.*, **21** (3), 414–421.
 120. Parra, L.P., Agudo, R., and Reetz, M.T. (2013) Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBioChem*, **14** (17), 2301–2309.
 121. Zhang, Z.G., Roiban, G.D., Acevedo, J.P., Polyak, I., and Reetz, M.T. (2013) A new type of stereoselectivity in Baeyer-Villiger reactions: access to *E*- and *Z*-olefins. *Adv. Synth. Catal.*, **355** (1), 99–106.
 122. Zhang, Z.G., Lonsdale, R., Sanchis, J., and Reetz, M.T. (2014) Extreme synergistic mutational effects in the directed evolution of a Baeyer-Villiger monooxygenase as catalyst for asymmetric sulfoxidation. *J. Am. Chem. Soc.*, **136** (49), 17262–17272.
 123. Wu, S., Acevedo, J.P., and Reetz, M.T. (2010) Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.*, **107** (7), 2775–2780.
 124. Bastian, S., Liu, X., Meyerowitz, J.T., Snow, C.D., Chen, M.M.Y., and Arnold, F.H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metab. Eng.*, **13** (3), 345–352.
 125. Liu, X., Bastian, S., Snow, C.D., Brustad, E.M., Saleski, T.E., Xu, J.-H., Meinhold, P., and Arnold, F.H. (2013) Structure-guided engineering of *Lactococcus lactis* alcohol dehydrogenase lldha for improved conversion of isobutyraldehyde to isobutanol. *J. Biotechnol.*, **164** (2), 188–195.
 126. Spickermann, D., Hausmann, S., Degering, C., Schwaneberg, U., and Leggewie, C. (2014) Engineering of highly selective variants of *Parvibaculum lavamentivorans* alcohol dehydrogenase. *ChemBioChem*, **15** (14), 2050–2052.
 127. Agudo, R., Roiban, G.-D., and Reetz, M.T. (2013) Induced axial chirality in biocatalytic asymmetric ketone reduction. *J. Am. Chem. Soc.*, **135** (5), 1665–1668.
 128. Yeon, Y.J., Park, H.-Y., and Yoo, Y.J. (2015) Engineering substrate specificity of succinic semialdehyde reductase (AKR7A5) for efficient conversion of levulinic acid to 4-hydroxyvaleric acid. *J. Biotechnol.*, **210**, 38–43.

129. Jakoblinnert, A., van den Wittenboer, A., Shivange, A.V., Bocola, M., Heffele, L., Ansorge-Schumacher, M., and Schwaneberg, U. (2013) Design of an activity and stability improved carbonyl reductase from *Candida parapsilosis*. *J. Biotechnol.*, **165** (1), 52–62.
130. (a) Patel, J.M., Musa, M.M., Rodriguez, L., Sutton, D.A., Popik, V., and Phillips, R.S. (2014) Mutation of *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase at Trp-110 affects stereoselectivity of aromatic ketone reduction. *Org. Biomol. Chem.*, **12** (31), 5905–5910; (b) Li, H., Yang, Y., Zhu, D., Hua, L., and Kantardjieff, K. (2010) Highly enantioselective mutant carbonyl reductases created via structure-based site-saturation mutagenesis. *J. Org. Chem.*, **75**, 7559–7564.
131. Huang, L., Ma, H.-M., Yu, H.-L., and Xu, J.-H. (2014) Altering the substrate specificity of reductase Cgkr1 from *Candida glabrata* by protein engineering for bioreduction of aromatic α -keto esters. *Adv. Synth. Catal.*, **356** (9), 1943–1948.
132. Wang, N.-Q., Sun, J., Huang, J., and Wang, P. (2014) Cloning, expression, and directed evolution of carbonyl reductase from *Leifsonia xyli* HS0904 with enhanced catalytic efficiency. *Appl. Microbiol. Biotechnol.*, **98** (20), 8591–8601.
133. Moon, J. and Liu, Z.L. (2012) Engineered NADPH-dependent GRE2 from *Saccharomyces cerevisiae* by directed enzyme evolution enhances HMF reduction using additional cofactor NADPH. *Enzyme Microb. Technol.*, **50** (2), 115–120.
134. Steffler, F., Guterl, J.-K., and Sieber, V. (2013) Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in synthetic cascade biomanufacturing. *Enzyme Microb. Technol.*, **53** (5), 307–314.
135. Urano, N., Fukui, S., Kumashiro, S., Ishige, T., Kita, S., Sakamoto, K., Kataoka, M., and Shimizu, S. (2011) Directed evolution of an aminoalcohol dehydrogenase for efficient production of double chiral aminoalcohols. *J. Biosci. Bioeng.*, **111** (3), 266–271.
136. Xu, X., Chen, J., Wang, Q., Duan, C., Li, Y., Wang, R., and Yang, S. (2015) Mutagenesis of key residues in the binding center of L-aspartate- β -semialdehyde dehydrogenase from *Escherichia coli* enhances utilization of the cofactor NAD(H). *ChemBioChem*. doi: 10.1002/cbic.201500534
137. Zheng, H., Wang, X., Yomano, L.P., Geddes, R.D., Shanmugam, K.T., and Ingram, L.O. (2013) Improving *Escherichia coli* FucO for furfural tolerance by saturation mutagenesis of individual acid positions. *Appl. Environ. Microbiol.*, **79**, 3202–3208.
138. Brouk, M., Nov, Y., and Fishman, A. (2010) Improving biocatalyst performance by integrating statistical methods into protein engineering. *Appl. Environ. Microbiol.*, **76** (19), 6397–6403.
139. Jensen, C.N., Ali, S.T., Allen, M.J., and Grogan, G. (2013) Mutations of an NAD(P)H-dependent flavoprotein monooxygenase that influence cofactor promiscuity and enantioselectivity. *FEBS Open Bio*, **3**, 473–478.
140. Sönmez, B., Yanik-Yildirim, K.C., Wood, T.K., and Vardar-Schara, G. (2014) The role of substrate binding pocket residues phenylalanine 176 and phenylalanine 196 on *Pseudomonas* sp. Oxl1 toluene o-xylene monooxygenase activity and regiospecificity. *Biotechnol. Bioeng.*, **111** (8), 1506–1512.
141. Yanik-Yildirim, K.C. and Vardar-Schara, G. (2014) Saturation mutagenesis of *Bradyrhizobium* sp. BTAi1 toluene 4-monooxygenase at alpha-subunit residues proline 101, proline 103, and histidine 214 for regiospecific oxidation of aromatics. *Appl. Microbiol. Biotechnol.*, **98** (21), 8975–8986.
142. Abrahamson, M.J., Vázquez-Figueroa, E., Woodall, N.B., Moore, J.C., and Bommarius, A.S. (2012) Development of an amine dehydrogenase for synthesis of chiral amines. *Angew. Chem. Int. Ed.*, **51** (16), 3969–3972.
143. Ye, L.J., Toh, H.H., Yang, Y., Adams, J.P., Snajdrova, R., and Li, Z. (2015)

- Engineering of amine dehydrogenase for asymmetric reductive amination of ketone by evolving *Rhodococcus* phenylalanine dehydrogenase. *ACS Catal.*, **5** (2), 1119–1122.
144. Abrahamson, M.J., Wong, J.W., and Bommarius, A.S. (2013) The evolution of an amine dehydrogenase biocatalyst for the asymmetric production of chiral amines. *Adv. Synth. Catal.*, **355** (9), 1780–1786.
 145. Yasukawa, K., Nakano, S., and Asano, Y. (2014) Tailoring D-amino acid oxidase from the pig kidney to R-stereoselective amine oxidase and its use in the deracemization of α -methylbenzylamine. *Angew. Chem. Int. Ed.*, **53** (17), 4428–4431.
 146. Fernandes, P., Aldeborgh, H., Carlucci, L., Walsh, L., Wasserman, J., Zhou, E., Lefurgy, S.T., and Mundorff, E.C. (2015) Alteration of substrate specificity of alanine dehydrogenase. *Protein Eng. Des. Sel.*, **28** (2), 29–35.
 147. 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.
 148. Li, H. and Liao, J.C. (2014) Development of an NADPH-dependent homophenylalanine dehydrogenase by protein engineering. *ACS Synth. Biol.*, **3** (1), 13–20.
 149. Ji, D., Wang, L., Hou, S., Liu, W., Wang, J., Wang, Q., and Zhao, Z.K. (2011) Creation of bioorthogonal redox systems depending on nicotinamide flucytosine dinucleotide. *J. Am. Chem. Soc.*, **133** (51), 20857–20862.
 150. Molloy, S., Nikodinovic-Runic, J., Martin, L.B., Hartmann, H., Solano, F., Decker, H., and O'Connor, K.E. (2013) Engineering of a bacterial tyrosinase for improved catalytic efficiency towards D-tyrosine using random and site directed mutagenesis approaches. *Biotechnol. Bioeng.*, **110** (7), 1849–1857.
 151. Bernath-Levin, K., Shainsky, J., Sigawi, L., and Fishman, A. (2014) Directed evolution of nitrobenzene dioxygenase for the synthesis of the antioxidant hydroxytyrosol. *Appl. Microbiol. Biotechnol.*, **98** (11), 4975–4985.
 152. Cao, Y.-D., He, Y.-C., Li, H., Kai, G.-Y., Xu, J.-H., and Yu, H.-L. (2015) Efficient biosynthesis of rare natural product scopolamine using *E. coli* cells expressing a S14P/K97A mutant of hyoscyamine 6 β -hydroxylase AaH6H. *J. Biotechnol.*, **211**, 123–129.
 153. Chen, Z., Ma, Y., He, M., Ren, H., Zhou, S., Lai, D., Wang, Z., and Jiang, L. (2015) Semi-rational directed evolution of monoamine oxidase for kinetic resolution of rac-mexiletine. *Appl. Biochem. Biotechnol.*, **176** (8), 2267–2278.
 154. (a) Ghislieri, D., Green, A.P., Pontini, M., Willies, S.C., Rowles, I., Frank, A., Grogan, G., and Turner, N.J. (2013) Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products. *J. Am. Chem. Soc.*, **135** (29), 10863–10869; (b) Ghislieri, D., Houghton, D., Green, A.P., Willies, S.C., and Turner, N.J. (2013) Monoamine oxidase (MAO-N) catalyzed deracemization of tetrahydro- β -carbolines: substrate dependent switch in enantioselectivity. *ACS Catal.*, **3** (12), 2869–2872.
 155. Qian, Y., Zheng, J., and Lin, Z. (2013) Loop engineering of amadoriase II and mutational cooperativity. *Appl. Microbiol. Biotechnol.*, **97** (19), 8599–8607.
 156. (a) Brenna, E., Crotti, M., Gatti, F.G., Monti, D., Parmeggiani, F., Powell, R.W., Santangelo, S., and Stewart, J.D. (2015) Opposite enantioselectivity in the bioreduction of (Z)- β -aryl- β -cyanoacrylates mediated by the tryptophan 116 mutants of old yellow enzyme 1: synthetic approach to (R)- and (S)- β -aryl- γ -lactams. *Adv. Synth. Catal.*, **357** (8), 1849–1860; (b) Walton, A.Z., Conerly, W.C., Pompeu, Y., Sullican, B., and Stewart, J.D. (2011) Biocatalytic reductions of Baylis-Hillman adducts. *ACS Catal.*, **1**, 989–993.
 157. Walton, A.Z., Sullivan, B., Patterson-Orazem, A.C., and Stewart, J.D. (2014) Residues controlling facial

- selectivity in an alkene reductase and semirational alterations to create stereo-complementary variants. *ACS Catal.*, **4** (7), 2307–2318.
158. Horita, S., Kataoka, M., Kitamura, N., Nakagawa, T., Miyakawa, T., Ohtsuka, J., Nagata, K., Shimizu, S., and Tanokura, M. (2015) An engineered old yellow enzyme that enables efficient synthesis of (4R,6R)-actinol in a one-pot reduction system. *ChemBioChem*, **16**, 440–445.
 159. Hwang, B.-Y. (2012) Directed evolution of cutinase using in vitro compartmentalization. *Biotechnol. Bioprocess Eng.*, **17** (3), 500–505.
 160. Swe, P.M., Copp, J.N., Green, L.K., Guise, C.P., Mowday, A.M., Smaill, J.B., Patterson, A.V., and Ackerley, D.F. (2012) Targeted mutagenesis of the *Vibrio fischeri* flavin reductase FRase I to improve activation of the anticancer prodrug CB1954. *Biochem. Pharmacol.*, **84** (6), 775–783.
 161. Bingham, A.S., Smith, P.R., and Swartz, J.R. (2012) Evolution of an [FeFe] hydrogenase with decreased oxygen sensitivity. *Int. J. Hydrogen Energy*, **37** (3), 2965–2976.
 162. Morra, S., Giraud, A., Di Nardo, G., King, P.W., Gilardi, G., and Valetti, F. (2012) Site saturation mutagenesis demonstrates a central role for cysteine 298 as proton donor to the catalytic site in CaHydA [FeFe]-hydrogenase. *PLoS One*, **7** (10), e48400.
 163. Baxter, S., Royer, S., Grogan, G., Brown, F., Holt-Tiffin, K.E., Taylor, I.N., Fotheringham, I.G., and Campopiano, D.J. (2012) An improved race-mase/acylase biotransformation for the preparation of enantiomerically pure amino acids. *J. Am. Chem. Soc.*, **134** (47), 19310–19313.
 164. Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., and Hughes, G.J. (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, **329** (5989), 305–309.
 165. Deszcz, D., Affaticati, P., Ladkau, N., Gegel, A., Ward, J.M., Hailes, H.C., and Dalby, P.A. (2015) Single active-site mutants are sufficient to enhance serine: pyruvate α -transaminase activity in an ω -transaminase. *FEBS J.*, **282** (13), 2512–2526.
 166. van Leeuwen, J.G.E., Wijma, H.J., Floor, R.J., van der Laan, J.-M., and Janssen, D.B. (2012) Directed evolution strategies for enantiocomplementary haloalkane dehalogenases: from chemical waste to enantiopure building blocks. *ChemBioChem*, **13** (1), 137–148.
 167. Guo, C., Chen, Y., Zheng, Y., Zhang, W., Tao, Y., Feng, J., and Tang, L. (2015) Exploring the enantioselective mechanism of halohydrin dehalogenase from *Agrobacterium radiobacter* AD1 by iterative saturation mutagenesis. *Appl. Environ. Microbiol.*, **81** (8), 2919–2926.
 168. Tang, L., Zhu, X., Zheng, H., Jiang, R., and Majerik Elenkov, M. (2012) Key residues for controlling enantioselectivity of halohydrin dehalogenase from *Arthrobacter* sp. strain AD2, revealed by structure-guided directed evolution. *Appl. Environ. Microbiol.*, **78** (8), 2631–2637.
 169. Payne, J.T., Poor, C.B., and Lewis, J.C. (2015) Directed evolution of RebH for site-selective halogenation of large biologically active molecules. *Angew. Chem. Int. Ed.*, **54** (14), 4226–4230.
 170. Kourist, R., Miyauchi, Y., Uemura, D., and Miyamoto, K. (2011) Engineering the promiscuous racemase activity of an arylmalonate decarboxylase. *Chem. Eur. J.*, **17** (2), 557–563.
 171. Gaßmeyer, S.K., Yoshikawa, H., Enoki, J., Hülsemann, N., Stoll, R., Miyamoto, K., and Kourist, R. (2015) STD-NMR-based protein engineering of the unique arylpropionate-racemase AMDase G74C. *ChemBioChem*, **16** (13), 1943–1949.
 172. Taylor, J.L., Price, J.E., and Toney, M.D. (2015) Directed evolution of the substrate specificity of dialkylglycine decarboxylase. *Biochim. Biophys. Acta*, **1854** (2), 146–155.

173. Wang, C., Zhang, K., Zhongjun, C., Cai, H., Honggui, W., and Ouyang, P. (2015) Directed evolution and mutagenesis of lysine decarboxylase from *Hafnia alvei* AS1.1009 to improve its activity toward efficient cadaverine production. *Biotechnol. Bioprocess Eng.*, **20** (3), 439–446.
174. Campeotto, I., Bolt, A.H., Harman, T.A., Dennis, C., Trinh, C.H., Phillips, S.E.V., Nelson, A., Pearson, A.R., and Berry, A. (2010) Structural insights into substrate specificity in variants of *N*-acetylneuraminic acid lyase produced by directed evolution. *J. Mol. Biol.*, **404** (1), 56–69.
175. Cheriyan, M., Toone, E.J., and Fierke, C.A. (2012) Improving upon nature: active site remodeling produces highly efficient aldolase activity toward hydrophobic electrophilic substrates. *Biochemistry*, **51** (8), 1658–1668.
176. Cheriyan, M., Walters, M.J., Kang, B.D., Anzaldi, L.L., Toone, E.J., and Fierke, C.A. (2011) Directed evolution of a pyruvate aldolase to recognize a long chain acyl substrate. *Bioorg. Med. Chem.*, **19** (21), 6447–6453.
177. Chou, C.-Y., Ko, T.-P., Wu, K.-J., Huang, K.-F., Lin, C.-H., and Wong, C.-H. (2011) Modulation of substrate specificities of D-sialic acid aldolase through single mutations at Val-251. *J. Biol. Chem.*, **286**, 14057–14064.
178. Ranoux, A., Karmee, S.K., Jin, J., Bhaduri, A., Caiazzo, A., Arends, I.W.C.E., and Hanefeld, U. (2012) Enhancement of the substrate scope of transketolase. *ChemBioChem*, **13** (13), 1921–1931.
179. Yi, D., Saravanan, T., Devamani, T., Charmantray, F., Hecquet, L., and Fessner, W.-D. (2015) A thermostable transketolase evolved for aliphatic aldehyde acceptors. *Chem. Commun.*, **51** (3), 480–483.
180. Cazares, A., Galman, J.L., Crago, L.G., Smith, M.E.B., Strafford, J., Rios-Solis, L., Lye, G.J., Dalby, P.A., and Hailes, H.C. (2010) Non- α -hydroxylated aldehydes with evolved transketolase enzymes. *Org. Biomol. Chem.*, **8** (6), 1301–1309.
181. Payongsri, P., Steadman, D., Hailes, H.C., and Dalby, P.A. (2015) Second generation engineering of transketolase for polar aromatic aldehyde substrates. *Enzyme Microb. Technol.*, **71**, 45–52.
182. Schneider, S., Gutiérrez, M., Sandalova, T., Schneider, G., Clapés, P., Sprenger, G.A., and Samland, A.K. (2010) Redesigning the active site of transaldolase talb from *Escherichia coli*: new variants with improved affinity towards nonphosphorylated substrates. *ChemBioChem*, **11** (5), 681–690.
183. Chen, S.-H., Hwang, D.-R., Chen, G.-H., Hsu, N.-S., Wu, Y.-T., Li, T.-L., and Wong, C.-H. (2012) Engineering transaldolase in *Pichia stipitis* to improve bioethanol production. *ACS Chem. Biol.*, **7** (3), 481–486.
184. Westphal, R., Waltzer, S., Mackfeld, U., Widmann, M., Pleiss, J., Beigi, M., Muller, M., Rother, D., and Pohl, M. (2013) (*S*)-selective mend variants from *Escherichia coli* provide access to new functionalized chiral α -hydroxy ketones. *Chem. Commun.*, **49** (20), 2061–2063.
185. (a) Kawahara, N. and Asano, Y. (2015) Mutagenesis of an Asn156 residue in a surface region of *S*-selective hydroxynitrile lyase from *Baliospermum montanum* enhances catalytic efficiency and enantioselectivity. *ChemBioChem*, **16** (13), 1891–1895; (b) Wiedner, R., Kothbauer, B., Pavkov-Keller, T., Gruber-Khadjawi, M., Gruber, K., Schwab, H., and Steiner, K. (2015) Improving the properties of bacterial R-selective hydroxynitrile lyases for industrial applications. *ChemCatChem*, **7**, 325–332.
186. Raj, H., Szymański, W., de Villiers, J., Rozeboom, H.J., Veetil, V.P., Reis, C.R., de Villiers, M., Dekker, F.J., de Wildeman, S., Quax, W.J., Thunnissen, A.-M.W.H., Feringa, B.L., Janssen, D.B., and Poelarends, G.J. (2012) Engineering methylaspartate ammonia lyase for the asymmetric synthesis of unnatural amino acids. *Nat. Chem.*, **4** (6), 478–484.
187. Wu, B., Szymański, W., Wybenga, G.G., Heberling, M.M., Bartsch, S., de Wildeman, S., Poelarends, G.J., Feringa,

- B.L., Dijkstra, B.W., and Janssen, D.B. (2012) Mechanism-inspired engineering of phenylalanine aminomutase for enhanced β -regioselective asymmetric amination of cinnamates. *Angew. Chem. Int. Ed.*, **51** (2), 482–486.
188. Vogel, A., Schmiedel, R., Hofmann, U., Gruber, K., and Zangger, K. (2014) Converting aspartase into a β -amino acid lyase by cluster screening. *ChemCatChem*, **6** (4), 965–968.
189. Kaya, E., Vrabel, M., Deiml, C., Prill, S., Fluxa, V.S., and Carell, T. (2012) A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. *Angew. Chem. Int. Ed.*, **51** (18), 4466–4469.
190. Zhang, W., Dourado, D.F.A.R., and Mannervik, B. (2015) Evolution of the active site of human glutathione transferase A2-2 for enhanced activity with dietary isothiocyanates. *Biochim. Biophys. Acta*, **1850** (4), 742–749.
191. Lu, W.-C., Levy, M., Kincaid, R., and Ellington, A.D. (2014) Directed evolution of the substrate specificity of biotin ligase. *Biotechnol. Bioeng.*, **111** (6), 1071–1081.
192. Amar, D., Berger, I., Amara, N., Tafa, G., Meijler, M.M., and Aharoni, A. (2012) The transition of human estrogen sulfotransferase from generalist to specialist using directed enzyme evolution. *J. Mol. Biol.*, **416** (1), 21–32.
193. Kramer, J.R. and Matsumura, I. (2013) Directed evolution of aminoglycoside phosphotransferase (3') type IIIa variants that inactivate amikacin but impose significant fitness costs. *PLoS One*, **8** (10), e76687.
194. Yongjun, C., Wei, B., Shujun, J., Meizhi, W., Yan, J., Yan, Y., Zhongliang, Z., and Goulin, Z. (2011) Directed evolution improves the fibrinolytic activity of nattokinase from *Bacillus natto*. *FEMS Microbiol. Lett.*, **325** (2), 155–161.
195. Ibach, J., Dietrich, L., Koopmans, K.R.M., Nöbel, N., Skoupi, M., and Brakmann, S. (2013) Identification of a T7 RNA polymerase variant that permits the enzymatic synthesis of fully 2'-O-methyl-modified RNA. *J. Biotechnol.*, **167** (3), 287–295.
196. Minnihan, E.C., Young, D.D., Schultz, P.G., and Stubbe, J. (2011) Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved polyspecific aminoacyl-tRNA synthetase. *J. Am. Chem. Soc.*, **133** (40), 15942–15945.
197. Uyeda, A., Watanabe, T., Kato, Y., Watanabe, H., Yomo, T., Hohsaka, T., and Matsuura, T. (2015) Liposome-based in vitro evolution of aminoacyl-tRNA synthetase for enhanced pyrrolysine derivative incorporation. *ChemBioChem*, **16** (12), 1797–1802.
198. Guo, L.-T., Wang, Y.-S., Nakamura, A., Eiler, D., Kavan, J.M., Wong, M., Kiessling, L.L., Steitz, T.A., O'Donoghue, P., and Söll, D. (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **111** (47), 16724–16729.
199. Obeid, S., Schnur, A., Gloeckner, C., Blatter, N., Welte, W., Diederichs, K., and Marx, A. (2011) Learning from directed evolution: *Thermus aquaticus* DNA polymerase mutants with translesion synthesis activity. *ChemBioChem*, **12** (10), 1574–1580.
200. 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**, 5288–5294.
201. Ji, J., Fan, K., Tian, X., Zhang, X., Zhang, Y., and Yang, K. (2012) Iterative combinatorial mutagenesis as an effective strategy for generation of deacetoxycephalosporin C synthase with improved activity toward penicillin G. *Appl. Environ. Microbiol.*, **78** (21), 7809–7812.
202. Wu, X.-B., Tian, X.-Y., Ji, J.-J., Wu, W.-B., Fan, K.-Q., and Yang, K.-Q. (2011) Saturation mutagenesis of *Acremonium chrysogenum* deacetoxy/deacetylcephalosporin C synthase R308 site confirms its role in controlling substrate specificity. *Biotechnol. Lett.*, **33** (4), 805–812.
203. Furubayashi, M., Saito, K., and Umeno, D. (2014) Evolutionary analysis of the

- functional plasticity of *Staphylococcus aureus* C₃₀ carotenoid synthase. *J. Biosci. Bioeng.*, **117** (4), 431–436.
204. (a) Sundermann, U., Bravo-Rodriguez, K., Klopries, S., Kushnir, S., Gomez, H., Sanchez-Garcia, E., and Schulz, F. (2013) Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem. Biol.*, **8** (2), 443–450; (b) Kushnir, S., Sundermann, U., Yahiaoui, S., Brockmeyer, A., Janning, P., and Schulz, F. (2012) Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library. *Angew. Chem. Int. Ed.*, **51** (42), 10664–10669.
 205. Koryakina, I. and Williams, G.J. (2011) Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *ChemBioChem*, **12** (15), 2289–2293.
 206. Wasiele, A.A., Baas, B.-J., Zandvoort, E., Quax, W.J., and Poelarends, G.J. (2012) Dehalogenation of an anthropogenic compound by an engineered variant of the mouse cytokine macrophage migration inhibitor factor. *ChemBioChem*, **13**, 1270–1273.
 207. Soskine, M., Biesemans, A., De Maeyer, M., and Maglia, G. (2013) Tuning the size and properties of ClyA nanopores assisted by directed evolution. *J. Am. Chem. Soc.*, **135** (36), 13456–13463.
 208. (a) Bhuiya, M.W., Liu, C.J. (2010). Engineering Monolignol 4-O-Methyltransferases to Modulate Lignin Biosynthesis. *J. Biol. Chem.*, **285** (1), 277–285; (b) Cai, Y., Zhang, K., Kim, H., Hou, G., Zhang, X., Yang, H., Feng, H., Miller, L., Ralph, J., and Liu, C.J. (2016). Enhancing digestibility and ethanol yield of *Populus* wood via expression of an engineered monolignol 4-O-methyltransferase. *Nat. Commun.*, **7**, 11989.
 209. Fujii, I. and Tsumuraya, T. (2014) Directed evolution of hydrolytic antibodies in phage-displayed combinatorial libraries. *Chem. Lett.*, 272–280.
 210. Lee, C.-H., Lee, Y.-C., Liang, H.-H., Leu, S.J., Lin, L.T., Chiang, J.-R., and Yang, Y.-Y. (2016) Antibodies against venom of the snake *Deinagkistrodon acutus*. *Appl. Environ. Microbiol.*, **82**, 71–80.

6

Directed Evolution of Enzyme Robustness

6.1

Introduction

As outlined in Chapter 1, increasing the robustness of enzymes under operating conditions has been the goal of protein engineers for more than three decades, simply because this property is a prerequisite for applications in biotechnology [1]. One of the earliest examples of directed evolution of protein robustness was the application of saturation mutagenesis (SM) at an oxidatively sensitive methionine residue of a protease, leading to two different mutants (Ser or Ala mutants, respectively) with dramatically enhanced stability in the presence of H_2O_2 [2]. Subsequently, numerous protein engineering studies appeared describing enhanced resistance to hostile organic solvents, thermostability, and/or oxidative stability. Most of these early efforts involved the application of site-directed mutagenesis based on rational design, a research area that has been summarized in review articles [3]. Since the degree of success was not uniformly high, many protein engineers (re)turned to directed evolution, especially in view of the fact that focused randomization had been shown to be successful [2]. In the 1990s, numerous studies appeared in which error-prone polymerase chain reaction (epPCR), SM at hot spots as revealed by such random mutagenesis, and/or DNA shuffling was applied [4a,b]. This development has continued to this day [4c–g]. The use of protein display systems such as yeast display for enhanced enzyme stability has been reviewed [4f].

Two other strategies also deserve mention. In one approach, a thermophilic error-prone strain from *Geobacillus kaustophilus* was engineered by deleting functional DNA repair genes, and then used in directed evolution of protein thermostabilization [4h]. This concept still needs to be tested on a broad basis. In a different development, *Bacillus subtilis* spore display of a laccase under extreme conditions of high concentrations of hostile organic solvents proved to be successful [5], but it is currently not clear how general this approach is.

In many, but not all cases, robustness increased without compromising activity. It is essentially impossible to compare the viability of the various strategies because different enzymes as well as different robustness indices were used. Many researchers reported melting temperatures (T_m), T_{50} values (temperature at which 50% of enzyme activity is lost after heating for a given period of time),

or half-life at a defined temperature. Moreover, different assessments result depending upon whether purified, partially purified enzymes, whole cells, or lysates are used in such measurements because robustness also depends upon the interaction of enzymes with other proteins and different biomolecules. In some studies, a correlation between enhanced thermostability and resistance to hostile organic solvents was found, but in other cases this aspect was not addressed. Strategies for stabilizing enzymes in organic solvents have been summarized, including protein engineering, chemical modification, immobilization, and the use of additives [6].

Lessons were learned by studying the origin of thermal stability of thermophilic enzymes [7], which indicated higher rigidity. Therefore, a central hypothesis was to increase the rigidity of thermally sensitive enzymes by introducing structure-guided point mutations that counteract the natural flexibility. In some of the early studies on directed evolution at least a few of the discovered point mutations were interpreted on a molecular level. Such factors as newly formed salt bridges, H-bonding interactions, disulfide bridges, introduction of proline for reducing conformational flexibility, and/or interior hydrophobic packing effects were invoked [3, 4]. However, many of the observed mutations could not be rationalized. The likely possibility that some of the accumulated point mutations are actually superfluous was not considered. In 2005, comprehensive review articles appeared, showing that epPCR and/or DNA shuffling, which require no structural data, were the most often used mutagenesis methods during the period 1995–2005 [4c–e]. These reviews also stress the importance of kinetic versus thermodynamic stability. Since then, a trend toward more “rational” strategies in laboratory evolution has emerged in which structural, sequence, and/or computational information is cleverly exploited. The sequence data-based consensus approach, first developed for stabilizing antibodies [8], has been applied in the quest to enhance enzyme robustness for a long time [9]. It is still used today, often augmented by structural data, which then constitutes a powerful strategy (see following text). One of the central themes has been and continues to be the controlled rigidification of flexible sites or regions in a protein [10].

Another early approach is the PROSIDE method, according to which large designed mutant libraries generated by random mutagenesis are evaluated by Q7 selection procedures based on phage or ribosome display systems [11]. In the absence of guiding information, epPCR and DNA shuffling are recommended even today. Several recent reviews cover all current approaches to protein engineering of enzyme robustness and consider the following properties [6, 10, 12].

- Kinetic and thermodynamic thermostability
- Resistance to hostile organic solvents including ionic liquids
- Oxidative stability
- Tolerance to different pH ranges.

A very different method for enhancing the robustness of proteins does not involve any changes in the sequence of the target enzyme and, therefore,

constitutes a completely different approach. Accordingly, insertional fusion of enzymes to thermophilic host proteins is the underlying strategy [13a]. Other techniques include enzyme immobilization [13b], an area which is not treated in this monograph.

Although enormous progress in developing advanced methods for protein stabilization has been made in recent years, it is not evident which approach promises to be truly general and most effective. Thus, the situation is quite different from evolving stereo- and regioselective enzymes in which SM at sites lining the binding pocket has emerged as the method of choice (see previous chapters). Several computational tools useful as guides when applying directed evolution for enhanced robustness have been developed (see Section 6.6). In the sections that follow, selected recent examples illustrating methodology development are critically analyzed. For the assessment of other case studies, the reader is referred to an excellent review [12a].

6.2

Application of epPCR and DNA Shuffling

If little or no structural or helpful sequence information is available, then epPCR, DNA shuffling, or mutator strains constitute the most important options for enhancing protein robustness. Even if such information is available, studies continue to appear in which this kind of data is ignored. Some degree of improvement can always be expected in laboratory evolution, irrespective of the mutagenesis strategy or method. In an interesting study, a mutant of the insect α -carboxylesterase from *Lucilia cuprina*, which had been evolved earlier to accept organophosphate insecticides, was subjected to directed evolution in order to increase its thermostability [14]. Four rounds of epPCR were traversed and assessed according to the following simple procedure. About 100 000 random variants were plated onto agar plates and then replicated on filter paper, which was subsequently incubated at defined temperatures for 1 h. Then the filter paper was immersed in a solution containing 2-naphthyl acetate. Colonies showing the highest activity were selected and pooled for the subsequent epPCR cycle. The final variant contained six new point mutations and displayed notable esterase activity after heat treatment at 54 °C for 1 h, whereas wild-type (WT) showed no activity under the same conditions. Three of the mutations occurred on the enzyme surface (Met364Leu, Lys530Glu, and Asp55Gly), the others were in the interior (Ile419Phe, Ala472Thr, and Ile505Thr). The structural basis for this notable improvement was partially elucidated with the realization that Ile419Phe fills space in a hydrophobic cavity [14] (Figure 6.1). The result of filling hydrophobic cavities with concomitant increase in thermal stability is a well-known mutational effect [15]. However, the interpretation of the other mutations proved to be difficult. Although an enormously high number of transformants had to be screened, the assay was quite simple to perform. This type of screening procedure

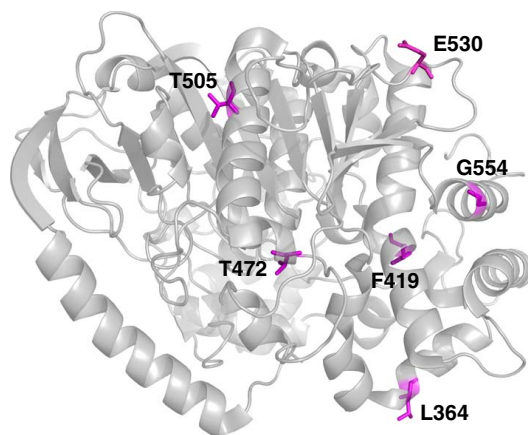


Figure 6.1 The six stabilizing mutations evolved in the insect α -carboxylesterase from *Lucilia cuprina* [14]. (Adapted from Jackson *et al.* [14]. Reproduced with permission of National Academy of Sciences.)

should be applicable to all esterases and lipases, irrespective of the mutagenesis method.

Another study also involves the thermostabilization of an esterase, in this case the feruloyl esterase from *Aspergillus niger*, an enzyme that is instrumental in the selective degradation of lignocellulose [16]. In an initial attempt using the PoP-MuSiC algorithm [17] as a computational guide, a slightly improved variant with two point mutations was generated, but the degree of thermostabilization was far from meeting industrial standards. Considerably higher robustness was needed in order to decrease the likelihood of microbial contamination, support the disorganization of the raw material (lignocellulose in biomass), and promote enzyme penetration. Using the previously generated double mutant as a template, epPCR was applied with formation of a 40 000-membered library. About 10 000 colonies were screened by a UV/vis-plate reader, which monitors the formation of 2-chloro-4-nitrophenolate (425 nm) formed upon hydrolysis of 2-chloro-4-nitrophenyl ferulate [16c] in a high-throughput manner. A mutant with 12 additional mutations was identified showing 80% residual activity after heat treatment at 90 °C for 15 min. It was also shown that ferulic acid was smoothly released from steam-exploded corn stalk as typical biomass (Figure 6.2).

In a different study, epPCR was likewise used in a successful attempt to enhance the robustness of the homodimeric haloperoxidase from *Streptomyces aureofaciens* [18]. A 1000-membered library was assayed in the halogenation of chlorodimedone. It was discovered that the best mutants show moderately enhanced thermostability, but increased resistance to a variety of different solvents such as 1-butanol was also detected (Table 6.1). Mainly, surface mutations occurred. It was deduced that the α -helix was stabilized and the interaction between subunits contribute to the improvements. Perhaps, screening a greater portion of sequence space using larger libraries would provide even better results.

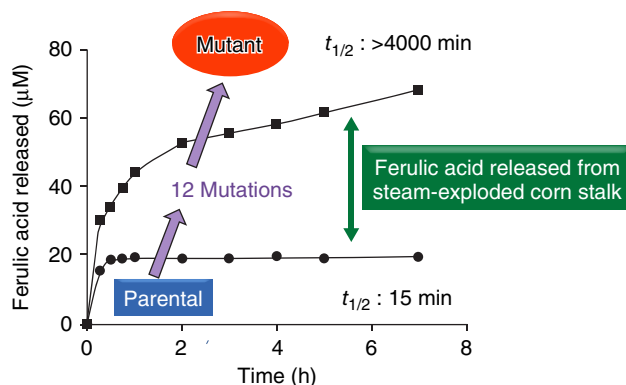


Figure 6.2 Catalytic performance of a variant of the feruloyl esterase from *Aspergillus niger* in the degradation of steam-exploded corn stalk as biomass [16a]. (Adapted and reproduced from Zhang *et al.* [16a]. Reproduced with permission of Elsevier.)

Table 6.1 Specific activity, T_{50} , and 1-propanol stability of WT and mutant BPO-A1 haloperoxidases [18].

BPO-A1 ^{a)}	Specific activity (kU g^{-1})	T_{50} ($^{\circ}\text{C}$)	Residual activity in 40% (v/v) 1-propanol (%) ^{b)}
Wild-type	56.4	80.6	25.9 ± 3.0
Mutant HT177	59.6	82.0	55.4 ± 4.1
Mutant R114H	50.3	81.9	63.2 ± 4.3
Mutant N146H	78.7	80.4	20.3 ± 5.3
Mutant HT507	101.7	84.5	15.2 ± 3.2
Mutant G106S	99.4	81.1	33.5 ± 2.4
Mutant V148I	117.5	83.7	7.60 ± 1.2

a) Mutants HT177 and HT507 were obtained by directed evolution. Mutants R114H, N146H, G106S, and V148I having single-amino-acid-substitutions were constructed by site-directed mutagenesis.

b) The data represent the averages of two independent experiments.

Source: Yamada *et al.* [18]. Reproduced with permission of Elsevier.

Several other studies based on random mutagenesis or DNA shuffling have appeared [19]. For example, in order to boost the robustness of the haloalkane dehalogenase from *Rhodococcus rhodochrous* toward heat and organic cosolvents, epPCR was first performed, leading to the identification of a stabilizing mutation at the access tunnel [19c]. Therefore, the access tunnel was explored mutationally more closely by SM, which provided the best mutant showing an increase in melting temperature of 19°C and a notable resistance to DMSO as cosolvent. Iterative saturation mutagenesis (ISM) was not tested, but could provide even better variants (see also Section 6.3 featuring the SM-based B-FIT approach).

In most studies directed toward enhancing the robustness of proteins, only one property such as thermostability was targeted. A particularly challenging task in directed evolution is to improve several properties (simultaneously), for example, thermostability, alkaline stability, and activity, all of which are needed for practical applications. The example featured here concerns the directed evolution study of the xylanase from *B. subtilis*, in which thermostability, stability at pH8, and activity were all improved [20]. In an earlier study, the authors had performed two rounds of epPCR, assaying in each case about 12 000 transformants with a pre-screen based on halo formation on solid agar containing Congo Red [20b]. This was followed by alternating DNA shuffling and further epPCR cycles, while varying the “selection pressure.” The workflow is shown in Figure 6.3.

The best variant was found to have eight point mutations, showing a temperature optimum at 80 °C under alkaline conditions (pH8) and a threefold increase in specific activity. The total number of transformants screened in the overall process was not reported, but this work shows that with a well-designed strategy and a great deal of experimental work the difficult problem of mastering several enzyme properties for real (industrial) applications can be solved. Several other

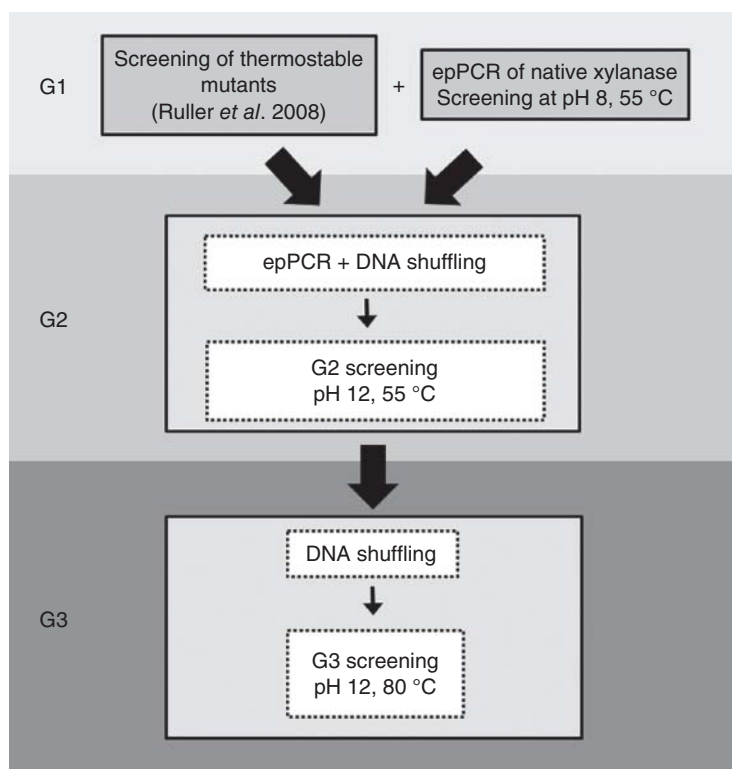


Figure 6.3 Workflow in the directed evolution of a highly improved xylanase [20]. (Ruller et al. [20]. Reproduced with permission of Oxford University Press.)

approaches to xylanase optimization (although not always addressing several parameters) have been reported, generally also relying on epPCR and DNA shuffling [21].

An example of shifting the pH optimum to suit industrial applications while maintaining (not increasing) high thermostability also deserves mention [22]. In order to enhance the activity of the thermostable β -1,3-1,4-glucanase from *Pae-cilomyces thermophila* at acidic pH where it is needed in such industrial applications as beer brewing and animal feed, the combination of epPCR and DNA shuffling was applied. A variant characterized by three point mutations shifted the optimum pH from 7.0 to 5.0 with no trade-off in activity or thermostability. Many practical applications can be envisioned.

The traditional “trick” to perform SM on the residues identified by epPCR for further thermostabilization [23a] or enhanced stereoselectivity [23b] is still used today, although many of these residues may not at all be true hot spots (superfluous mutations). The current literature abounds with further thermostability studies in which epPCR and/or DNA shuffling were applied; only a few representative studies are cited here [19a,e, 24]. In some of these studies hot spots identified by epPCR or DNA shuffling were subjected to SM for further improvement in enzyme robustness. For example, in an illuminating study of the lipase from *B. subtilis* (Lip A) using recursive epPCR followed by SM, the T_m value was boosted to 78 °C, the optimal activity temperature of the variant being 65 °C [24j]. Undesired aggregation upon heat treatment did not occur. X-ray structures proved to be helpful in interpreting the results. Another prominent example pertains to the thermostabilization of an aldehyde dehydrogenase by epPCR and SM at hot spots for application in synthetic cascade biomanufacturing [19e].

Lip A was also subjected to SM without performing any epPCR, specifically by randomizing each of the 181 residues separately [25]. The purpose was to enhance detergent tolerance, which proved to be successful. The most robust variants remained active as catalysts in the presence of anionic, cationic, zwitterionic, and nonionic surfactants.

6.3

B-FIT Approach

When turning away from “blind” directed evolution based on epPCR and/or DNA shuffling for enhancing thermostability, the B-FIT approach utilizing SM and, optionally, ISM is one of several logical options [26]. The basic idea is to utilize SM for rigidifying certain regions in a protein, which requires a criterion for choosing appropriate randomization sites. The choice is made based on the B-factor values available from X-ray data. In a first step, the average B-factor values of all residues are obtained with the help of the computer aid B-FITTER, available free of charge from the author (<http://www.kofo.mpg.de/en/research/biocatalysis>). It automatically lists the average B-factor values of all residues in a protein, starting from the highest to the lowest values. Residues exhibiting high

average B-factors indicate positions of flexibility. Therefore, six to eight such residues are normally chosen for SM. As in the case of applying SM for enhancing or inverting stereoselectivity, statistical factors concerning library coverage and oversampling need to be considered (Chapters 3 and 4). Typically, residues exhibiting the highest average B-factors are grouped into two- or three-residue randomization sites, followed by SM and ISM. It should be mentioned that B-FIT is not the only way to introduce rigidity; computational techniques are options as well (see Section 6.6).

In the first case study, B-FIT- based ISM was applied to the lipase from *B. subtilis* (Lip A), leading to a variant characterized by five point mutations with significantly increased apparent thermostability [26]. Complete deconvolution allowed the construction of a fitness pathway landscape with $5! = 120$ trajectories leading from WT Lip A to the best evolved mutant. A theoretical analysis of the results revealed pronounced cooperative mutational effects occurring on the surface of the lipase in the form of a continuous H-bond network [27]. Biophysical and biochemical characterization including protein NMR spectroscopic studies, circular dichroism, X-ray structural analyses, and combining thermal inactivation profiles uncovered an interesting surface effect that results from the mutations [28]: Upon heat treatment, WT Lip A undergoes undesired irreversible aggregation with precipitation, whereas the “robust” variant aggregates at higher temperatures to a significantly lower extent. Reduced aggregation and precipitation of the unfolding intermediates were shown to be responsible for activity retention at higher temperatures. The NMR spectroscopic investigation of the native and the thermally treated ^{15}N -labeled Lip A variant showed that it recovers almost all of the natural conformation after heat treatment and cooling to ambient temperature, in agreement with activity recovering. More specifically, after heating and cooling the variant shows a conformation that is almost identical to the heat untreated enzyme, as identified by peaks in $1\text{D}^1\text{H}$ and $2\text{D}[^{15}\text{N}, ^1\text{H}]\text{-HSQC}$ spectra (Figure 6.4) [28].

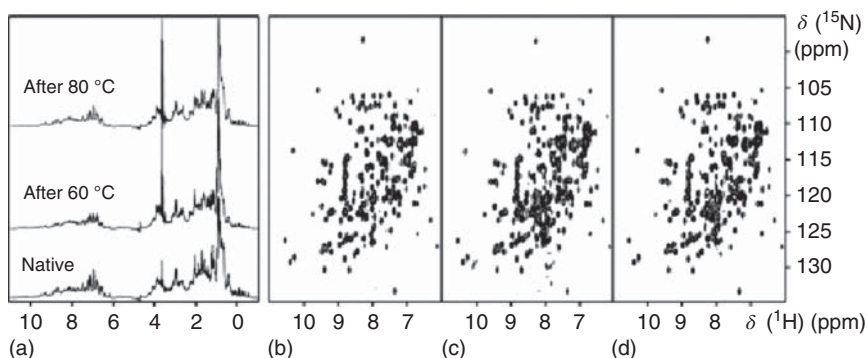


Figure 6.4 NMR spectra recorded for native and thermally treated ^{15}N -labeled Lipase A mutant XI. (a) $1\text{D}^1\text{H}$ spectra. (b–d) $2\text{D}[^{15}\text{N}, ^1\text{H}]\text{-HSQC}$ spectra of mutant XI Lipase A: (b) native; (c) recovered after 60 °C treatment; (d) recovered after 80 °C treatment [28].

Furthermore, it was discovered that not 100% of the enzyme is recovered in its native fold. Interestingly, a small amount of soluble unfolded and/or degraded species was seen in the spectra, which indicates that the fraction of unfolded/degraded and also irreversibly aggregated (but still soluble) species is very small [28]. In a different study, it was demonstrated that the same evolved Lip A variant is also tolerant to hostile organic solvents such as acetonitrile, dimethylsulfoxide, and dimethylformamide [29].

The possible role of irreversible aggregation and precipitation propensity of thermally unfolded states of WT enzymes has not been addressed very often when performing directed evolution [24j, 29], a subject that needs further attention in the future. In fact, when studying only T_{50} values, the true cause of enzyme improvement may not be fully evident, a caution that pertains to essentially all directed evolution studies.

In another B-FIT- based ISM study, the thermostability of the epoxide hydrolase from *A. niger* was increased to a significant degree [30]. Eight of the 356 residues showing the highest calculated average B-factor values as well as four crystallographically unresolved and possibly likewise flexible residues were considered for SM. The 12 residues were grouped into six two-residue randomization sites A, B, C, D, E, and F for NDT-based SM. Several but not all theoretically possible upward ISM pathways were explored, leading to the best variant with a 21 °C increase in the T_{50}^{60} value, an 80-fold improvement in half-life at 60 °C, and a 44 kcal mol⁻¹ improvement in inactivation energy. ISM exploration also provided seven other variants characterized by 10–14 °C increases in T_{50}^{60} values, 20–30-fold increase in half-lives at 60 °C and 15–20 kcal mol⁻¹ elevations in inactivation energy. In several cases, a given ISM library failed to contain improved variants, which means that a local minimum on the fitness landscape (dead end) was encountered. Such an event is not at all rare in the directed evolution in general, irrespective of the mutagenesis method. In such situations a simple strategy to escape from these local minima was applied [31], namely by utilizing a nonimproved or even inferior mutant in the library as a template for the subsequent ISM step. In a curiosity-driven experiment, a neutral variant and an inferior one were used in otherwise the same ISM scheme. Importantly, the inferior template led to better results (Figure 6.5) [30].

Another noteworthy example of B-FIT concerns directed evolution of the cellulose from *Trichoderma reesei*, a key enzyme for economically saccharifying cellulosic biomass in the production of biofuels [32]. Using the B-FITTER computer aid (<http://www.kofo.mpg.de/en/research/biocatalysis>), 10 residues with high B-factors were identified and grouped into seven randomization sites A, B, C, D, E, F, and G (Figure 6.6). Amino acids with high B-factors that are spatially close to the N- or C-termini, disulfide bridges, or N-glycosylation sites were not considered in this analysis. A total of 11 000 transformants were screened, which led to about 500 variants showing higher activity in the reaction of carboxymethyl cellulose following pre-incubation at 50 °C for 45 min. From this collection, 70 variants were sequenced, many of which showed notably enhanced robustness. Combining point mutations led to further improvements in some cases. The triple

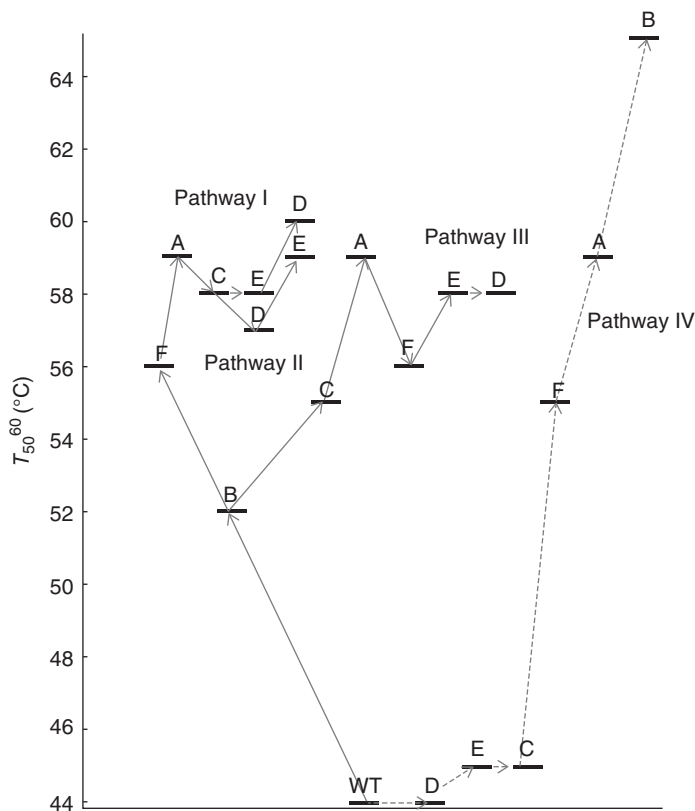


Figure 6.5 Results of limited ISM exploration starting from the best mutant, GUY-003 (site B), and the worst mutant, GUY-007, in the initial round of saturation mutagenesis at sites A–F. In all cases, NDT codon degeneracy was used except when

performing saturation mutagenesis at site D, in which case NNG codon degeneracy was applied [30]. (Gumulya and Reetz [30]. Reproduced with permission of John Wiley & Sons.)

mutant G230A/D113S/D115T displayed a higher melting temperature by 3 °C and increased half-life at 60 °C ($t_{1/2}$ = 161 vs 74 h of WT).

A number of other B-FIT studies have been reported leading to the thermostabilization of enzymes, sometimes utilizing ISM. Examples include *Pseudomonas fluorescens* esterase [33], feruloyl esterase from *Aspergillus usamii* [34], *Penicillium expansum* [35], endo-1,4- β -galactanase from *Talaromyces stipitatus* [36], ADP-glucose pyrophosphorylase [37], *Pseudomonas aeruginosa* lipase C [38], *Burkholderia* lipase [39], rhamnogalacturonan I lyase from *Bacillus licheniformis* [40] and *Bacillus acidopullulyticus* pullulanase [41]. In the last two cited studies, B-FIT was compared with two other strategies based on the consensus approach (see Section 6.5) and the PoPMuSiC computational aid. B-FIT proved to be the superior strategy, although the authors wisely emphasize that this should not be

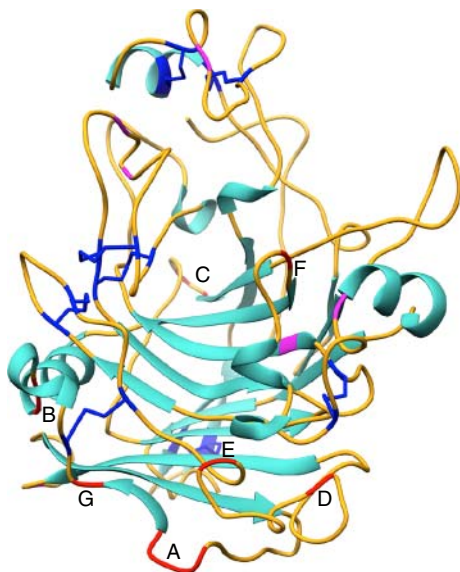


Figure 6.6 B-FIT based thermostabilization of endoglucanase I from *Trichoderma reesei* [32]. Disulfide bonds are shown in blue. N-glycosylation sites are shown in magenta. Mutagenesis sites are shown in red and are labeled as follows: A (aa 284–287), B (aa 301–302), C (aa 113, 115), D (aa 238),

E (aa 230), F (aa 323), and G (aa 291). Mutations at site C and site E resulted in improved TrEGI enzyme variants. PDB code 1EG1. (Chokhawala, <http://bmcbiotechnol.biomedcentral.com/articles/10.1186/s12896-015-0118-z>. Used under CC-BY 4.0, <http://creativecommons.org/licenses/by/4.0/>.)

generalized [40, 41]. Finally, as mentioned in Section 6.2, SM has been applied to residues lining the entrance tunnel of a haloalkane dehalogenase which led to significant thermostabilization [19c]. It is currently not clear whether such an approach can be extended to other enzyme types.

6.4

Iterative Saturation Mutagenesis (ISM) at Protein–Protein Interfacial Sites for Multimeric Enzymes

A novel and highly useful application of ISM was developed in which neither CASTing nor B-FIT was involved. Rather, randomization sites were chosen at a protein–protein interface of a dimeric enzyme [42]. The basic idea was to introduce mutations that prevent dimer dissociation, because strengthening the inter-subunit interface would counteract disintegration. The homodimeric D-tagatose 3-epimerase from *Pseudomonas chichorii* (PcDTE) was chosen as the model enzyme. Initial application of B-FIT provided only a slightly improved variant. The introduction of inter-subunit disulfide bonds failed completely. Therefore, the PcDTE crystal structure was analyzed for interfacial interactions

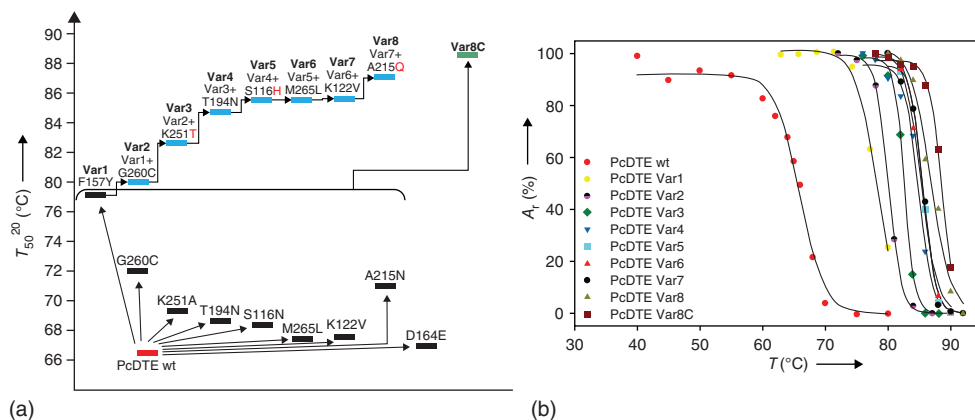


Figure 6.7 Thermostabilization of PcDTE following application of ISM [42]. (a) Thermostability, expressed as the T_{50}^{20} value, of all variants involved in this study: PcDTE wild-type (red bar), hits obtained in the first SM round (black bars), variants 2–8 obtained by ISM (blue bars), and variant 8C obtained by combination of the eight mutations from the

first round (green bar). Mutation D164E was excluded in combinations as no improved variant could be identified during ISM. (b) Residual activity curves of WT PcDTE, variants 1–8, and variant 8C, fitted to a second-order sigmoidal function [42]. (Bosshart *et al.* [42]. Reproduced with permission of John Wiley & Sons.)

using the software PDBePISA [43], giving rise to 44 residues as potential candidates for SM. Three of these residues hardly contribute to the buried surface area, and were therefore excluded from further consideration, as were 10 highly conserved residues that were thought to fulfill other important functions. Each of the remaining 31 residues was subjected to SM using NNK codon degeneracy. In nine libraries thermally improved hits were identified, the best variant being F157Y. Subsequently, ISM was applied as shown in Figure 6.7. As can be seen, an enormous increase in thermostability was evolved. The total number of transformants screened did not exceed 4000. This approach is likely to be general for improving the robustness of multimeric proteins.

6.5

Ancestral and Consensus Approaches and their Structure-Guided Extensions

The original embodiments of ancestral and consensus methods for enhancing the thermal stability of proteins require no structural knowledge. The ancestral approach, first reported in 1990 [44], rests upon the assumption that early ancestors of today's organisms had higher thermostability than extant homologs. The method has not been applied very often, but interesting variations continue to be published [45]. In a study focusing on the B-subunit of DNA gyrase, the ancestral method was compared experimentally with the consensus strategy [45a].

The consensus approach for increasing the robustness of proteins has been used in numerous studies for many years, and recent refinements that include structural data and computational aids such as molecular dynamics (MD) simulations have contributed to its success (see further text). The method is based on the alignment of orthologous proteins leading to the identification of the most prevalent amino acid at a given position; these “suggested” mutations are then introduced in a small mutant library which requires little screening [8]. Early examples of using such multiple sequence alignments (MSAs) focused on the thermostabilization of a fungal phytase [46] and a β -lactamase [47]. Interesting variations followed [48], including the use of Bayesian sequence-based algorithms applied to serine protease sequences in the quest to identify stabilizing interactions in subtilisin E (increase in melting temperature by 13 °C) [49]. In another study, the esterase from *Aspergillus fumigatus* was stabilized by applying the consensus approach, only two (successful) suggestions for point mutations (not a library) being tested experimentally [50]. A 24-fold longer half-life at 50 °C relative to the WT esterase was observed, which was related to the formation of new hydrogen bonds.

Unfortunately, low levels of sequence identity often exist, which means that consensus residues at a given position cannot be determined reliably. A major step forward in this respect was the development of the structure-guided consensus approach [8, 9] in a systematic and reliable manner [51]. In the latter study, the challenging problem of thermostabilizing penicillin G acylase (PGA) was solved. It is a difficult endeavor because it involves a large heterodimeric enzyme (α subunit 23 kDa and β subunit 63 kDa). The two subunits of eight known class IIa PGA sequences were separately aligned. This enabled the identification of 109 out of the 766 possible positions in *E. coli* PGA, which are different in *E. coli*. They occurred in more than 50% of the other sequences. In order to reduce the number of substitutions, structural criteria were applied [51]:

- Only mutations that are more than 10 Å away from the active site were considered.
- Mutations in a helix were considered only if they are not helix-destabilizing amino acids.
- Existing H-bonds or salt bridges were left intact.
- Stabilizing mutations previously reported in the literature were not considered.

This cleverly designed data-driven consensus approach led to a dozen variants (out of the predicted 21) with enhanced thermostability and little tradeoff in activity. Two of them displayed an almost threefold higher half-life at 50 °C. It is important to note that none of the computational guides available at the time (FoldX, PROSA, or SCRATCH) predicted these variants (Table 6.2) [51]. The same group also championed the structure-guided consensus approach in the thermostabilization of glucose dehydrogenase [52]. Respective criteria for reducing the number of predicted amino acid candidates were developed, including previous information regarding amino acid propensities toward secondary structures. In this way, the consensus cut-off percentage was defined.

Table 6.2 Thermal stability results in comparison to computational predictions [51].^{a)}

Residue position	Secondary structure	Distance to active site (Å)	RSA (%)	Experimental thermostability ^{b)}	I-Mutant (kcal mol ⁻¹) ^{c)}	Fold-X (kcal mol ⁻¹)	PROSA 2003 ^{d)}	SCRATCH (kcal mol ⁻¹) ^{e)}
α W25Y	H	31.2	5.4	+	-1.41	2.22	-	-1.55
α A80R	H	29.9	58.1	+	-0.34	0.09	=	-0.83
α M90R	H	29.5	24.5	=	-0.93	0.31	+	-0.96
α Q118E	H	22.6	5.0	Inactive	-0.46	1.66	+	-1.10
α T121D	H	27.7	57.0	-	0.45	-0.05	+	-1.13
α E130T	U	27.9	50.4	Very low activity	0.38	0.14	+	-1.22
α T150N	U	15.5	22.4	+	-0.98	-0.72	=	-1.00
β S9E	T	21.9	1.1	Very low activity	-1.96	-0.06	+	-0.24
β A84P	B	42.2	70.8	+	0.49	-0.24	+	-0.85
β V97K	B	37.2	31.2	Inactive	-1.48	-0.02	+	-1.24
β L100E	B	32.6	53.0	+	-1.16	-0.59	+	-1.17
β A149Q	H	20.1	41.7	Inactive	-0.31	0.34	+	-0.50
β V184K	T	22.7	18.2	-	-2.45	0.39	-	-1.01
β A305D	H	29.9	58.1	+	-0.55	0.31	-	-0.45
β T311P/Q312A	U/T	38.2/36.9	67.8/93.5	+	-1.09/-0.52	-1.08	=	-1.26/-0.36
β N348D	H	18.3	32.8	+	-0.49	1.53	=	-0.62
β V359L	H	19.4	0.0	+	-0.02	-1.55	=	-0.15
β S374T	U	15.6	0.0	=	-0.16	-0.61	+	-1.23
β V400L	T	20.8	9.8	+	-0.20	-1.63	+	-0.51
β A545K	H	32.1	62.6	-	0.15	-0.11	+	-1.19

a) H, helix; B, strand; T, turn; U, unstructured; +, increased thermal stability; -, decreased thermal stability; =, equivalent thermal stability; and RSA, relative solvent accessible area.

b) Thermostability prediction of I-Mutant, Fold-X, PROSA, and SCRATCH is based on the PDB 1PNK structure.

c) I-Mutant2.0: pH 7.0; T, 25 °C; $\Delta\Delta G$ is given in comparison to wild-type, mutants at 311 and 312 were analyzed separately.

d) PROSA: Mutants were labeled based on a clear trend of Z-score, 0.5, or greater, in the case of two off-setting peaks, the mutants were labeled as equivalent, α and β subunits were treated separately.

e) SCRATCH: α and β subunits were treated separately; mutants at 311 and 312 were analyzed separately.

Source: Polizzi *et al.* [51]. Reproduced with permission of John Wiley & Sons.

In an interesting twist on the subject of consensus-based thermostabilization, design using binary polar/nonpolar patterning [53] was explored without phylogenetic bias [54]. The authors called attention to the fact that “because the sequences of natural proteins generally derive from a common ancestor, they tend to be heavily biased by evolutionary relationships.” This is something to keep in mind. Nevertheless, numerous studies utilizing the consensus approach for thermostabilization have proven to be successful [6, 10, 12, 51, 55].

6.6

Computationally Guided Methods

As summarized by recent reviews, various computationally guided approaches to protein thermostabilization have been developed, sometimes in combination with structural and/or sequence data [56]. In most computational setups, the respective algorithms predict a certain number of point mutations that are expected to be stabilizing, typically several dozen. Subsequently, a library comprising these single mutants are generated in the laboratory, of which only a fraction usually prove to be hits, which are then combined. The DeGrado-algorithm, which can be used for various applications including thermostabilization, is an early example [57]. Another computational package is the set of Rosetta algorithms that have been used successfully a number of times for increasing thermostability [58]. An automated protocol as part of Rosetta-design has been developed, which can predict stabilizing point mutations that improve insufficiently packed protein cores [58d]. In many studies, rational design and directed evolution begin to merge, and indeed SM at hot spots can also be included in such studies. Several other approaches are highlighted in more detail in the following sections.

6.6.1

SCHEMA Approach

The SCHEMA approach to directed evolution is described in Section 3.7 [59]. Basically, blocks of sequences are identified in the attempt to minimize structural disruption when recombination with formation of chimeric proteins is induced. Then pairs of interacting residues within 4.5 Å of each other are found computationally, which serve as a basis for constructing contact matrices. An optimization algorithm is included, which selects optimal crossovers that minimize the average disruption of a library. Interactions that break upon recombination are considered in a disruption score. This information is employed when designing optimal shuffling experiments. The technique was used in the thermostabilization of a fungal cellulase [59b], a cellobiohydrolase [60], and a human arginase [61]. More recently it was modified with the development of the so-called non-contiguous recombination (NCR) as a method for rapidly identifying stabilizing mutations needed for enhancing the robustness of proteins [62]. In an initial application, the cellobiohydrolase I from *Hypocrea jecorina* was subjected to NCR [63].

Unlike traditional SCHEMA recombination libraries based on swapping elements of sequence, elements of structure in the form of “blocks” are shuffled among homologous proteins. These elements need not involve contiguous polypeptide sequences. Structural blocks among *H. jecorina* CBH 1 and two thermostable homologs from *Talaromyces emersonii* and *Chaetomium thermophilum* were subjected to swapping, and a subset of CBHs from a library of 5 000 000 possible chimeric sequences was analyzed. Central to NCR is the creation of a graph from nonnative residue–residue contacts with nodes that correspond to residues and edges corresponding to nonnative contact. In this approach, minimal cuts that partition the graph are identified, thereby minimizing the SCHEMA disruption. The blocks are then shuffled which generates a library of noncontiguous chimeras [63]. In this study, several parental enzymes were chosen, the design of a 12-block library having an average SCHEMA disruption of 24.8 and an average of 83.4 mutations from the closest parent (Figure 6.8).

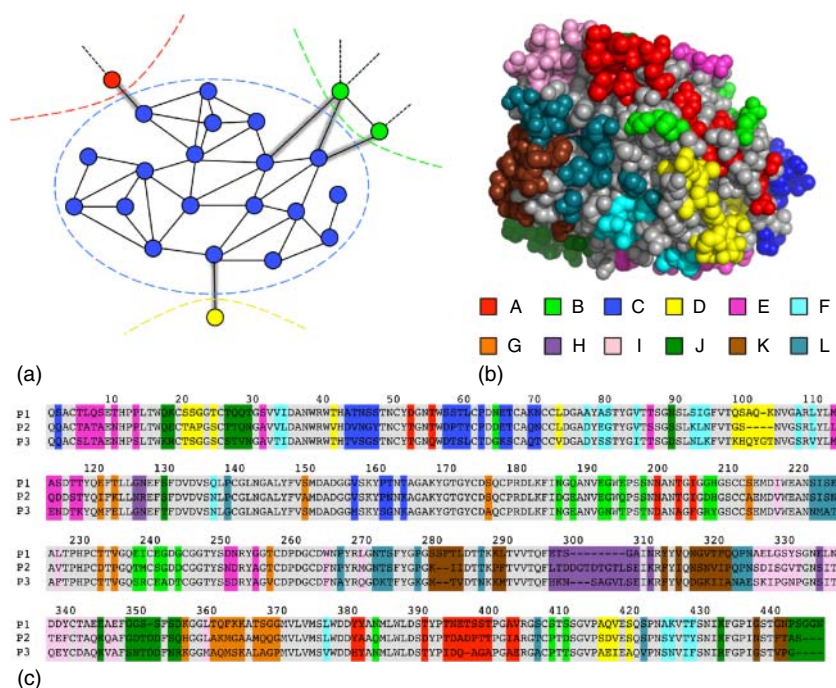


Figure 6.8 SCHEMA-based noncontiguous recombination library design [63]. (a) A graph view of the blue block and neighboring residues, with nodes representing residues, and edges representing residue – residue contacts. Colored, dashed lines define the graph partitions for each block. Contacts to residues from other blocks (highlighted) are broken upon recombination. (b) The

12-block design displayed on the structure of P2 (1Q9H.pdb). Each block (labeled A–L) is represented by a different color, and conserved residues are in gray. (c) The 12-block design displayed on the numbered sequence alignment of the catalytic domains of the three parental enzymes. (Smith *et al.* [63]. Reproduced with permission of American Chemical Society.)

Following a limited amount of lab work, several variants within two blocks were found to stabilize *H. jecorina* CBHI, the best one being characterized by one point mutation that results in the stabilization of this industrially important enzyme by 3 °C. Unfortunately, for an unknown reason combining other stabilizing mutations failed to increase thermostability. As an outlook, the authors suggest that NCR can be improved by prioritizing point mutations on the basis of the consensus approach [63].

6.6.2

FRESCO Approach

Recently, yet another approach to protein thermostabilization was developed, called Framework for Rapid Enzyme Stabilization by Computational libraries (FRESCO) [64a]. This promising concept is illustrated in Figure 6.9. In the first step, multiple algorithms such as Rosetta ddg [64b], FoldX [65], and the newly developed Dynamic Disulfide Discovery (DDD) are used to identify stabilizing mutations (orthogonal *in silico* assaying). These computational tools for choosing stabilizing mutations predict the respective change in the energy of folding ($\Delta\Delta G^{\text{Fold}}$). In steps 2 and 3, false positives are removed by filtering out unreasonable mutations and eliminating variants in which protein flexibility is increased. In step 4, experimental screening is performed prior to combining the most stabilizing mutations in the fifth step (Figure 6.9).

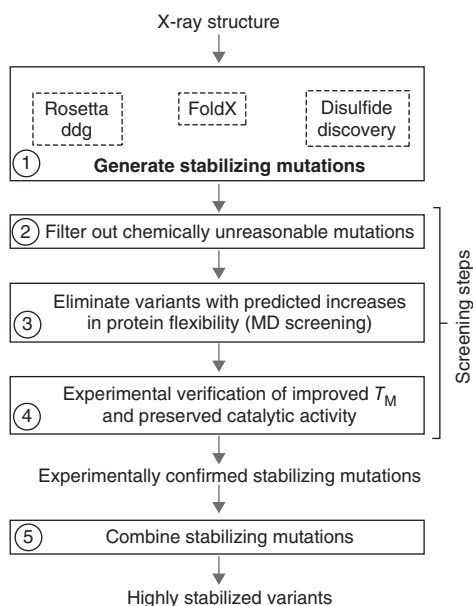


Figure 6.9 FRESCO strategy for protein thermostabilization [64]. (Wijma *et al.* [64]. Reproduced with permission of Oxford University Press.)

FRESCO was first applied to limonene epoxide hydrolase from *Rhodococcus erythropolis* DCL14, which led to dramatic improvements in thermostability [64a]. As a result of screening a library of 64 predicted variants, 21 stabilizing mutations were found in flexible as well as in rigid regions of the enzyme. Thus, the computational “error-rate” amounts to 43 false predictions. Upon combining 10–12 of the experimentally stabilizing mutations, several multi-site mutants resulted showing an impressive increase in apparent melting temperature from WT 50 to 85 °C and a more than 250-fold longer half-life, enhanced catalytic activity, and maintained regioselectivity also being observed. Control experiments showed that the orthogonal *in silico* screening is necessary for success. For example, when applying only FoldX, a significant number of predicted stabilizing mutations proved to be neutral or even destabilizing. According to the authors, the sole use of Rosetta ddg would have predicted only the introduction of aromatic amino acids on the protein surface, a known problem when attempting *in silico* design.

In a follow-up study, the crystal structures of two of the most stable epoxide hydrolase variants allowed a detailed structural analysis of the source of thermostabilization [66]. Variants LEH-P and LEH-F1b are characterized by 8 and 12 mutations, respectively. Both carry the identical eight point mutations S15P, A19K, E45K, T76K, T85V, N92K, Y96F, and E124D, the latter variant also having four point mutations that allow the formation of two disulfide bonds (I5C-E84C and G89C-S91C). The crystal structure of LEH-F1b is shown in Figure 6.10.

The most general insight gained by the structural analysis is the finding that all of the stabilizing point mutations occur either on or near the surface of the

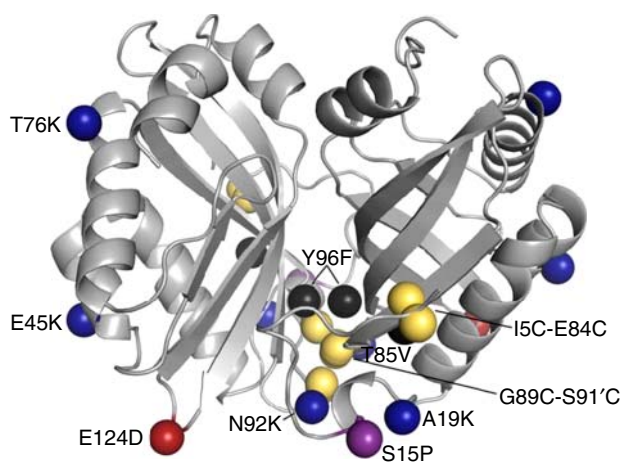


Figure 6.10 Positions of 12 stabilizing mutations as revealed by the crystal structure of the LEH-F1b and P dimers [66]. Mutations introducing surface-located positively charged residues are indicated in blue, surface-located negative charges are shown

in red, and buried hydrophobic residues in black. Proline residues in loops are in purple and disulfide bonds in yellow. Mutations are indicated once per dimer. (Floor *et al.* [66]. Reproduced with permission of John Wiley & Sons.)

enzyme. Many of the beneficial mutations lead to stronger H-bonds that stabilize the local protein structure, the most important ones being localized inside or near the flexible *N*-terminus, and to a lesser degree in helices 3 and 4. The S15P mutation enhances the enzyme's melting temperature by 1.0 °C, whereas the formation of a disulfide bridge shifts it by about 14–15 °C. The mutations do not cause any geometric changes at the active site. A comparison of FRESKO with B-FIT and consensus was also made, although these methods were not directly applied to this epoxide hydrolase. It was shown that both FRESKO and B-FIT predict amino acid substitutions at positions 5, 15, and 92. However, FRESKO also led to mutations at positions that are not flexible and would have missed by B-FIT. FRESKO in turn misses some of the flexible positions predicted by B-FIT. Although in some cases FRESKO and the consensus approach led to similar predictions, important additional mutations are not accessible by the latter method [66].

FRESKO was also successful in the thermostabilization of the haloalkane dehalogenase from *Sphingomonas paucimobilis* [67]. A mere 150 predicted mutants were experimentally tested, of which 18 were found to be stabilizing. Thus, the vast majority of suggested mutants do not fulfill the predictions, which may be theoretically disturbing, but causes no problems in a practical way. Indeed, screening a mere 150 variants is trivial. A combined mutant, generated by selecting three stabilizing point mutations remote from the active site and one disulfide bond, led to an 11 °C increase in thermostability. The best variant was produced by choosing other combinations of stabilizing point mutations, including several that are close to the active site. This resulted in a 23 °C increase in apparent unfolding temperature over WT and a 200-fold lower rate of inactivation. Some trade-off in activity was found as monitored in the dehalogenation of 1-bromopropane [67].

The underlying concept of FRESKO is based on the use of three computational guides, rather than just one as in previous approaches. This means that the user needs to become acquainted with three different computational software packages, Rosetta, FoldX, and DDD. Fast and optimal application of these algorithms requires some experience.

6.6.3

FireProt Approach

In the last step of FRESKO, predicted and generated point mutations are combined into a final variant, which may indeed result in improved protein stability, but also in the opposite effect due to possible nonadditivity. In contrast, it has been claimed that an alternative method called *FireProt* directly “delivers” variants characterized by multiple mutations [68]. The method combines two concepts, the energy- and evolution-based approaches (Figure 6.11). Among other steps, the performances of four prediction tools, FoldX [5], Rosetta [64b], ERIS [69], and CUPSAT [70], are evaluated using the Protherm data set. Additionally, the best multiple-point mutants derived from predictions of $\Delta\Delta G$ following mutation based on a set of crystal structures and evolutionary information obtained

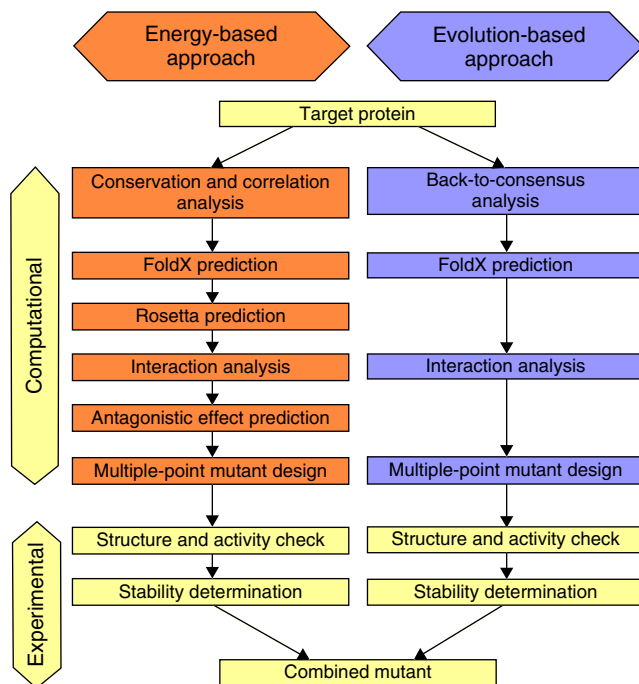


Figure 6.11 Workflow of the FireProt method. Individual steps involved in the energy- and evolution-based approaches [68]. (Bednar, <http://journals.plos.org/>

ploscompbiol/article?id=10.1371/journal.pcbi.1004556. Used under CC-BY 4.0, <http://creativecommons.org/licenses/by/4.0/>.)

from MSA are combined. Intelligent pre- and post-processing filters need to be employed in order to maximize the reliability of prediction.

FireProt was applied successfully to the haloalkane dehalogenase DhaA, which resulted in a notable increase in thermostability ($\Delta T_m = 24.6^\circ\text{C}$), albeit at a 3.2-fold reduction in enzyme activity. A second enzyme, hexachlorocyclohexane dechlorinase (LinA) was also subjected to FireProt, which resulted in $\Delta T_m = 20.9^\circ\text{C}$ [68]. Both results document the effectiveness of this approach to protein thermostabilization.

6.6.4

Constrained Network Analysis (CNA) Approach

A very different computational guide is based on the Constrained Network Analysis (CNA), which can be used to identify, inter alia, structural weak spots in a protein. The graph-theory based rigidity analysis therefore sets the stage for “rational” site-specific mutagenesis or SM for thermostabilization [71]. The correlation of protein rigidity and thermodynamic stability is the basis of the method. Accordingly, a protein is modeled as a constraint network in a “body-and bar”

representation followed by rigidity analysis leading to the identification of rigid clusters of atoms having no internal motion and flexible links in between. The rigidity analysis is performed by application of the pebble game algorithm. In an intriguing study [72], CNA nicely explained a number of mutations in the lipase A from *B. subtilis*, which had been engineered earlier by other research groups [24j, 26, 28, 73]. The primary features of CNA are promising because it:

- Offers a refined modeling of thermal unfolding simulations, which include consideration of temperature dependence of hydrophobic tethers.
- Enables rigidity analyses on ensembles of network topologies on the basis of either structural ensembles or fuzzy noncovalent constraints.
- Provides a computed set of global and local indices needed in quantifying protein stability.

Most recently, VisualCNA has been developed, which is a user-friendly PyMOL plug helpful in rapidly establishing a setup for CNA runs and analyses thereof [74]. Figure 6.12 outlines the general scheme of VisualCNA. The *B. subtilis* lipase was used in the first application of this computational approach to thermostabilization, which proved to be successful [72a]. It will be interesting to see how often this approach will be applied in future directed evolution studies, and whether it can be generalized to include other types of enzymes.

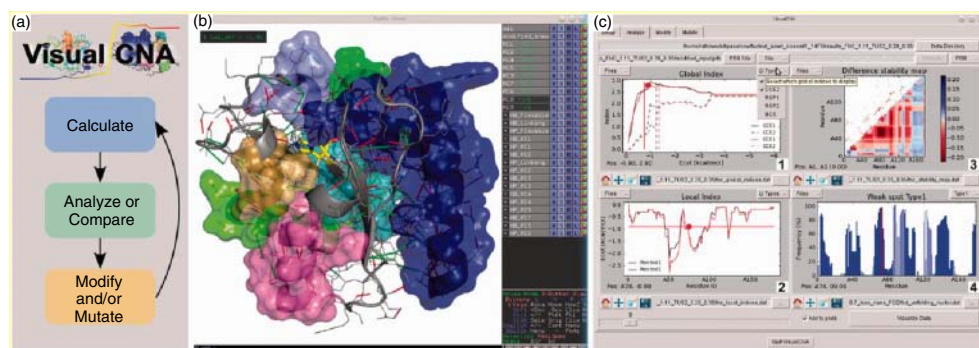


Figure 6.12 Schematic representation of VisualCNA [74]. (a) Illustration of the technique's iterative work flow for optimization of protein thermostability. (b) PyMOL window showing the 3D protein structure at the melting point. Rigid clusters are shown as uniformly colored semi-transparent bodies. Constraints due to hydrogen bonds, salt bridges, and hydrophobic contacts are shown as red, magenta, and green sticks, respectively. A mutation is shown in yellow stick representation. Flexible regions are shown in gray. (c) The VisualCNA Analyze

panel shows a comparison of multiple graphs from wild-type (black) and mutant (red) analyses. (1) Global indices with transition points are indicated as vertical lines. (2) Local index with a red circle indicates the mutation and a horizontal red line shows the unfolding state. (3) Difference stability map between wild-type and mutant. (4) Likelihood of a residue of being a structural weak spot with the mutant is shown in red. (Rathi *et al.* [74]. Reproduced with permission of Oxford University Press.)

6.6.5

Alternative Approaches

Several other computationally guided studies reporting protein thermostabilization have appeared, sometimes utilizing strategies similar to previous ones. In an interesting contribution, a statistical computationally assisted design strategy (SCADS) was applied to the terpene synthase TEAS (tobacco epi-aristolochene synthase) with the aim of enhancing its thermostability [75]. The DeGrado-algorithm was applied [57], which enables predictions on the basis of side chain interactions with the neighboring protein backbone. The activity of WT TEAS as catalyst in the cyclization of FPP to 5-epi-aristolochene breaks down at temperatures above 40 °C, in contrast to a mutant having 12 point mutations, and which is active at 65 °C. Some problems with inclusion bodies were encountered. SM experiments at selected hot spots were not performed, but such an option would probably lead to further improvements [75].

In an integrated approach for thermal stabilization of a mesophilic adenylate kinase, multiple stabilization techniques were applied, including bioinformatics- and structure-guided protocols [76]. Local structural entropy (LSE) played a major role in this endeavor, which is an empirical descriptor for describing conformational variability in short stretches of protein sequence computed on the basis of structural information available in the Protein Data Bank [77]. It was extended by including sequence-based techniques for stabilization [78]. Using previous mutational data [79], the integrated approach led to a variant showing a 25 °C increase in its thermal denaturation midpoint [76]. The crystal structures of three variants were obtained, which led to an insight into the origin of enhanced robustness. A different computational approach provided an adenylate kinase displaying a 21.5 °C increase in T_m [80]. In this study, 100 predicted variants were tested experimentally, the best one having mutations quite different from the ones generated by the integrated approach. This shows that very different strategies may lead to notably stabilized proteins, but the effects on a molecular level can be very different.

In what has been termed “*a divide and combine approach*” to thermal stabilization, the focus shifted to larger and nonfully cooperative proteins that are actually more abundant [81]. The key to success in such systems is to identify the regions of lower stability (Figure 6.13). A stepwise combination of structure-based and rationally designed mutations at such regions was employed in the thermostabilization of an engineered apoflavodoxin, the final variant displaying a 32 °C increase in melting temperature. The underlying concept has great promise for the future.

Finally, protein thermostabilization has been achieved by systematically designing salt bridges [82]. Using this computational approach for enhancing the robustness of a β -glycosidase, 10 556 surface salt bridges in 6493 X-ray structures were analyzed statistically. The model was coupled with B-factors, weighted contact number, relative solvent accessibility, and conservation screening. Electrostatic pairs at five positions were identified, and the combination of three of them led to an increase in T_m of 15.7 °C [82]. As in many other new approaches, it remains to be seen to what extent this technique can be generalized.

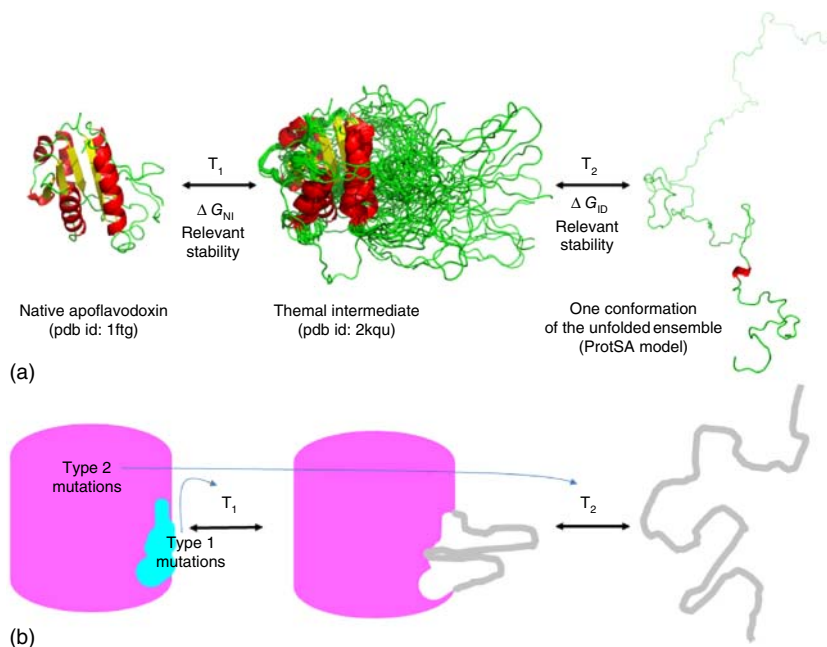


Figure 6.13 Divide and combine approach to protein thermostabilization featuring unfolding equilibria of a three-state protein [81]. (a) Ribbon cartoons represent the conformation of apoflavodoxin in the three states populated in its thermal unfolding equilibrium. The native state is represented by the crystal structure of the WT protein (pdb id: 1ftg); the intermediate state by the solution structure of the F98N variant (pdb id: 2kqu), and the unfolded state by one of the 2000 conformations calculated for the unfolded ensemble using the ProtSA server. The low temperature transition (T_1) signals the unfolding of the less stable region leading to an equilibrium intermediate. The higher temperature transition (T_2) represents the unfolding of the intermediate, leading to the unfolded state. The free energy difference between the native

and the intermediate conformation (ΔG_{NI}) is termed *relevant stability of the protein* while that between the intermediate and the fully unfolded conformation (ΔG_{IU}) is termed *residual stability of the protein*. (b) Simplified scheme depicting a protein with two structural regions of different stability (less stable region in cyan, and more stable one in pink) and the likely effects of mutations on T_1 and T_2 are shown. Type 1 mutations, those introduced in the unstable region or at its interface with the more stable one, will mainly modify the relevant stability of the protein. Type 2 mutations, those introduced in the more stable region, will only modify the residual stability of the protein. (Lamazares, http://www.nature.com/articles/srep09129?WT.ec_id=SREP-20150317. Used under CC-BY 4.0, <http://creativecommons.org/licenses/by/4.0/>.)

References

1. (a) Lotti, M. and Secundo, F. (2015) Editorial: protein stabilization – crossroad for protein-based processes and products. *Biotechnol. J.*, **10** (3), 341–342; (b) Drauz, K., Gröger, H., and May, O. (eds) (2012) *Enzyme Catalysis in Organic Synthesis*, 3rd edn, Wiley-VCH Verlag GmbH, Weinheim; (c) Liese, A., Seelbach, K., and Wandrey, C. (2006) *Industrial Biotransformations*, Wiley-VCH Verlag GmbH, Weinheim.

2. 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.
3. (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.
4. (a) Arnold, F.H. (1998) Design by directed evolution. *Acc. Chem. Res.*, **31**, 125–131; (b) Petrounia, I.P. and Arnold, F.H. (2000) Designed evolution of enzymatic properties. *Curr. Opin. Biotechnol.*, **11** (4), 325–330; (c) Eijsink, V.G.H., Gåseidnes, S., Borchert, T.V., and van den Burg, B. (2005) Directed evolution of enzyme stability. *Biomol. Eng.*, **22** (1–3), 21–30; (d) Bommarius, A.S. and Broering, J.M. (2005) Established and novel tools to investigate biocatalyst stability. *Biocatal. Biotransform.*, **23** (3–4), 125–139; (e) Polizzi, K.M., Bommarius, A.S., Broering, J.M., and Chaparro-Riggers, J.F. (2007) Stability of biocatalysts. *Curr. Opin. Chem. Biol.*, **11** (2), 220–225; (f) Tokuriki, N. and Tawfik, D.S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.*, **19** (5), 596–604; (g) Traxlmayr, M.W. and Obinger, C. (2012) Directed evolution of proteins for increased stability and expression using yeast display. *Arch. Biochem. Biophys.*, **526** (2), 174–180; (h) Suzuki, H., Kobayashi, J., Wada, K., Furukawa, M., and Doi, K. (2015) Thermoadaptation-directed enzyme evolution in an error-prone thermophile derived from *Geobacillus kaustophilus* HTA426. *Appl. Environ. Microbiol.*, **81**, 149–158.
5. Jia, H., Lee, F.S., and Farinas, E.T. (2014) *Bacillus subtilis* spore display of laccase for evolution under extreme conditions of high concentrations of organic solvent. *ACS Comb. Sci.*, **16** (12), 665–669.
6. Stepankova, V., Bidmanova, S., Koudelakova, T., Prokop, Z., Chaloupkova, R., and Damborsky, J. (2013) Strategies for stabilization of enzymes in organic solvents. *ACS Catal.*, **3** (12), 2823–2836.
7. (a) Szilágyi, A. and Závodszy, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, **8** (5), 493–504; (b) Kumar, S., Tsai, C.-J., and Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Eng.*, **13** (3), 179–191; (c) Purmonen, M., Valjakka, J., Takkinen, K., Laitinen, T., and Rouvinen, J. (2007) Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Eng. Des. Sel.*, **20** (11), 551–559.
8. Steipe, B., Schiller, B., Plückthun, A., and Steinbacher, S. (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.*, **240** (3), 188–192.
9. Lehmann, M., Pasamontes, L., Lassen, S.F., and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta*, **1543** (2), 408–415.
10. Yu, H. and Huang, H. (2014) Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnol. Adv.*, **32** (2), 308–315.
11. (a) Sieber, V., Plückthun, A., and Schmid, F.X. (1998) Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.*, **16** (10), 955–960; (b) Schmid, F.-X. (2011) Lessons about protein stability from in vitro selections. *ChemBioChem*, **12** (10), 1501–1507; (c) Kristensen, P. and Winter, G. (1998) Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des.*, **3** (5), 321–328.
12. (a) Bommarius, A.S. and Paye, M.F. (2013) Stabilizing biocatalysts. *Chem. Soc. Rev.*, **42** (15), 6534–6565; (b) Liszka, M.J., Clark, M.E., Schneider, E., and Clark, D.S. (2012) Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.*, **3** (1), 77–102; (c) Suplatov, D., Voevodin, V., and Švedas, V. (2015) Robust enzyme design: bioinformatic tools for improved protein stability. *Biotechnol. J.*, **10** (3), 344–355;

- (d) Socha, R.D. and Tokuriki, N. (2013) Modulating protein stability – directed evolution strategies for improved protein function. *FEBS J.*, **280** (22), 5582–5595.
13. (a) Pierre, B., Labonte, J.W., Xiong, T., Aoraha, E., Williams, A., Shah, V., Chau, E., Helal, K.Y., Gray, J.J., and Kim, J.R. (2015) Molecular determinants for protein stabilization by insertional fusion to a thermophilic host protein. *ChemBioChem*, **16** (16), 2392–2402; (b) Guisan, J.M. (ed.) (2013) *Immobilization of Enzymes and Cells*, 2nd edn, Humana Press, Totowa.
 14. Jackson, C.J., Liu, J.-W., Carr, P.D., Younus, F., Coppin, C., Meirelles, T., Lethier, M., Pandey, G., Ollis, D.L., Russell, R.J., Weik, M., and Oakeshott, J.G. (2013) Structure and function of an insect α -carboxylesterase (α -esterase-7) associated with insecticide resistance. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (25), 10177–10182.
 15. (a) Ishikawa, K., Nakamura, H., Morikawa, K., and Kanaya, S. (1993) Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry*, **32** (24), 6171–6178; (b) Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R., and Gaillardian, C. (2003) Hyperthermostabilization of *Bacillus licheniformis* α -amylase and modulation of its stability over 50 °C temperature range. *Protein Eng.*, **16**, 287–293.
 16. (a) Zhang, S.-B., Pei, X.-Q., and Wu, Z.-L. (2012) Multiple amino acid substitutions significantly improve the thermostability of feruloyl esterase A from *Aspergillus niger*. *Bioresour. Technol.*, **117**, 140–147; (b) Zhang, S.-B. and Wu, Z.-L. (2011) Identification of amino acid residues responsible for increased thermostability of feruloyl esterase A from *Aspergillus niger* using the PoP-MuSiC algorithm. *Bioresour. Technol.*, **102** (2), 2093–2096; (c) Zhang, S.-B., Ma, X.-F., Pei, X.-Q., Liu, J.-Y., Shao, H.-W., and Wu, Z.-L. (2012) A practical high-throughput screening system for feruloyl esterases: substrate design and evaluation. *J. Mol. Catal. B: Enzym.*, **74** (1–2), 36–40.
 17. Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., and Rooman, M. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: Popmusic-2.0. *Bioinformatics*, **25** (19), 2537–2543.
 18. Yamada, R., Higo, T., Yoshikawa, C., China, H., and Ogino, H. (2014) Improvement of the stability and activity of the BPO-A1 haloperoxidase from *Streptomyces aureofaciens* by directed evolution. *J. Biotechnol.*, **192**, 248–254.
 19. (a) Dana, C.M., Saija, P., Kal, S.M., Bryan, M.B., Blanch, H.W., and Clark, D.S. (2012) Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnol. Bioeng.*, **109** (11), 2710–2719; (b) Voutilainen, S., Boer, H., Alapuranen, M., Jänis, J., Vehmaanperä, J., and Koivula, A. (2009) Improving the thermostability and activity of *Melanocarpus albomyces* cellobiohydrolase Cel7B. *Appl. Microbiol. Biotechnol.*, **83** (2), 261–272; (c) Koudelakova, T., Chaloupkova, R., Brezovsky, J., Prokop, Z., Sebestova, E., Hesseler, M., Khabiri, M., Plevaka, M., Kulik, D., Kuta Smatanova, I., Rezacova, P., Ettrich, R., Bornscheuer, U.T., and Damborsky, J. (2013) Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chem. Int. Ed.*, **52** (7), 1959–1963; (d) Yong, K.J. and Scott, D.J. (2015) Rapid directed evolution of stabilized proteins with cellular high-throughput encapsulation solubilization and screening (CHESS). *Biotechnol. Bioeng.*, **112** (3), 438–446; (e) Steffler, E., Guterl, J.-K., and Sieber, V. (2013) Improvement of thermostable aldehyde dehydrogenase by directed evolution for application in synthetic cascade biomanufacturing. *Enzyme Microb. Technol.*, **53**, 307–314.
 20. (a) Ruller, R., Alpointi, J., Deliberto, L.A., Zanthorlin, L.M., Machado, C.B., and Ward, R.J. (2014) Concomitant adaptation of a GH11 xylanase by directed evolution to create an alkali-tolerant/thermophilic enzyme. *Protein Eng. Des. Sel.*, **27** (8), 255–262; (b) Ruller, R., Deliberto, L., Ferreira, T.L.,

- and Ward, R.J. (2008) Thermostable variants of the recombinant xylanase from *Bacillus subtilis* produced by directed evolution show reduced heat capacity changes. *Proteins: Struct. Funct. Bioinf.*, **70** (4), 1280–1293.
21. (a) Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X., Tomlinson, G.L., Verruto, J., Wong, V.W.K., Mathur, E.J., Short, J.M., Robertson, D.E., and Steer, B.A. (2004) An evolutionary route to xylanase process fitness. *Protein Sci.*, **13** (2), 494–503; (b) McHunu, N.P., Singh, S., and Permaul, K. (2009) Expression of an alkali-tolerant fungal xylanase enhanced by directed evolution in *Pichia pastoris* and *Escherichia coli*. *J. Biotechnol.*, **141** (1–2), 26–30; (c) Qaim, M., Subramanian, A., and Sadashivappa, P. (2009) Commercialized GM crops and yield. *Nat. Biotechnol.*, **27** (9), 803–804; (d) Hokanson, C.A., Cappuccilli, G., Odineca, T., Bozic, M., Behnke, C.A., Mendez, M., Coleman, W.J., and Crea, R. (2011) Engineering highly thermostable xylanase variants using an enhanced combinatorial library method. *Protein Eng. Des. Sel.*, **24** (8), 597–605; (e) Wang, Y., Feng, S., Zhan, T., Huang, Z., Wu, G., and Liu, Z. (2013) Improving catalytic efficiency of endo- β -1, 4-xylanase from *Geobacillus stearothermophilus* by directed evolution and H179 saturation mutagenesis. *J. Biotechnol.*, **168** (4), 341–347; (f) Zheng, H., Liu, Y., Sun, M., Han, Y., Wang, J., Sun, J., and Lu, F. (2014) Improvement of alkali stability and thermostability of *Paenibacillus campinasensis* family-11 xylanase by directed evolution and site-directed mutagenesis. *J. Ind. Microbiol. Biotechnol.*, **41** (1), 153–162; (g) Qian, C., Liu, N., Yan, X., Wang, Q., Zhou, Z., and Wang, Q. (2015) Engineering a high-performance, metagenomic-derived novel xylanase with improved soluble protein yield and thermostability. *Enzyme Microb. Technol.*, **70**, 35–41.
 22. Jia, H., Li, Y., Liu, Y., Yan, Q., Yang, S., and Jiang, Z. (2012) Engineering a thermostable β -1,3-1,4-glucanase from *Paecilomyces thermophila* to improve catalytic efficiency at acidic pH. *J. Biotechnol.*, **159** (1–2), 50–55.
 23. (a) Miyazaki, K. and Arnold, F.H. (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J. Mol. Evol.*, **49** (6), 716–720; (b) Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T., and Jaeger, K.E. (2000) Directed evolution of an enantioselective lipase. *Chem. Biol.*, **7** (9), 709–718.
 24. (a) Niederhauser, B., Siivonen, J., Määttä, J.A., Jänis, J., Kulomaa, M.S., and Hytönen, V.P. (2012) DNA family shuffling within the chicken avidin protein family – a shortcut to more powerful protein tools. *J. Biotechnol.*, **157** (1), 38–49; (b) Taskinen, B., Airenne, T.T., Jänis, J., Rahikainen, R., Johnson, M.S., Kulomaa, M.S., and Hytönen, V.P. (2014) A novel chimeric avidin with increased thermal stability using DNA shuffling. *PLoS One*, **9** (3), e92058; (c) Wu, I. and Arnold, F.H. (2013) Engineered thermostable fungal Cel6A and Cel7A cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. *Biotechnol. Bioeng.*, **110** (7), 1874–1883; (d) Liang, C., Gui, X., Zhou, C., Xue, Y., Ma, Y., and Tang, S.-Y. (2015) Improving the thermoactivity and thermostability of pectate lyase from *Bacillus pumilus* for ramie degumming. *Appl. Microbiol. Biotechnol.*, **99** (6), 2673–2682; (e) Stephens, D.E., Khan, F.I., Singh, P., Bisetty, K., Singh, S., and Permaul, K. (2014) Creation of thermostable and alkaline stable xylanase variants by DNA shuffling. *J. Biotechnol.*, **187**, 139–146; (f) Buettner, K., Hertel, T., and Pietzsch, M. (2012) Increased thermostability of microbial transglutaminase by combination of several hot spots evolved by random and saturation mutagenesis. *Amino Acids*, **42** (2-3), 987–996; (g) Gonzalez-Perez, D., Garcia-Ruiz, E., Ruiz-Dueñas, F.J., Martinez, A.T., and Alcalde, M. (2014) Structural determinants of oxidative stabilization in an evolved versatile peroxidase. *ACS Catal.*, **4** (11), 3891–3901; (h) Huang, L., Xu, J.-H., and Yu, H.-L.

- (2015) Significantly improved thermostability of a reductase CgSKR1 from *Candida glabrata* with a key mutation at Asp 138 for enhancing bioreduction of aromatic α -keto esters. *J. Biotechnol.*, **203**, 54–61; (i) Parker, B.M., Taylor, I.N., Woodley, J.M., Ward, J.M., and Dalby, P.A. (2011) Directed evolution of a thermostable L-aminoacylase biocatalyst. *J. Biotechnol.*, **155** (4), 396–405; (j) Kamal, M.Z., Ahmad, S., Molugu, T.R., Vijayalakshmi, A., Deshmukh, M.V., Sankaranarayanan, R., and Rao, N.M. (2011) In vitro evolved non-aggregating and thermostable lipase: structural and thermodynamic investigation. *J. Mol. Biol.*, **413** (3), 726–741.
25. Fulton, A., Frauenkron-Machedjou, V.J., Skoczinski, P., Wilhelm, S., Zhu, L., Schwaneberg, U., and Jaeger, K.E. (2015) Exploring the protein stability landscape: *Bacillus subtilis* lipase A as a model for detergent tolerance. *ChemBioChem*, **16** (6), 930–936.
 26. (a) Reetz, M.T., Carballera, J.D., and Vogel, A. (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.*, **45** (46), 7745–7751; (b) Reetz, M.T. and Carballera, J.D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protocols*, **2** (4), 891–903.
 27. Reetz, M.T., Soni, P., Acevedo, J.P., and Sanchis, J. (2009) Creation of an amino acid network of structurally coupled residues in the directed evolution of a thermostable enzyme. *Angew. Chem. Int. Ed.*, **48** (44), 8268–8272.
 28. (a) Augustyniak, W., Brzezinska, A.A., Pijning, T., Wienk, H., Boelens, R., Dijkstra, B.W., and Reetz, M.T. (2012) Biophysical characterization of mutants of *Bacillus subtilis* lipase evolved for thermostability: factors contributing to increased activity retention. *Protein Sci.*, **21** (4), 487–497; (b) Augustyniak, W., Wienk, H., Boelens, R., and Reetz, M.T. (2013) ^1H , ^{13}C and ^{15}N resonance assignments of wild-type *Bacillus subtilis* Lipase A and its mutant evolved towards thermostability. *Biomol. NMR Assign.*, **7** (2), 249–252.
 29. Reetz, M.T., Soni, P., Fernandez, L., Gumulya, Y., and Carballera, J.D. (2010) Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem. Commun.*, **46** (45), 8657–8658.
 30. Gumulya, Y. and Reetz, M.T. (2011) Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *ChemBioChem*, **12** (16), 2502–2510.
 31. Gumulya, Y., Sanchis, J., and Reetz, M.T. (2012) Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *ChemBioChem*, **13** (7), 1060–1066.
 32. Chokhawala, H., Roche, C., Kim, T.-W., Atreya, M., Vegesna, N., Dana, C., Blanch, H., and Clark, D. (2015) Mutagenesis of *Trichoderma reesei* endoglucanase I: impact of expression host on activity and stability at elevated temperatures. *BMC Biotechnol.*, **15** (1), 11.
 33. Jochens, H., Aerts, D., and Bornscheuer, U.T. (2010) Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Protein Eng. Des. Sel.*, **23** (12), 903–909.
 34. Yin, X., Li, J.-F., Wang, C.-J., Hu, D., Wu, Q., Gu, Y., and Wu, M.C. (2015) Improvement in the thermostability of a type A feruloyl esterase, AuFaeA, from *Aspergillus usamii* by iterative saturation mutagenesis. *Appl. Microbiol. Biotechnol.*, **99** (23), 10047–10056.
 35. Frascari, D., Zannoni, A., Pinelli, D., and Nocentini, M. (2007) Chloroform aerobic cometabolism by butane-utilizing bacteria in bioaugmented and non-bioaugmented soil/groundwater microcosms. *Process Biochem.*, **42** (8), 1218–1228.
 36. Larsen, D., Nyffenegger, C., Swiniarska, M., Thygesen, A., Strube, M., Meyer, A., and Mikkelsen, J. (2015) Thermostability enhancement of an endo-1,4- β -galactanase from *Talaromyces stipitatus* by site-directed mutagenesis.

- Appl. Microbiol. Biotechnol.*, **99** (10), 4245–4253.
37. Boehlein, S.K., Shaw, J.R., Stewart, J.D., Sullivan, B., and Hannah, L.C. (2015) Enhancing the heat stability and kinetic parameters of the maize endosperm ADP-glucose pyrophosphorylase using iterative saturation mutagenesis. *Arch. Biochem. Biophys.*, **568**, 28–37.
 38. Cesarini, S., Bofill, C., Pastor, F.I.J., Reetz, M.T., and Diaz, P. (2012) A thermostable variant of *P. aeruginosa* cold-adapted Lip C obtained by rational design and saturation mutagenesis. *Process Biochem.*, **47** (12), 2064–2071.
 39. Liu, Y., Qiu, L., Huang, J., Zhao, B., Wang, Z., Zhu, X., Gao, Y., and Shu, Z. (2015) Screening for mutants with thermostable lipase A from *Burgholderia* sp ZYB002. *Acta Microbiol. Sin.*, **55**, 748–754.
 40. Silva, I., Larsen, D., Jers, C., Derkx, P., Meyer, A., and Mikkelsen, J. (2013) Enhancing RGI lyase thermostability by targeted single point mutations. *Appl. Microbiol. Biotechnol.*, **97** (22), 9727–9735.
 41. Chen, A., Li, Y., Nie, J., McNeil, B., Jeffrey, L., Yang, Y., and Bai, Z. (2015) Protein engineering of *Bacillus acidopullulyticus* pullulanase for enhanced thermostability using in silico data driven rational design methods. *Enzyme Microb. Technol.*, **78**, 74–83.
 42. Bosshart, A., Panke, S., and Bechtold, M. (2013) Systematic optimization of interface interactions increases the thermostability of a multimeric enzyme. *Angew. Chem. Int. Ed.*, **52** (37), 9673–9676.
 43. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372** (3), 774–797.
 44. Malcolm, B.A., Wilson, K.P., Matthews, B.W., Kirsch, J.F., and Wilson, A.C. (1990) Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature*, **345**, 86–89.
 45. (a) Akanuma, S., Iwami, S., Yokoi, T., Nakamura, N., Watanabe, H., Yokobori, S., and Yamagishi, A. (2011) Phylogeny-based design of a B-subunit of DNA gyrase and its ATPase domain using a small set of homologous amino acid sequences. *J. Mol. Biol.*, **412** (2), 212–225; (b) Perez-Jimenez, R., Ingles-Prieto, A., Zhao, Z.M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T.J., Tanokura, M., Holmgren, A., Sanchez-Ruiz, J.M., Gaucher, E.A., and Fernandez, J.M. (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.*, **18** (5), 592–596.
 46. (a) Lehmann, M. and Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.*, **12** (4), 371–375; (b) Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.*, **15** (5), 403–411.
 47. Amin, N., Liu, A.D., Ramer, S., Ahle, W., Meijer, D., Metin, M., Wong, S., Gualfetti, P., and Schellenberger, V. (2004) Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.*, **17** (11), 787–793.
 48. (a) Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Plückthun, A. (2003) Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.*, **332** (2), 489–503; (b) Watanabe, K., Ohkuri, T., Yokobori, S.-I., and Yamagishi, A. (2006) Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J. Mol. Biol.*, **355** (4), 664–674.
 49. DiTursi, M.K., Kwon, S.-J., Reeder, P.J., and Dordick, J.S. (2006) Bioinformatics-driven, rational engineering of protein thermostability. *Protein Eng. Des. Sel.*, **19** (11), 517–524.
 50. Zhang, S., Wu, G., Feng, S., and Liu, Z. (2014) Improved thermostability of esterase from *Aspergillus fumigatus*

- by site-directed mutagenesis. *Enzyme Microb. Technol.*, **64–65**, 11–16.
51. Polizzi, K.M., Chaparro-Riggers, J.F., Vázquez-Figueroa, E., and Bommarius, A.S. (2006) Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Biotechnol. J.*, **1** (5), 531–536.
 52. Vázquez-Figueroa, E., Chaparro-Riggers, J., and Bommarius, A.S. (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *ChemBioChem*, **8** (18), 2295–2301.
 53. Kamtekar, S., Schiffer, J., Xiong, H., Babik, J., and Hecht, M. (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science*, **262** (5140), 1680–1685.
 54. (a) Besenmatter, W., Kast, P., and Hilvert, D. (2007) Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins*, **66** (2), 500–506; (b) Jäckel, C., Bloom, J.D., Kast, P., Arnold, F.H., and Hilvert, D. (2010) Consensus protein design without phylogenetic bias. *J. Mol. Biol.*, **399** (4), 541–546.
 55. (a) Huang, L., Ma, H.-M., Yu, H.-L., and Xu, J.-H. (2014) Altering the substrate specificity of reductase CgKR1 from *Candida glabrata* by protein engineering for bioreduction of aromatic α -keto esters. *Adv. Synth. Catal.*, **356** (9), 1943–1948; (b) Anbar, M., Gul, O., Lamed, R., Sezerman, U.O., and Bayer, E.A. (2012) Improved thermostability of clostridium thermocellum endoglucanase cel8a by using consensus-guided mutagenesis. *Appl. Environ. Microbiol.*, **78** (9), 3458–3464; (c) Zhang, D., Zhu, F., Fan, W., Tao, R., Yu, H., Yang, Y., Jiang, W., and Yang, S. (2011) Gradually accumulating beneficial mutations to improve the thermostability of N-carbamoyl-D-amino acid amidohydrolase by step-wise evolution. *Appl. Microbiol. Biotechnol.*, **90** (4), 1361–1371; (d) Trudeau, D.L., Lee, T.M., and Arnold, F.H. (2014) Engineered thermostable fungal cellulases exhibit efficient synergistic cellulose hydrolysis at elevated temperatures. *Biotechnol. Bioeng.*, **111** (12), 2390–2397; (e) Komor, R.S., Romero, P.A., Xie, C.B., and Arnold, F.H. (2012) Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng. Des. Sel.*, **25** (12), 827–833.
 56. (a) Pantazes, R.J., Grisewood, M.J., and Maranas, C.D. (2011) Recent advances in computational protein design. *Curr. Opin. Struct. Biol.*, **21** (4), 467–472; (b) Wijma, H.J., Floor, R.J., and Janssen, D.B. (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.*, **23** (4), 588–594.
 57. (a) Bender, G.M., Lehmann, A., Zou, H., Cheng, H., Fry, H.C., Engel, D., Therien, M.J., Blasie, J.K., Roder, H., Saven, J.G., and DeGrado, W.F. (2007) De novo design of a single-chain diphenylporphyrin metalloprotein. *J. Am. Chem. Soc.*, **129** (35), 10732–10740; (b) Nanda, V., Rosenblatt, M.M., Osyczka, A., Kono, H., Getahun, Z., Dutton, P.L., Saven, J.G., and DeGrado, W.F. (2005) De novo design of a redox-active minimal rubredoxin mimic. *J. Am. Chem. Soc.*, **127** (16), 5804–5805; (c) Calhoun, J.R., Kono, H., Lahr, S., Wang, W., DeGrado, W.F., and Saven, J.G. (2003) Computational design and characterization of a monomeric helical dinuclear metalloprotein. *J. Mol. Biol.*, **334** (5), 1101–1115.
 58. (a) Korkegian, A., Black, M.E., Baker, D., and Stoddard, B.L. (2005) Computational thermostabilization of an enzyme. *Science*, **308**, 857–860; (b) Sheffler, W. and Baker, D. (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.*, **18** (1), 229–239; (c) Sheffler, W. and Baker, D. (2010) RosettaHoles2: a volumetric packing measure for protein structure refinement and validation. *Protein Sci.*, **19** (10), 1991–1995; (d) Borgo, B. and Havranek, J.J. (2012) Automated selection of stabilizing mutations in designed and natural proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (5), 1494–1499; (e) Kiss, G., Celebi-Olcum, N., Moretti, R., Baker, D., and Houk, K.N. (2013) Computational enzyme design. *Angew. Chem. Int. Ed.*, **52** (22), 5700–5725.

59. (a) Saab-Rincon, G., Li, Y., Meyer, M., Carbone, M., Landwehr, M., and Arnold, F.H. (2009) in *Protein Engineering Handbook* (eds S. Lutz and U.T. Bornscheuer), Wiley-VCH Verlag GmbH, Weinheim, pp. 481–492; (b) Heinzelman, P., Snow, C.D., Smith, M.A., Yu, X., Kannan, A., Boulware, K., Villalobos, A., Govindarajan, S., Minshull, J., and Arnold, F.H. (2009) SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J. Biol. Chem.*, **284**, 26229–26233; (c) Pantazes, R.J., Saraf, M.C., and Maranas, C.D. (2007) Optimal protein library design using recombination or point mutations based on sequence based scoring functions. *Protein Eng. Des. Sel.*, **20**, 361–373.
60. Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C., and Arnold, F. (2010) Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. *Protein Eng. Des. Sel.*, **23** (11), 871–880.
61. Romero, P.A., Stone, E., Lamb, C., Chantranupong, L., Krause, A., Miklos, A.E., Hughes, R.A., Fechtel, B., Ellington, A.D., Arnold, F.H., and Georgiou, G. (2012) SCHEMA-designed variants of human arginase I and II reveal sequence elements important to stability and catalysis. *ACS Synth. Biol.*, **1** (6), 221–228.
62. Smith, M.A., Romero, P.A., Wu, T., Brustad, E.M., and Arnold, F.H. (2013) Chimera genesis of distantly-related proteins by noncontiguous recombination. *Protein Sci.*, **22** (2), 231–238.
63. Smith, M.A., Bedbrook, C.N., Wu, T., and Arnold, F.H. (2013) *Hypocrea jecorina* cellobiohydrolase I stabilizing mutations identified using noncontiguous recombination. *ACS Synth. Biol.*, **2** (12), 690–696.
64. (a) Wijma, H.J., Floor, R.J., Jekel, P.A., Baker, D., Marrink, S.J., and Janssen, D.B. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.*, **27** (2), 49–58; (b) Kellogg, E.H., Leaver-Fay, A., and Baker, D. (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins*, **79** (3), 830–838.
65. Guerois, R., Nielsen, J.E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320** (2), 369–387.
66. Floor, R.J., Wijma, H.J., Jekel, P.A., Terwisscha van Scheltinga, A.C., Dijkstra, B.W., and Janssen, D.B. (2015) X-ray crystallographic validation of structure predictions used in computational design for protein stabilization. *Proteins*, **83** (5), 940–951.
67. Floor, R.J., Wijma, H.J., Colpa, D.I., Ramos-Silva, A., Jekel, P.A., Szymanski, W., Feringa, B.L., Marrink, S.J., and Janssen, D.B. (2014) Computational library design for increasing haloalkane dehalogenase stability. *ChemBioChem*, **15** (11), 1660–1672.
68. Bednar, D., Beerens, K., Sebestova, E., Bendl, J., Khare, S., Chaloupkova, R., Prokop, Z., Brezovsky, J., Baker, D., and Damborsky, J. (2015) FireProt: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.*, **11** (11), e1004556.
69. Yin, S., Ding, F., and Dokholyan, N.V. (2007) Eris: An automated estimator of protein stability. *Nat. Methods*, **4** (6), 466–467.
70. Parthiban, V., Gromiha, M.M., and Schomburg, D. (2006) CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.*, **34** (11), W239–W242.
71. (a) Rath, P.C., Radestock, S., and Gohlke, H. (2012) Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio. *J. Biotechnol.*, **159** (3), 135–144; (b) Kruger, D.M., Rath, P.C., Pfleger, C., and Gohlke, H. (2013) CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability, and function. *Nucleic Acids Res.*, **41**, W340–W348; (c) Pfleger, C., Rath, P.C., Klein, D.L., Radestock, S., and Gohlke, H. (2013) Constraint Network Analysis

- (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function. *J. Chem. Inf. Model.*, **53** (4), 1007–1015.
72. (a) Rath, P.C., Jaeger, K.E., and Gohlke, H. (2015) Structural rigidity and protein thermostability in variants of lipase A from *Bacillus subtilis*. *PLoS One*, **10**, e0130289; (b) Rath, P.C., Fulton, A., Jaeger, K.-E., and Gohlke, H. (2016) Application of rigidity theory to the thermostabilization of lipase A from *Bacillus subtilis*. *PLoS Comput. Biol.*, **12**, e1004754.
 73. (a) Ahmad, S. and Rao, N.M. (2009) Thermally denatured state determines refolding in lipase: mutational analysis. *Protein Sci.*, **18** (6), 1183–1196; (b) Srivastava, A. and Sinha, S. (2014) Thermostability of in vitro evolved *Bacillus subtilis* lipase A: a network and dynamics perspective. *PLoS One*, **9**, e102856.
 74. Rath, P.C., Mulnaes, D., and Gohlke, H. (2015) VisualCNA: A GUI for interactive constraint network analysis and protein engineering for improving thermostability. *Bioinformatics*, **31** (14), 2394–2396.
 75. Diaz, J.E., Lin, C.S., Kunishiro, K., Feld, B.K., Avranitis, S.K., Bronson, J., Greaves, J., Saven, J.G., and Weiss, G.A. (2011) Computational design and selections for an engineered, thermostable terpene synthase. *Protein Sci.*, **20** (9), 1597–1606.
 76. Moon, S., Jung, D.K., Phillips, G.N. Jr., and Bae, E. (2014) An integrated approach for thermal stabilization of a mesophilic adenylate kinase. *Proteins*, **82** (9), 1947–1959.
 77. Chan, C.H., Liang, H.K., Hsiao, N.W., Ko, M.T., Lyu, P.C., and Hwang, J.K. (2004) Relationship between local structural entropy and protein thermostability. *Proteins*, **57** (4), 684–691.
 78. (a) Bae, E., Bannen, R.M., and Phillips, G.N. Jr., (2008) Bioinformatic method for protein thermal stabilization by structural entropy optimization. *Proc. Natl. Acad. Sci. U.S.A.*, **105** (28), 9594–9597; (b) Bannen, R.M., Suresh, V., Phillips, G.N. Jr., Wright, S.J., and Mitchell, J.C. (2008) Optimal design of thermally stable proteins. *Bioinformatics*, **24** (20), 2339–2343.
 79. Bae, E. and Phillips, G.N. Jr., (2006) Roles of static and dynamic domains in stability and catalysis of adenylate kinase. *Proc. Natl. Acad. Sci. U.S.A.*, **103** (7), 2132–2137.
 80. Howell, S.C., Inampudi, K.K., Bean, D.P., and Wilson, C.J. (2014) Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure*, **22** (2), 218–229.
 81. Lamazares, E., Clemente, I., Bueno, M., Velazquez-Campoy, A., and Sancho, J. (2015) Rational stabilization of complex proteins: a divide and combine approach. *Sci. Rep.*, **5**, 9129.
 82. Lee, C.W., Wang, H.J., Hwang, J.K., and Tseng, C.P. (2014) Protein thermal stability enhancement by designing salt bridges: a combined computational and experimental study. *PLoS One*, **9** (11), e112751.

7

Directed Evolution of Promiscuity: Artificial Enzymes as Catalysts in Organic Chemistry

7.1

Introductory Background Information

The term “*promiscuous activity*” in enzymology has been used in different ways. Originally, biochemists and biotechnologists applied it whenever an enzyme was found to catalyze the reaction of a natural compound that is not the actual natural substrate, but later the term was extended to include the transformation of unnatural compounds [1]. It also became clear that nature has allowed the evolution of a given enzyme not just for a single purpose, but surprisingly often for reaching several goals. Jensen was the first to postulate that broad specificity and promiscuous activities of enzymes constitute the starting points for the evolution of new and selective functions [2]. This phenomenon was later mimicked in the laboratory by applying directed evolution to the human estrogen sulfotransferase with the aim of converting it from a generalist to a specialist [3]. The idea of enzyme ancestor reconstruction or resurrection, proposed by Benner [4], has been implemented in laboratory experiments using directed evolution methods [5]. These and other insights are not only of central importance in evolutionary biology, they are also of practical interest in biotechnology, specifically when aiming to expand or shift the range of substrate acceptance [6].

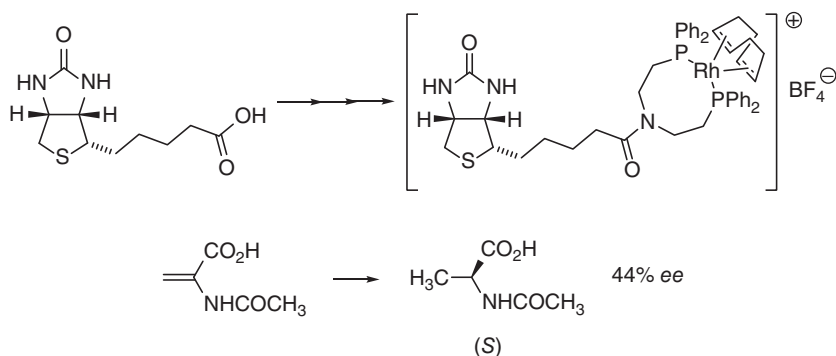
It was also discovered that sometimes a given enzyme can catalyze a transformation that is usually catalyzed by another type of enzyme, one of many examples being the ability of a decarboxylase to catalyze acyloin condensation [7]. Aside from the theoretical interest in this phenomenon, the practical benefit needs to be assessed in each case. This also applies to those studies in which such a switch has been achieved by protein engineering, as in the conversion of an esterase into an epoxide hydrolase [8] (see Section 7.2), or in the conversion of a phosphotriesterase into an arylesterase [9]. These and other aspects of enzyme promiscuity have been addressed in numerous studies [10]. The ability of a protein or a chemically and/or genetically modified protein to catalyze a reaction type that does not occur in nature has also been described as being promiscuous [11]. Minimalist

active-site redesign based on the introduction of a limited number of rationally chosen point mutations has proven to be successful in a number of studies as summarized in an extensive review [12].

In addition to various review articles covering specific aspects [6], the general area of enzyme promiscuity has been summarized [13].

Many cases of promiscuity have been discovered simply by screening a set of enzymes (or proteins in general) for a certain reaction type not known to be catalyzed by these biocatalysts, guided by mechanistic considerations and sometimes accompanied by designed point mutations (but less often by directed evolution). Seminal examples include lipases (or esterases) as catalysts in aldol and Michael reactions [14], epoxidation reactions [15], Mannich reactions [16], Markovnikov additions [17], and Baylis–Hillman reactions [18]. Progress in this research area has been reviewed [19].

Promiscuous catalytic behavior can also be induced by attaching a ligand/transition metal moiety to a host protein either covalently or noncovalently. Such artificial metalloenzymes are sometimes called *hybrid catalysts*. The Whitesides system comprising a biotinylated rhodium/diphosphine-complex, which is conjugated noncovalently to avidin (or streptavidin) is a seminal example [20]. The bound achiral Rh-complex was employed as the catalyst in asymmetric olefin hydrogenation of *N*-acyl acrylic acid (Scheme 7.1).



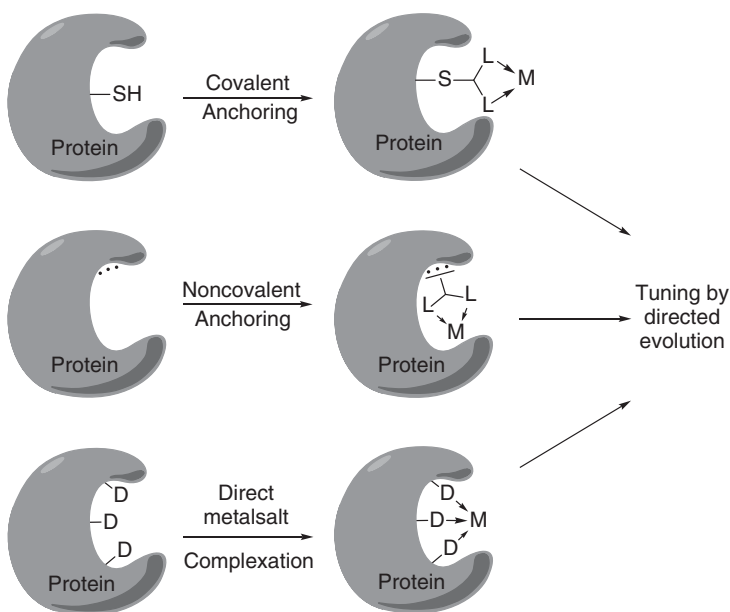
Scheme 7.1 Whitesides system comprising a biotinylated achiral diphosphine/Rh-complex noncovalently bound to avidin, which was used as the catalyst in the asymmetric olefin-hydrogenation of *N*-acyl acrylic acid [20]. Later streptavidin was employed as host [21, 22].

Although enantioselectivity proved to be moderate, decades later this study opened the door to many interesting extensions that include a variety of different transition metal catalyzed reaction types [21, 22]. Using wild-type (WT) streptavidin (or avidin), a single transition metal catalyst is produced, the catalytic profile of which being a matter of fortune. Fortunately, as first suggested in 2001–2002 [23] and further explained in Section 7.2, tuning by directed evolution offers exciting opportunities for developing truly efficient biocatalysts in organic chemistry, stereo- and/or regioselectivity being the primary parameters of interest. However,

this endeavor is not as easy as it may appear because several prerequisites have to be fulfilled, which will be discussed in more detail in Section 7.2:

- The host protein needs to be stable under operating conditions.
- The host protein needs to be expressed efficiently to provide sufficient quantities of protein in miniaturized and parallelized form.
- A simple and efficient protein purification in miniaturized form (microplates) has to be available.
- Bioconjugation should be regiospecific and essentially quantitative.

Following the original Whitesides publication, many research groups focused on devising other types of artificial metalloenzymes as catalysts in synthetic organic chemistry, numerous examples appearing in the 1980s and 1990s [24]. Today, this exciting research area continues to arouse the imagination of chemists and biotechnologists alike, as summarized by review articles that have emphasized different strategies [21b,c, 22a,b, 25, 26]. The general approach to creating artificial metalloenzymes by anchoring a designed ligand/metal moiety to an appropriate protein serving as a host has been systematized according to Scheme 7.2 [21b,c, 25].



Scheme 7.2 Systematization for generating artificial metalloenzymes as hybrid catalysts [21b,c, 25]. L, synthetic ligand; M, transition metal; D, donor atoms of side-chains of appropriate amino acids such as aspartate or cysteine which bind transition metals M directly.

Several other concepts for noncovalent bioconjugation with introduction of transition metal catalysts into host proteins have been developed, which

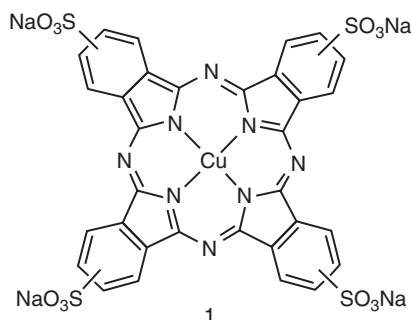
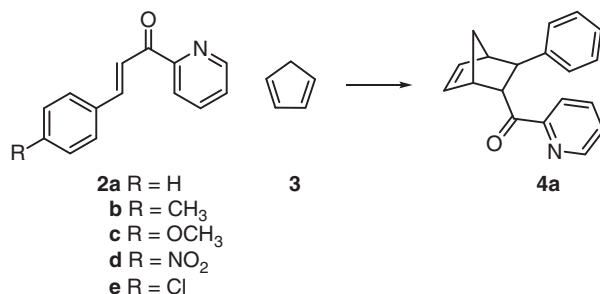


Figure 7.1 Water-soluble Cu(II)-phthalocyanine used in bioconjugation to serum albumins [29]. (Ohashi *et al.* [29]. Reproduced with permission of John Wiley & Sons.)

in principle can also be tuned by directed evolution [21b,c, 25]. For example, serum albumins binding noncovalently water-soluble sulfonylated Fe(III)- and Mn(III)-corroles are active catalysts in H_2O_2 -mediated asymmetric sulfoxidation of prochiral thio-ethers (up to 74% ee) [26]. This work was based on the earlier finding that the sodium salts of di-, tri-, and tetrasulfonic acid derivatives of porphyrins, phthalocyanines, and corroles bind strongly to serum albumins in a noncovalent manner [27]. Moreover, in the case of human serum albumin (HSA) harboring protoporphyrin dimethyl ester, an X-ray crystal structure was obtained, revealing supramolecular bonding in the subdomain IB with weak axial coordination by Tyr161 [28]. This information served as a guide in developing an artificial Cu(II)-based metalloenzyme as the catalyst for asymmetric Diels–Alder reactions, specifically by using the commercially available water-soluble amphiphilic phthalocyanine **1** as the achiral transition metal catalyst (Figure 7.1) [30].

Guided by the crystal structure of Fe-protoporphyrin dimethyl ester in HSA [30], a model for the binding mode of **1** was proposed, which reveals a cavity directly opposite to the Cu(II) where reactions such as Diels–Alder cycloadditions could occur (Figure 7.2) [29].

This artificial metalloenzyme was then used as the catalyst in asymmetric Diels–Alder reactions of azachalcone **2a** with cyclopentadiene (**3**) leading to adduct **4a** (Scheme 7.3). Up to 93% ee with an endo/exo ratio of 95:5 was



Scheme 7.3 Diels–Alder reaction of azachalcones **2** with **3** leading to endoperoxides **4** [29]. (Ohashi *et al.* [29]. Reproduced with permission of John Wiley and Sons.)

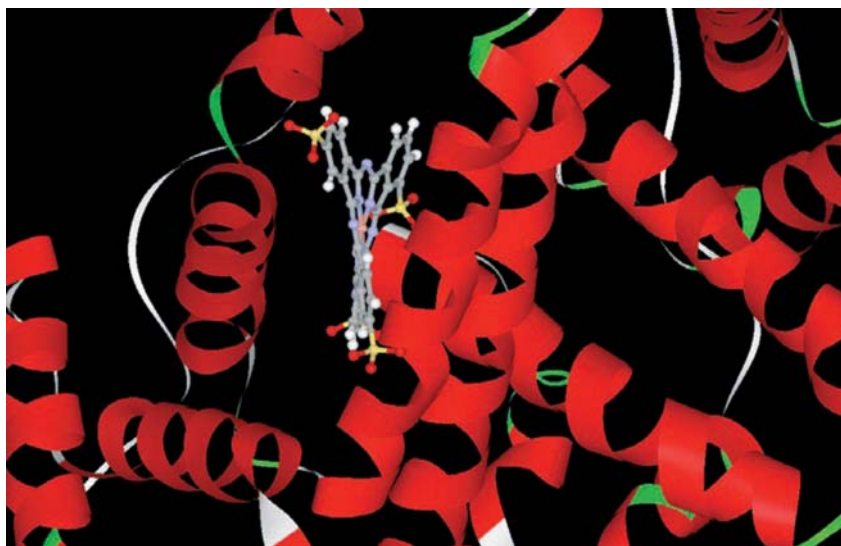


Figure 7.2 Model of 1-HSA [29] based on the crystal structure of human serum albumin (HSA) harboring Fe-protoporphyrin dimethyl ester [30].

observed under optimized reaction conditions (Table 7.1) [29]. The other substrates **2b–e** reacted similarly with enantioselectivities in the range 85–98% ee. Mechanistically, it was postulated that the lone electron pair of the carbonyl O-atom and that of the pyridine N-atom form a chelate to Cu(II) of the catalytic center, thereby causing activation by means of LUMO-lowering [29]. Although the Diels–Alder reaction proved to be somewhat slow, the system seems to be ideal

Table 7.1 Optimization of the Diels–Alder reaction of **2a** with **3** in the presence of Cu(II)-salts and BSA in water.^{a)}

Entry	Cu(II)-salt	Buffer (pH)	Conversion (%)	<i>endo/exo</i>	% ee of <i>endo-4a</i>
1	1 (crude) ^{b)}	MOPS (6.5) ^{c)}	30	92 : 8	64
2	1 (crude)	Acetate (5.0)	74	86 : 14	81
3	1 (crude)	Formate (4.5)	74	93 : 7	82
4	1 (crude)	Formate (4.0)	80	96 : 4	93
5	1 (crude) ^{d)}	Formate (4.0)	80	95 : 5	89
6	1 ^{e)}	Formate (4.0)	78	96 : 4	93

a) Purchased from Sigma [A-6793/Fraction 5]; **2a**:**3**:BSA:Cu(II) = 50 : 300 : 1.2 : 1; 20 mM scale; concentration of BSA is 0.3 mM in buffer (30 mM); 3 °C; **3**.

b) Purchased from Aldrich [27360-85-6] as a mixture of regioisomeric tetrasulfonic acid sodium salts.

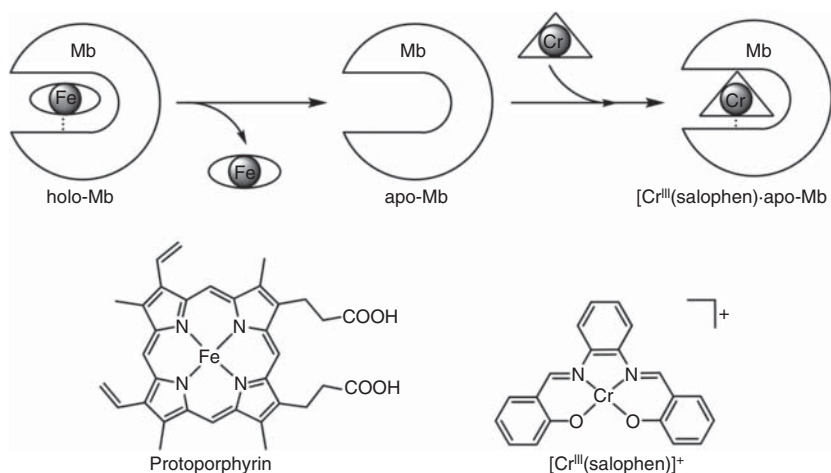
c) MOPS = 3-[*N*-morpholino] propane sulfonic acid.

d) Reaction performed at 5 °C.

e) Purchased from Aldrich [123439-80-5] as the 3,4',4'',4'''-tetrasulfonic acid sodium salt.

for further improvement and substrate scope tuning by directed evolution. The use of other transition metals for a variety of other reaction types can be envisioned.

Covalent bonding to the metal of porphyrin-type or structurally related transition metal complexes can also be envisioned in appropriate protein hosts, possibly supported by additional noncovalent interactions. One of several examples is the Watanabe system in which an achiral chromium(III) Schiff base catalyst was anchored to apo-myoglobin (Scheme 7.4) [29]. This artificial metalloenzyme was used as the catalyst in the asymmetric sulfoxidation of thioanisole (13% ee). Again, this is a case for directed evolution, especially if structural information could be obtained by X-ray crystallography.



Scheme 7.4 Schematic representation for the generation of a myoglobin-based chromium-salen hybrid catalyst for the asymmetric sulfoxidation of thioanisole [31]. (Carey *et al.* [31a]. Reproduced with permission of American Chemical Society.)

Yet another potential case for directed evolution pertains to the Lu-system in which apo-myoglobin is used again, but contains a designed covalent anchor between a manganese Schiff base complex and two cysteines introduced by site-directed mutagenesis (Figure 7.3) [31]. This artificial metalloenzyme proved to be more active and enantioselective in the sulfoxidation of thioanisole (52% ee).

Covalent bioconjugation of proteins by way of C–S bond formation at a cysteine residue for a variety of purposes has been practiced for decades by S_N2 alkylation or Michael addition. This traditional technique was applied a number of times in order to anchor achiral ligand/metal moieties (or organocatalysts) in protein hosts [21b,c, 25]. When aiming for this kind of artificial metalloenzyme, it is best to choose a host that is thermally robust and capable of tolerating the reaction conditions used in traditional aqueous transition metal catalysis [32]. An excellent protein host for this kind of endeavor is an enzyme called *tHisF* from the thermophilic organism *Thermotoga maritima*, which is involved in the biosynthesis of histidine [33]. It can be heated in aqueous phase at 75 °C

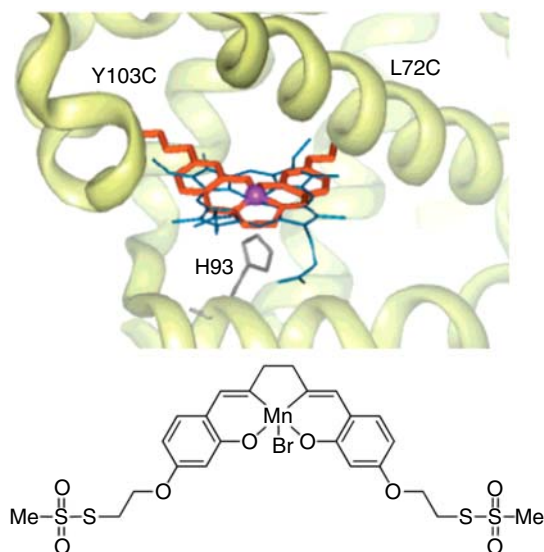
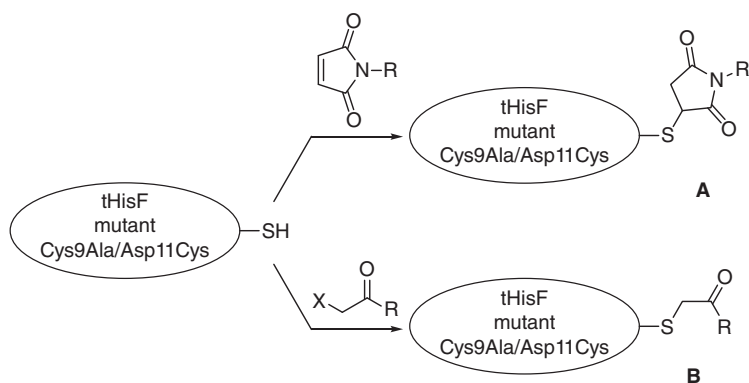


Figure 7.3 Manganese Schiff base complex introduced into apo-myoglobin [32]. (Cornils and Herrmann [32]. Reproduced with permission of John Wiley & Sons.)

for 1 h without any sign of denaturing. Moreover, expression in *Escherichia coli* is unusually efficient, and its X-ray structure has been analyzed [34]. tHisF has a barrel-like structure with one cysteine located fairly deep in the channel (Cys9). In order to enable smooth site-specific covalent anchoring, a double mutant Cys9Ala/Asp11Cys was created, the reactive cysteine needed in bioconjugation being just below the rim of the barrel-like structure [25]. Appropriate Michael additions and S_N2 reactions were applied in order to anchor ligands, ligand/metal entities, and even organocatalysts (Scheme 7.5) [25]. A platform for

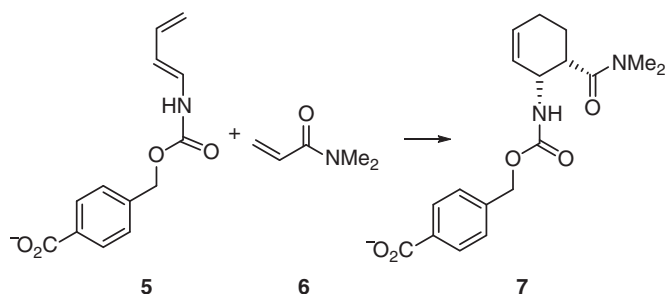


Scheme 7.5 Chemical modification of tHisF mutant Cys9Ala/Asp11Cys by means of Michael additions that lead to bioconjugates A, and S_N2 -reactions that provide bioconjugates B [25]. (Reetz *et al.* [25]. Reproduced with permission of John Wiley and Sons.)

potential directed evolution was developed, which includes miniaturization and parallelization of fermentation and mutant enzyme purification by simple heat treatment of 24-format deep-well plates. This means that all foreign protein is denatured and precipitated from solution, leaving behind tHisF for clean and selective bioconjugation. It would be interesting to apply directed evolution using a variety of different transition metal catalyzed transformations.

Developing artificial metalloenzymes by designing transition metal binding sites using His/Cys or other motifs inspired from natural proteins with the help of site-directed mutagenesis (Scheme 7.2) has a number of advantages, because it does not require a bioconjugation step. One example of this approach pertains to the use of the thermostable tHisF as the protein, in which a Cu(II)-binding site was implemented by designing mutant Asp11/His50/His52 [35]. The apo-form was treated with CuSO_4 leading to the formation of the desired artificial metalloenzyme. It was characterized by EPR, proving that two histidines are indeed involved in metal complexation. The catalyst was used in the asymmetric Diels–Alder reaction shown in Scheme 7.3, leading to an enantioselectivity of 46% ee and an endo/exo ratio of 13:1 [36]. Directed evolution of the catalyst in this asymmetric transformation or in other Cu(II)-catalyzed reactions has not been attempted to date.

Impressive examples of computational design of artificial metalloenzymes and/or promiscuous enzymes in general based on the Rosetta algorithms have continued to appear in the literature as documented in a review article [36]. The multi-step computational procedure utilizes, inter alia, *ab initio* protein structure prediction, quantum mechanics (QM) energy refinement, and sequence design. Examples of the Rosetta-approach are designed proteins that catalyze the Kemp elimination, the Diels–Alder cycloaddition, and the Baylis–Hillman reaction. In the case of the Diels–Alder reaction shown in Scheme 7.6, the endo-(3*S*,4*R*)-configuration in adduct 7 was the desired outcome as stipulated by design [38]. From a collection of 84 computed and then in *E. coli* expressed enzymes, two showed Diels–Alder activity, one of them leading to the predicted stereoselectivity. In the case of the Baylis–Hillman reaction, the results appear to be less successful, no information regarding the enantioselectivity of the asymmetric



Scheme 7.6 Model Diels–Alder cycloaddition used in Rosetta-design [37]. (Bjelic *et al.* [37]. Reproduced with permission of American Chemical Society.)

transformation being reported [37]. Directed evolution was not attempted in this case, in contrast to the project focusing on the Kemp elimination (see Section 7.2).

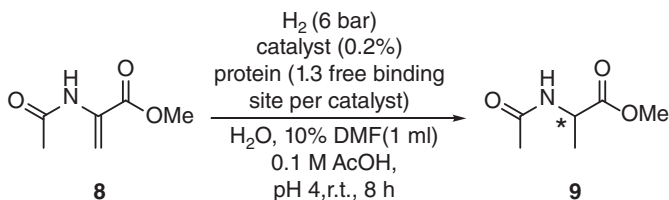
Along a different line, promiscuous enzymes in unculturable bacterial communities promise to be a rich source of new biocatalysts, but their discovery by functional metagenomics has proven to be problematic. In order to solve this problem, the technique of microfluidic picoliter oil-in-water droplets (Chapter 2) has been applied [39]. Million-membered metagenome libraries can be assessed in this way. This advancement connects enzymes with distantly related sequences. It was found that most of the hits could not have been identified or predicted solely on the basis of sequence data.

7.2

Tuning the Catalytic Profile of Promiscuous Enzymes by Directed Evolution

As outlined in Section 7.1, the term promiscuous behavior has been applied in some studies when engineering an expanded substrate scope, which leaves behind the misleading impression that directed evolution is used routinely in the area of promiscuity. The majority of researchers performing directed evolution for manipulating substrate acceptance generally do not use this terminology, a convention that is adhered to in this monograph. Directed evolution of substrate acceptance (or enhancement of activity of a sluggish enzyme) is treated in Chapter 5.

As pointed out in Chapters 3 and 4, saturation mutagenesis and iterative saturation mutagenesis (ISM) at sites lining the active site as part of the combinatorial active-site saturation test (CAST) constitute reliable tools for reshaping the binding pocket of enzymes in the quest to manipulate activity, stereo-, and/or regioselectivity. This was not foreseen when the concept of directed evolution of artificial metalloenzymes (hybrid catalysts) was proposed [23]. At the time, mainly error-prone polymerase chain reaction (epPCR) and DNA shuffling were used for tuning stereoselectivity by directed evolution [40]. The first example of directed evolution of an artificial metalloenzyme as catalyst in a stereoselective transformation made use of the Whitesides system (Section 7.1), but in this case utilizing the ester of *N*-acyl acrylic acid because its reduction product is easier to extract from the aqueous phase (Scheme 7.7) [21].



Scheme 7.7 Model reaction used in the directed evolution of the Whitesides system [21]. (Reetz *et al.* [21a]. Reproduced with permission of Royal Society of Chemistry.)

At the time of this study, several problems hampered the envisioned application of laboratory evolution. The major difficulty was the fact that all available expression systems of streptavidin were not very efficient. It was not possible to produce enough protein in the conventional small wells of 96-format micro-titer plates. Therefore, expression, bioconjugation, and reaction were performed in 500 ml Erlenmeyer flasks! As a consequence, only very small mutant libraries were possible [21]. In order to identify potential randomization residues for saturation mutagenesis, the biotinylated Rh-diphosphine complex was modeled in streptavidin, revealing several residues about 4–5 Å from the Rh-center (Asn49, Leu110, Ser112, and Leu124) as well as second sphere CAST sites Glu51, Tyr54, Trp79, Asn81, Arg84, Asn85, and His87 (Figure 7.4). WT streptavidin delivers an ee-value of only 23% in favor of (*R*)-**9**.

A limited number of very small libraries in a restricted ISM process provided the results shown in Scheme 7.8. It can be seen that ISM leads stepwise to clearly improved enantioselectivity (65% ee), and that even reversal of enantioselectivity is possible. This study constitutes proof-of-principle regarding the use of directed evolution as a means to tune the stereoselectivity of promiscuous metalloenzymes, but it is far from being practical [21]. It is likely that “normal-sized”

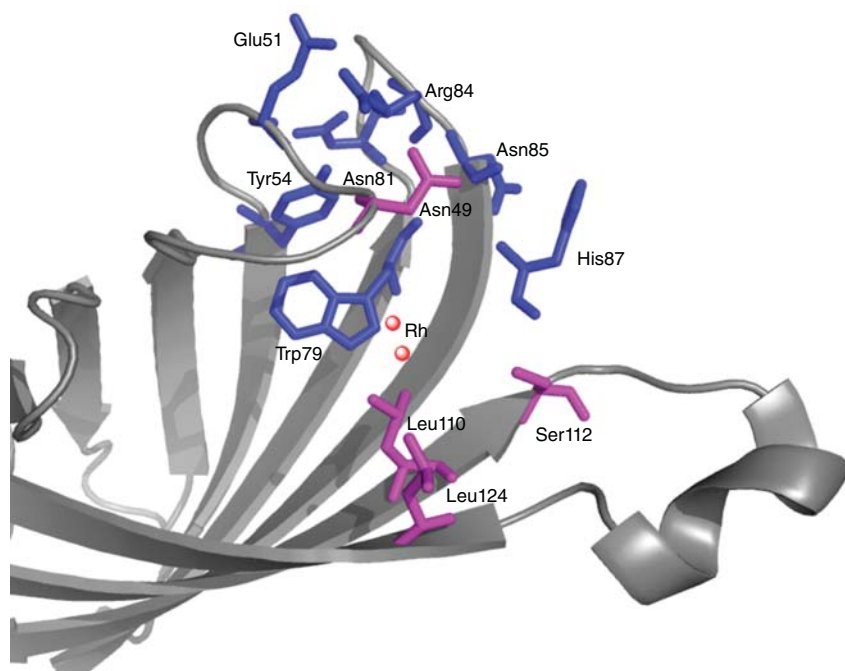
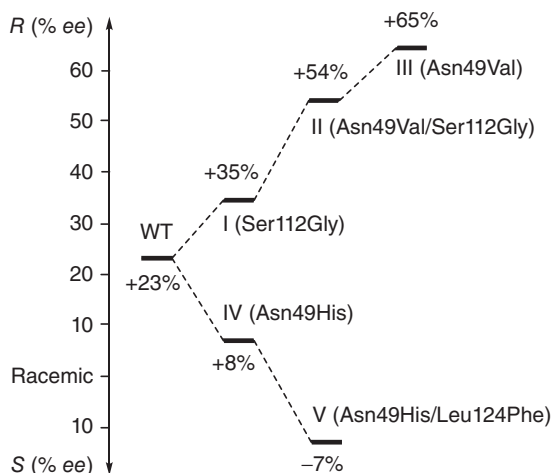


Figure 7.4 Model of the biotinylated diphosphine-Rh-complex in streptavidin [21]. (Reetz *et al.* [21a]. Reproduced with permission of Royal Society of Chemistry.)

saturation mutagenesis libraries amounting to a thousand or so transformants would provide high stereoselectivity.

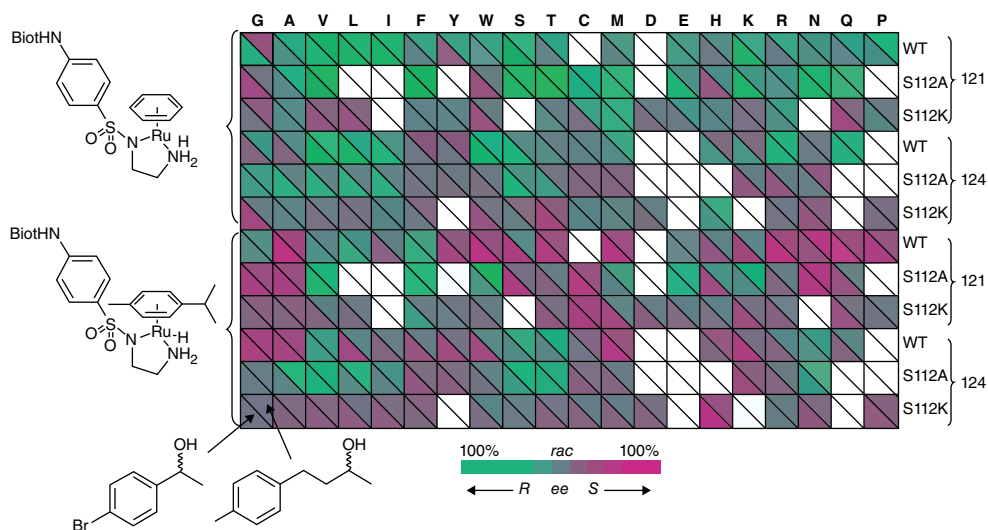


Scheme 7.8 Directed evolution of stereoselectivity of a promiscuous enzyme based on the Whitesides system, iterative saturation mutagenesis (ISM) being employed as the genetic tool and the Rh-catalyzed

hydrogenation of substrate **8** with formation of **9** serving as the model reaction [21]. (Reetz *et al.* [21a]. Reproduced with permission of Royal Society of Chemistry.)

The Whitesides system has been extended to include many different transition metal catalyzed stereoselective transformations, generally by optimizing the spacer length between biotin and the ligand/metal entity and applying CAST saturation mutagenesis [41]. In a study focusing on Noyori-type transfer hydrogenation of prochiral ketones, the initial CAST libraries did not provide sufficiently improved variants and, therefore, ISM was employed [42]. Several achiral Ru-complexes were biotinylated with variation of the spacer length, and first tested as catalysts in the asymmetric reduction of seven different prochiral ketones. The fingerprint display of two of these substrates, *p*-bromoacetophenone and 4-phenyl-2-butanone is displayed in Scheme 7.9 [43].

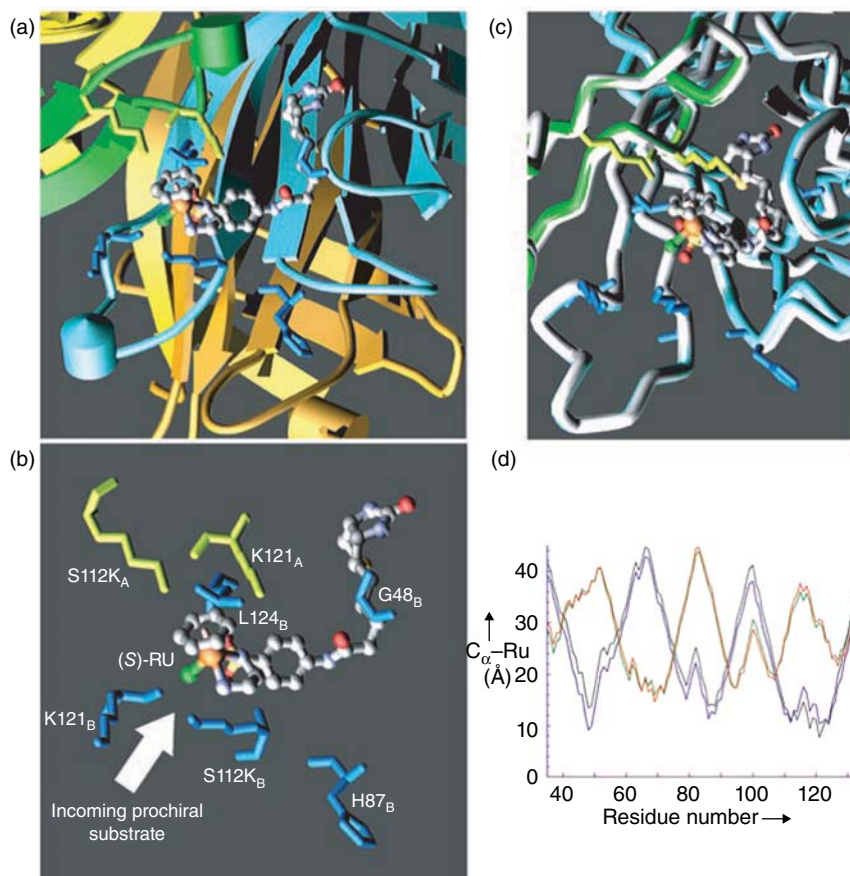
In order to enhance enantioselectivity, two different piano stool Ru-complexes were employed in ISM experiments, leading to the structural results summarized in Scheme 7.10 [43]. One of the fascinating results of this study concerns the crystal structure of a mutant, the metal (Ru) itself being a center of chirality. Thus, the stereoselective outcome is not solely due to the immediate chiral protein environment, although it induces chirality at the metal. Another embodiment of the Whitesides streptavidin system concerns regioselective Rh(III)-catalyzed CH-activation, in which site-specific mutagenesis was successfully employed in order to enhance activity [43]. This is a crucial issue, because promiscuous metalloenzymes generated by the different approaches as highlighted here are generally characterized by low activity.



Scheme 7.9 Fingerprint display of the results for the chemogenetic optimization of the reduction of ketones 1 and 3 in the presence of biotin-sepharose-immobilized artificial metalloenzymes $[\eta^6\text{-(arene)RuH(Biot-p-L)}]\text{C Sav mutant}$. Catalytic runs which could not be performed (insufficient soluble protein expression are represented by white triangles) [43]. Substrates, reduction products, and operating conditions used for the designed evolution of artificial transfer hydrogenases. $\eta^6\text{-arene}$ = benzene,

p-cymene; Sav mutant: K121X, L124X, S112A K121X, S112K K121X, S112A L124X, S112K L124X. The catalytic runs were performed at 558° C for 64 h using the mixed buffer NaO_2CH (0.48 m), B(OH)_3 (0.41 m), and 3-(*N*-morpholino)propanesulfonic acid (MOPS, 0.16 m) at pH initial 6.25. Ru/substrate/formate ratio 1 : 100 : 4000. (Adapted and reproduced from Hyster *et al.* [43]. Reproduced with permission of The American Association of the Advancement of Science.)

In the attempt to convert the esterase from *Pseudomonas fluorescens* (PFE) into an epoxide hydrolase, its mechanism was compared to that of the epoxide hydrolase from *Agrobacterium radiobacter* (EchA) (Scheme 7.11), which formed the basis for subsequent directed evolution experiments [8]. Sequence data of esterases and epoxide hydrolases were consulted, as was the structural comparison of six epoxide hydrolases. Several rationally designed mutants were generated, for example, by switching serine to aspartate or by introducing two tyrosines, but these showed no activity in the hydrolysis of styrene epoxide. Finally, the whole loop in a PFE mutant containing mutations L29P, F93H, S94D, F125Y, V139Y, and V195Y was replaced by the corresponding element in EchA by a polymerase chain reaction (PCR)-based procedure. This provided a variant that could be expressed as a soluble protein in *E. coli*, showing low epoxide hydrolase activity (initial activity of 9 mU mg^{-1} and a turnover number of 0.01 s^{-1}). Unfortunately, it was not possible to measure K_m nor $V_{\max}$ values. This study underscores the challenge in applying directed evolution to this type of enzyme promiscuity [8]. In other cases, it appears to be much easier, as in the

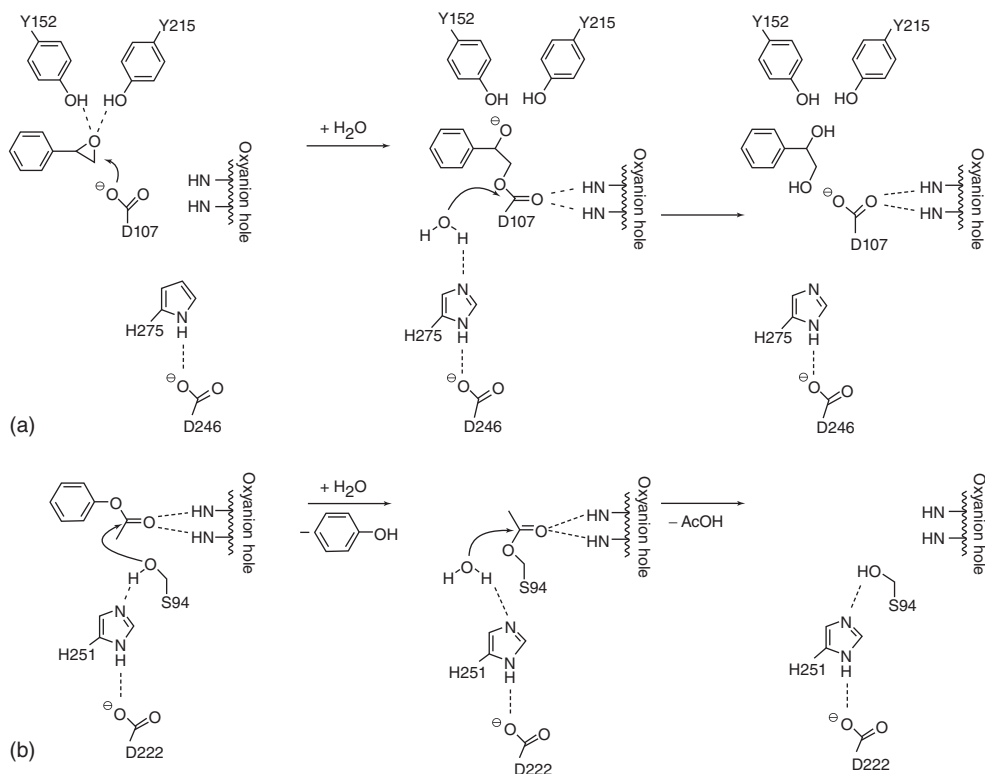


Scheme 7.10 X-ray crystal structure of $[\eta^6\text{-(benzene)RuCl(Biot-}p\text{-L)}] \subset \text{S112K Sav}$. (a) Close-up view (only monomer B (blue) occupied by the biotinylated catalyst (ball-and-stick representation); monomers A (green), C (orange), and D (yellow)). (b) Highlight of amino acid sidechain residues displaying short contacts with Ru. The absolute configuration at ruthenium is S. (c) Superimposition of the structure of $[\eta^6\text{-(benzene)RuCl(Biot-}p\text{-L)}] \subset \text{S112K Sav}$ with the structure of

biotin $\subset$ core streptavidin (PDB reference code 1STP, only monomers A and B displayed for clarity; biotin: white stick, core streptavidin: white tube). (d) Ru–Ca distances extracted from the X-ray structure of $[\eta^6\text{-(benzene)RuCl(Biot-}p\text{-L)}] \subset \text{S112K Sav}$; monomers: A, black; B, blue; C, green; and D, red [43]. (Adapted and reproduced from Hyster *et al.* [43]. Reproduced with permission of The American Association of the Advancement of Science.)

highly successful conversion of a phosphotriesterase into an aryylesterase [9]. This may be because the respective enzyme mechanisms share greater similarity.

Switching cofactor dependency from NADPH to NADH, or reverse, has been accomplished several times (Table 5.1). Usually the goal is to exchange the more expensive NADPH by NADH. In one study concerning cofactor switch, a change



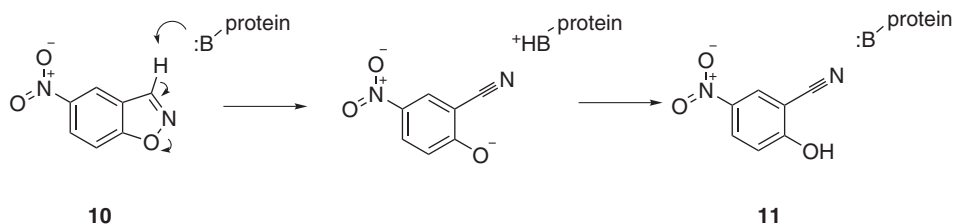
Scheme 7.11 Mechanistically essential amino acid residues in *Agrobacterium radiobacter* epoxide hydrolase (EchA). (a) Formation and liberation of the alkyl enzyme intermediate derived from styrene oxide as substrate and *Pseudomonas fluorescens* esterase (PFE).

(b) Formation and liberation of the acetyl enzyme intermediate derived from phenyl acetate as substrate [8]. (Jochens *et al.* [8]. Reproduced with permission of John Wiley & Sons.)

in enantioselectivity in sulfoxidation was unexpectedly observed upon introducing mutations by site-directed mutagenesis and saturation mutagenesis, the result being termed “*cofactor promiscuity*” [44]. The flavoprotein monooxygenase from *Stenotrophomonas maltophilia* (SMFMO) is unusual because it can utilize either the non-phosphorylated cofactor NADH or NADPH for the reduction of the FAD coenzyme, specificity being 1.5 : 1 in favor of NADH. Following protein engineering, the specificity switched to 1 : 3.5 in favor of NADPH. Seven different prochiral thio-ethers were tested using the evolved mutant, resulting either in reduction or in reversal of enantioselectivity. Although practical applications did not result from this study, it shows that switching cofactor specificity can influence stereoselectivity in a manner that is currently not well understood [44].

Directed evolution has also been applied in the successful attempt to boost the activity of a designed promiscuous enzyme called *Kemp eliminase*, which had been produced by computational design on the basis of the Rosetta algorithm [45].

The Kemp elimination [46] is a synthetically trivial base-catalyzed transformation leading to ring-opened products of benzisoxazoles, but much can be learned from designing protein-mediated catalysts. A review is available, which outlines the benchmarks set by the different approaches in the race to generate the most active Kemp eliminase, the particular transformation **10** $\rightarrow$ **11** serving as the standard model reaction (Scheme 7.12) [47].



Scheme 7.12 Protein-catalyzed Kemp elimination **10** $\rightarrow$ **11**.

One of the best original Rosetta design motifs predicted a TIM barrel scaffold of HisF with glutamine serving as the base (Figure 7.5) [45b]. Following the detection of very low activity, directed evolution was applied. After seven rounds of epPCR and DNA shuffling, a mutant characterized by eight point mutations was evolved, showing a 200-fold increase in catalytic efficiency ($k_{\text{cat}}/K_m = 2600 \text{ M}^{-1} \text{ s}^{-1}$) [45a,b].

Considering the simplicity of Kemp-type elimination with many common bases being highly active in the absence of any protein, the catalytic efficiency achieved by this eliminase is far from optimal despite extensive design and protein engineering. Therefore, a second attempt was undertaken to further improve activity [45c]. One of the earlier designed Kemp eliminases based on the TIM barrel scaffold of the deoxyribose phosphate aldolase of *E. coli*, KE70, was subjected to nine rounds of directed evolution, resulting in the presently known best protein-based catalyst ($k_{\text{cat}}/K_m = 5 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$). This extensive experimental effort (Table 7.2) was flanked by computational design including ensemble generation, β -strand perturbations, $\text{p}K_a$ modulation, and loop redesign. The best mutant was characterized structurally and biochemically, which led to the conclusion that (i) the active-site cavity was reshaped to achieve tighter substrate binding, (ii) the electrostatics around the catalytic dyad His-Asp was optimized, and (iii) the active-site dyad was stabilized in a conformation, which is optimal for catalysis [45c]. Thus, these are valuable mechanistic lessons learned as a consequence of such an undertaking, which underscore the primal characteristics of natural enzyme catalyzed reactions as involving precisely positioned substrates as defined by the Jencks hypothesis [48] and electrostatic stabilization of the transition state according to the Warshel hypothesis [49]. The results are indeed impressive, because catalytic efficiency begins to approach that of enzyme-catalyzed natural transformations that are known to be in the range $k_{\text{cat}}/K_m = 10^5 - 10^8 \text{ M}^{-1} \text{ s}^{-1}$. Nevertheless, there is still plenty of room for further rate enhancement.

Table 7.2 Summary of the directed evolution of the Kemp eliminase KE70 [45b].

Round	Random mutagenesis	Recombination	Mutation spiking by ISOR	Fold improvement measured with crude lysates ^{a)}
1	2 ± 1 random mutations per gene	—	—	≤6-fold relative to designed KE70; best variant is R1 8/9C
2	—	Shuffling of the 15 best variants from Round 1	—	≤1.5-fold relative to R1 8/9C; best variant is R2 7/12F
3	—	—	Shuffling of the 15 best variants from Round 2, with incorporation of designed mutations: Library 1 (design categories 1a + 2) – Met161Leu/Leu/Val/Phe, Leu181Ile/Leu/Val/Phe, Trp72Ser/Cys/His/Leu, Gly101Glu/Gln/Ala/Ser, Ser138Ala, His166Tyr/Asp/Asn/Ala/Ser; Library 2 (design category 4a) – insertions after the residues Thr20 (Gly/Ser), Asn22 (Gly/Ser), Thr171 (Asn/Ala/Pro/Gly/Ser), Val204 (Ala/Pro/Gly/Ser), and Ser239 (Asn/Ala/Pro/Gly/Ser)	≤3-fold relative to R2 7/12F; best variants are R3 2/6D (Library 1) and R3 9/3B (Library 2)

4	—	Shuffling of the 18 best variants from Round 3, both from Libraries 1 and 2	—	≤3-fold relative to R3 9/3B; best variants are R4 4/1B and R4 4/5B
5	—	Shuffling of the 10 best variants from Round 5	Shuffling of the 12 best variants from Round 4, with incorporation of designed mutations: design category 1b – Ser74Ala/Gly, Phe77Tyr, Leu136Trp, Ala178Ser, Lys173Asn/Thr, Ala231Ser, Ala238Ser, Ser239Thr/Asn/His/Arg; design category 4a – Ala21Asn/Gln/Arg, Asn22Gln/Arg	≤3-fold relative to R4 4/1B; best variant is R5 7/4A
6	—	Shuffling of the 10 best variants from Round 5	—	≤1.2-fold relative to R5 7/4A; best variants are R6 6/10A and R6 4/8B

(continued overleaf)

Table 7.2 (Continued)

Round	Random mutagenesis	Recombination	Mutation spiking by ISOR	Fold improvement measured with crude lysates ^{a)}
7	Four best variants from Round 5 and four variants from Round 6; 2 ± 1 mutations per gene	—	—	≤1.2-fold relative to R6 6/10A; best variant is R7 7/1C
8	—	—	Shuffling of the 14 best variants of Round 7, with the incorporation of designed mutations: design category 2 – Met16Ala/ Val, Leu18Ile/Val/Leu/Phe; design category 4b – Ala238Met, Ser240Gly, Leu241Ala	≤1.2-fold relative to R7 4/2E and 3/2B; best variants are R8 12/12B and R8 15/11E
9	—	—	—	≤1.2-fold relative to R8 12/12B and R8 15/11E

a) The activity improvement measured in crude lysates is not corrected for protein expression and is therefore only a preliminary measure for an increase in protein activity.
(Adapted and reproduced from Khersonsky *et al.* [45b]. Reproduced with permission of Elsevier.)

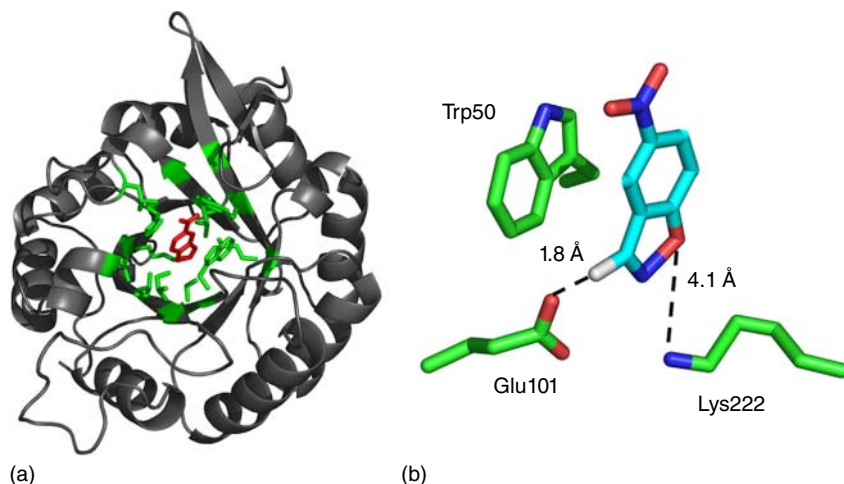


Figure 7.5 (a) The KE07 design, showing the TIM barrel scaffold of HisF (PDB accession code 1THF), the modeled 5-nitrobenzisoxazole substrate (red), and the 13 residues that were replaced to create the designed Kemp eliminase active site (green). (b) Details of the active site of the designed

KE07. Shown are the 5-nitrobenzisoxazole substrate (cyan), the catalytic base (Glu101), the general acid/H-bond donor (Lys222), and the stacking residue (Trp50) [45b]. (Adapted and reproduced from Khersonsky *et al.* [45b]. Reproduced with permission of Elsevier.)

Following these reports, a considerably shorter approach was undertaken, which also involved directed evolution, but was accompanied by less mutagenesis efforts [50]. In this case, a noncatalytic calmodulin scaffold was employed, a fairly small (16 706 Da, 148 amino acid) regulatory binding protein (CaM). In a simple and fast computational design based on the interplay of protein folding and functional group tuning, CaM was converted into a Kemp eliminase called *AlleyCat*. It was truncated in the process (74 amino acids), only the C-terminal domain of CaM being utilized. It showed (low) activity in the reaction of the standard substrate **10** → **11** (Scheme 7.12). A single point mutation was shown to be essential [50a]. In order to boost activity, directed evolution was subsequently applied [50b]. First, the N-terminal domain was re-introduced in order to enhance protein expression efficiency in *E. coli*. This protein was shown to have the same activity as the originally designed *AlleyCat*. Then eight positions lining the binding pocket were subjected individually to saturation mutagenesis using NNK codon degeneracy encoding all 20 canonical amino acids. Several variants displayed more than a sixfold increase in enzymatic efficiency. This was followed by one round of DNA shuffling and two subsequent epPCR cycles. The N-terminal domain was then removed, affording the catalyst *AlleyCat7*. A total of only seven rounds of mutagenesis and screening less than 6000 transformants were required, leading to a catalytic efficiency of $k_{\text{cat}}/K_m = 1283 \text{ M}^{-1} \text{ s}^{-1}$ [50b]. The simple introduction of glutamine at the correct position was shown to be responsible for activity. Interestingly, it was shown that the Kemp eliminase

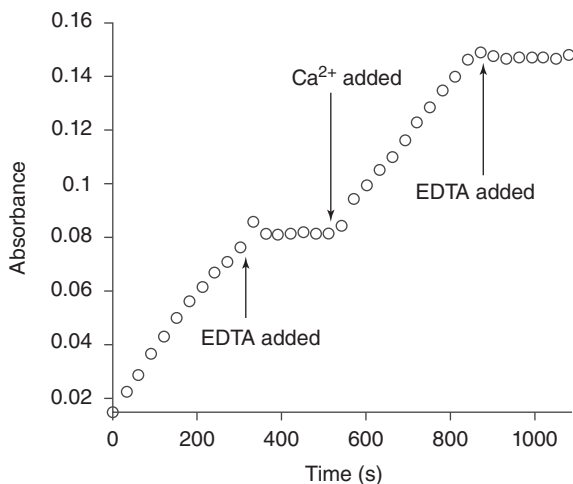


Figure 7.6 Allosteric regulation of AlleyCat7. Experimental conditions: initial concentrations: 130 nm protein, 100 mm NaCl, 20 mm HEPES buffer, pH 7.0, 0.1 mm CaCl_2 , 0.1 mm substrate. At 300 s EDTA (ethylenediaminetetraacetic acid) was added to the final

concentration of 0.2 mm, followed by addition of CaCl_2 at 540 s to the final concentration of 0.3 mm and, again, EDTA at 840 s to the final concentration of 0.5 mm. (Moroz *et al.* [50b]).

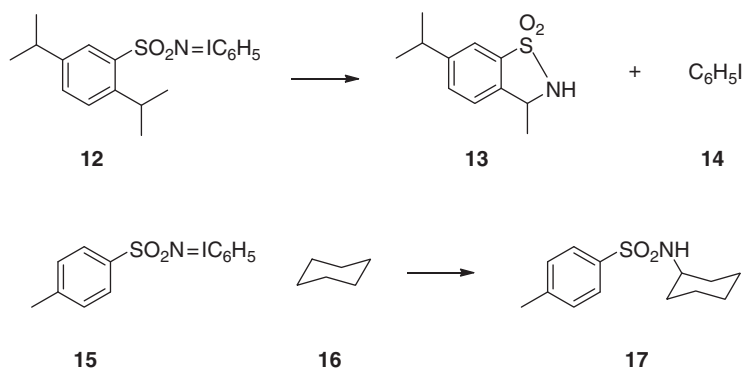
AlleyCat 7 can be cycled through the on and off states, which does not result in any activity loss or product inhibition (Figure 7.6) [50b]. This approach shows that a fairly simple strategy can be successful, but the final result in terms of catalytic efficiency does not match the best Kemp eliminases generated by a more elaborate and labor-intensive procedure [45c]. It can be concluded that more research is necessary, which entails less effort but higher catalytic efficiency. If successful, valuable new insights can be expected to emerge.

The world record concerning activity of a Kemp eliminase was set by applying design and extensive directed evolution using a xylanase as scaffold [51]. Starting with the mutant HG3 designed and evolved previously, 17 further rounds of mutagenesis were performed using epPCR, DNA shuffling, saturation mutagenesis at hot spots identified by epPCR, and ISM. This provided variant HG3-17 showing a catalytic efficiency of $k_{\text{cat}}/K_m = 230\,000\text{ M}^{-1}\text{ s}^{-1}$ [51]. On the basis of X-ray structural analysis, it was concluded that precise positioning of the catalytic machinery with respect to the substrate is essential for approaching typical natural enzyme activity.

Several other studies describing the use of directed evolution in the quest to increase promiscuous enzyme activity have appeared, in addition to the report describing directed evolution for the functional analysis of multi-specific proteins [52]. Although involving “only” substrate acceptance (which is treated in Chapters 4 and 5), a study describing the directed evolution of a nucleotidyltransferase deserves mention here. In the attempt to expand, the nucleotide and sugar

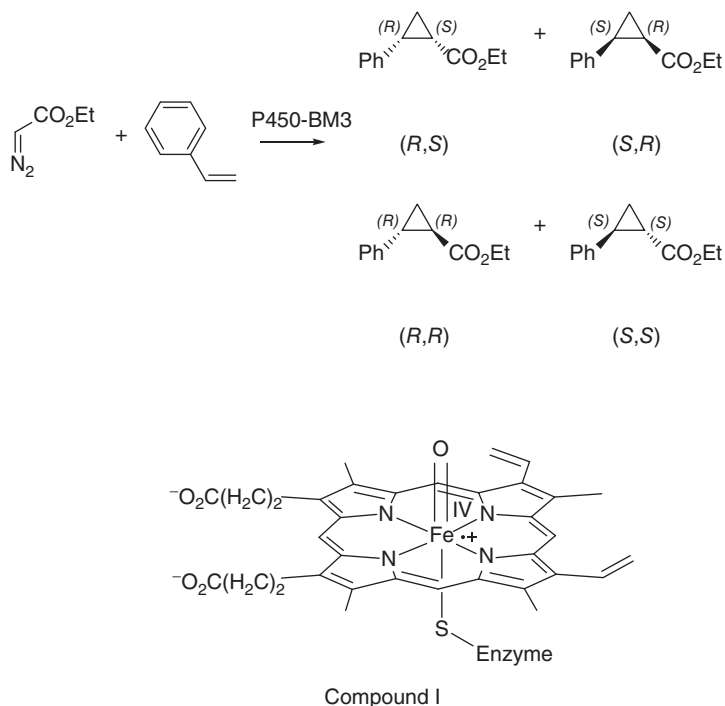
1-phosphate promiscuity of the nucleotidyltransferase RmIA, a combination of epPCR and saturation mutagenesis was employed [52]. Based on structural data, 22 residues lining the binding pocket (CAST sites) were selected for individual NNK-based randomization. The final variants were shown to accept non-native substrates such as pyrimidine and purine-based nucleotides as well as nonnative D- and L-sugars including α - and β -isomers, which has obvious practical implications.

Finally, in a seminal study 30 years ago by Dawson and coworkers, it was shown that P450 enzymes show promiscuous reactivity by catalyzing inter- and intramolecular insertion of nitrenes into nonactivated C–H bonds [53]. The source of the *N*-sulfonyl-nitrenes were compounds of the type $\text{ArS(O)}_2\text{N=I-C}_6\text{H}_5$, which were transferred onto Fe(II) as metal carbenes. The respective Fe-bound nitrenes underwent C–H activating amidations (Scheme 7.13).



Scheme 7.13 P450-catalyzed insertion of nitrenes into nonactivated C–H bonds [53].

Mutations for tuning activity were not introduced into the enzyme, but this report set the stage for systematic studies of promiscuous P450 monooxygenases, which in turn provided intriguing opportunities. A different cytochrome P450 monooxygenase, specifically P450-BM3, was genetically modified so that several different reaction types became amenable to catalysis that traditionally belong to the realm of conventional transition metal catalysis. This new and fast growing research area has already been reviewed [13, 54]. It is also reminiscent of other transition metal porphyrin-type promiscuous biocatalysts (Figures 7.2 and 7.3, Scheme 7.4), and yet of another study [55], but goes beyond these previous approaches. The first of a series of studies in this area focused on Fe-catalyzed carbene transfer reactions using ethyl diazo acetate as the carbene source and styrene as the olefinic substrate with formation of the respective cyclopropane derivative (Scheme 7.14, top) [56], analogous to the Fe-nitrene formation [53]. It was speculated that carbene transfer could occur in analogy to the known P450-BM3 catalyzed epoxidation via oxygen transfer by the catalytically active species (Compound I) (Scheme 7.14).



Scheme 7.14 Promiscuous reactivity of P450-BM3. Top: Envisioned catalytic cyclopropanation; below: Intermediate (Compound I) known to be the active species in the catalytic P450-BM3 catalyzed epoxidation of olefins, serving as a conceptual guide in devising Fe-catalyzed cyclopropanation [56]. (Coelho *et al.* [56]. Reproduced with permission of American Association for the Advancement of Science.)

Prior to this report, it had been shown that numerous porphyrin-based transition metal complexes are active catalysts in a variety of different transition metal catalyzed reactions. For example, an achiral porphyrin-Fe(II) complex was known to catalyze the cyclopropanation reaction depicted in Scheme 7.14 with formation of all stereoisomers, provided an oxygen-free system was applied [55]. In the presence of air, Fe(II) is oxidized to Fe(III) which was shown to be inactive. Moreover, in the P450-catalyzed metabolism of 1,3-benzodioxole, the intermediacy of a 1,3-benzodioxole carbenoid at the heme-Fe center had been postulated [57]. In the new approach (Scheme 7.14), air was not excluded, but the reducing agent $\text{Na}_2\text{S}_2\text{O}_4$ was added under aerobic conditions. In spite of some mechanistic uncertainty as pointed out in a highlight [58], WT P450-BM3 proved to be active, but not very stereoselective. Therefore, a library of P450-BM3 variants evolved earlier for other purposes was screened, leading to improved enantio- and diastereoselectivity. Final optimization was accomplished by saturation mutagenesis at some of the sites near the binding pocket. The final variant showed high diastereoselectivity (cis:trans = 92 : 8) and high enantioselectivity (97% ee) in favor of the (*S,R*)-adduct. Several other styrene-type substrates were also subjected to

cyclopropanation, but these reactions occurred with moderate to poor diastereo- and enantioselectivity [56].

It is conceivable that the use of transition metals other than Fe would lead to new reactivities. The researchers chose another strategy by optimizing the axial ligand at heme-Fe, that is, by replacing cysteine by other amino acids, guided by the known electronic influences of ligands in simple porphyrin-Fe catalysis [13, 59]. Amino acids that impart very different promiscuous catalytic properties to P450-BM3 are serine or histidine. In the case of the cyclopropanation reaction, activity was enhanced dramatically, probably because these mutations allowed ready reduction of the heme-Fe to Fe(II) [59]. This also enabled whole cell catalysis of carbene transfer at high rate, which begins to compete with the most active Rh-catalysts. Cyclopropanation was also applied to the formal synthesis of the therapeutic drug levomilnacipran by using a P450-BM3 variant with only five amino acid substitutions [60]. Finally, carbene insertion into aryl N–H bonds using variants of P450-BM3 was also demonstrated [61].

Inspired by the early report on P450-catalyzed nitrene insertions via intermediate Fe-nitrenoids (Scheme 7.14) [53], it appeared logical to test the P450-BM3 system for nitrene transfer reactions [61]. C–H activating amidation was expected, rather than aziridine formation, which would be analogous to the Fe-carbenoid mediated cyclopropanation reactions. Using sulfonyl azides as the nitrene precursors, this was indeed observed. Upon optimizing the structure of the aryl group in the nitrene precursor and testing a variety of P450-BM3 mutants as catalysts, smooth intramolecular amidation was observed with enantioselectivity up to 89% ee [61]. Point mutations Thr268Ala and Cys400Ser were shown to be particularly important, the latter occurring as the axial ligand at heme-Fe. In an independent study, different P450-BM3 mutants, originally evolved by systematic saturation mutagenesis at sites lining the binding pocket of P450-BM3, provided even better results [62]. In a study directed toward switching the regioselectivity of intramolecular C–H activating amidation, six-membered instead of the usual five-membered cyclic sulfonamides were observed as products using a P450-BM3 variant characterized by a dozen point mutations [63].

Along a different and likewise intriguing line, artificial metalloenzymes were devised for olefin metathesis in water [64]. In one study, a Grubbs–Hoveyda type Ru-catalyst was anchored covalently according to Scheme 7.5A to the β -barrel protein nitrobindin, this artificial metalloenzyme mediating olefin metathesis in aqueous medium at an unusually high catalytic rate (Figure 7.7) [64a]. Some mutational effects were observed. In a second study, noncanonical amino acids were incorporated site-specifically into the ribosomal lasso peptide capistruin, which enabled the anchoring of the Grubbs–Hoveyda Ru-catalyst [64b].

7.3

Conclusions and Perspectives

Finding promiscuous enzymes in nature or creating such biocatalysts by designing artificial metalloenzymes or enabling new and unusual reactivities

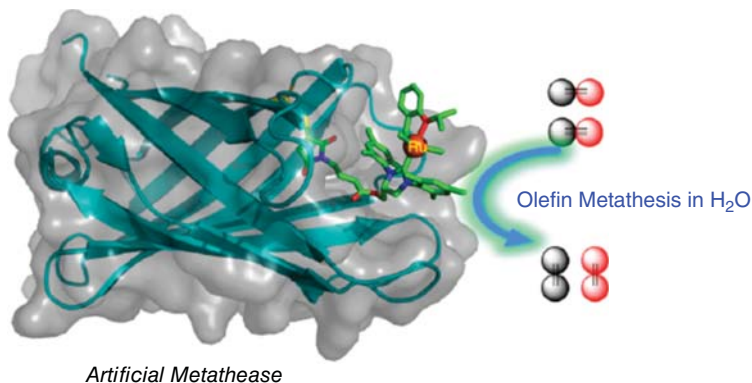


Figure 7.7 An artificial olefin metathese based on anchoring a Grubbs–Hoveyda Ru-catalyst covalently to the protein nitrobindin [64a]. (Sauer *et al.* [64a]. Reproduced with permission of American Chemical Society.)

by computational means continues to be an intriguing research area. In the vast majority of cases studied reported thus far, activity proved to be very low. As shown in Section 7.2, the application of directed evolution provides the potential to remedy the situation [23]. Indeed, a few examples have been reported in which notable enhancement of activity was accomplished, sometimes accompanied by excellent stereo- and/or regioselectivity. On a critical note, it needs to be stated that such cases are the exception [64a]. Moreover, many more improvements are necessary before real (industrial) applications become reality in any of the approaches delineated in this chapter. This pertains especially to artificial metalloenzymes, because these biocatalysts have to be compared to the growing number of efficient man-made transition metal catalysts characterized by much lower molecular weights and higher activities. As an example, directed evolution based mutants of a biotinylated Rh-diphosphine complex anchored to streptavidin led to enhanced enantioselectivity at each evolutionary stage, but activity was not improved (Scheme 7.8) [21]. Thus, in the area of artificial metalloenzymes the primary challenge revolves around the following basic question, which calls for intensified future research:

- How can directed evolution of the protein environment around a synthetic transition metal center induce notable enhancement of intrinsic catalytic activity?

References

- (a) O'Brien, P.J. and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.*, **6** (4), R91–R105; (b) Copley, S.D. (2003) Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Biol.*, **7** (2), 265–272; (c) Khersonsky, O., Roodveldt, C., and Tawfik, D.S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.*, **10** (5), 498–508.

2. Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.*, **30** (1), 409–425.
3. Amar, D., Berger, I., Amara, N., Tafa, G., Meijler, M.M., and Aharoni, A. (2012) The transition of human estrogen sulfo-transferase from generalist to specialist using directed enzyme evolution. *J. Mol. Biol.*, **416** (1), 21–32.
4. Benner, S.A. (2002) The past as the key to the present: resurrection of ancient proteins from eosinophils. *Proc. Natl. Acad. Sci. U.S.A.*, **99** (8), 4760–4761.
5. Bar-Rogovsky, H., Stern, A., Penn, O., Kobl, I., Pupko, T., and Tawfik, D.S. (2015) Assessing the prediction fidelity of ancestral reconstruction by a library approach. *Protein Eng. Des. Sel.*, **28** (11), 507–518.
6. (a) Tawfik, D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.*, **79** (1), 471–505; (b) Nobeli, I., Favia, A.D., and Thornton, J.M. (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.*, **27** (2), 157–167; (c) Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Gaucher, E.A., and Sanchez-Ruiz, J.M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β -lactamases. *J. Am. Chem. Soc.*, **135** (8), 2899–2902; (d) Babbitt, A., Tokuriki, N., and Hollfelder, F. (2010) What makes an enzyme promiscuous? *Curr. Opin. Chem. Biol.*, **14** (2), 200–207; (e) Atkins, W.M. (2015) Biological messiness vs. biological genius: mechanistic aspects and roles of protein promiscuity. *J. Steroid Biochem. Mol. Biol.*, **151**, 3–11.
7. Ward, O.P. and Singh, A. (2000) Enzymatic asymmetric synthesis by decarboxylases. *Curr. Opin. Biotechnol.*, **11** (6), 520–526.
8. Jochens, H., Stiba, K., Savile, C., Fujii, R., Yu, J.-G., Gerassenkova, T., Kazlauskas, R.J., and Bornscheuer, U.T. (2009) Converting an esterase into an epoxide hydrolase. *Angew. Chem. Int. Ed.*, **48** (19), 3532–3535.
9. (a) Roodveldt, C. and Tawfik, D.S. (2005) Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *Protein Eng. Des. Sel.*, **18** (1), 51–58; (b) Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R., and Tawfik, D.S. (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.*, **3**, 1257.
10. (a) Pandya, C., Farelli, J.D., Dunaway-Mariano, D., and Allen, K.N. (2014) Enzyme promiscuity: engine of evolutionary innovation. *J. Biol. Chem.*, **289** (44), 30229–30236; *J. Mol. Biol.* 2014, **289**, 30229–30236; (b) Hiblot, J., Gotthardt, G., Elias, M., and Chabriere, E. (2013) Differential active site loop conformations mediate promiscuous activities in the lactonase SsoPox. *PLoS One*, **8** (9), e75272; (c) Brizendine, A.M., Odokonyero, D., McMillan, A.W., Zhu, M., Hult, K., Romo, D., and Glasner, M.E. (2014) Promiscuity of *Exiguobacterium* sp. At1b o-succinylbenzoate synthase illustrates evolutionary transitions in the osbs family. *Biochem. Biophys. Res. Commun.*, **450** (1), 679–684; (d) Noda-García, L., Juárez-Vázquez, A.L., Ávila-Arcos, M.C., Verduzco-Castro, E.A., Montero-Morán, G., Gaytán, P., Carrillo-Tripp, M., and Barona-Gómez, F. (2015) Insights into the evolution of enzyme substrate promiscuity after the discovery of (β)8 isomerase evolutionary intermediates from a diverse metagenome. *BMC Evol. Biol.*, **15** (1), 1–14.
11. (a) Bornscheuer, U.T. and Kazlauskas, R.J. (2004) Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angew. Chem. Int. Ed.*, **43** (45), 6032–6040; (b) Berglund, P. and Park, S. (2005) Strategies for altering enzyme reaction specificity for applied biocatalysis. *Curr. Org. Chem.*, **9**, 325–336; (c) Hult, K. and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.*, **25** (5), 231–238.
12. Toscano, M.D., Woycechowsky, K.J., and Hilvert, D. (2007) Minimalist active-site redesign: teaching old enzymes new tricks. *Angew. Chem. Int. Ed.*, **46** (18), 3212–3236.

13. Renata, H., Wang, Z.J., and Arnold, F.H. (2015) Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed.*, **54** (11), 3351–3367.
14. (a) Branneby, C., Carlqvist, P., Magnusson, A., Hult, K., Brinck, T., and Berglund, P. (2003) Carbon–carbon bonds by hydrolytic enzymes. *J. Am. Chem. Soc.*, **125** (4), 874–875; (b) Li, C., Feng, X.-W., Wang, N., Zhou, Y.-J., and Yu, X.-Q. (2008) Biocatalytic promiscuity: the first lipase-catalysed asymmetric aldol reaction. *Green Chem.*, **10** (6), 616–618; (c) Torre, O., Gotor-Fernández, V., Alfonso, I., García-Alles, L.F., and Gotor, V. (2005) Study of the chemoselectivity in the aminolysis reaction of methyl acrylate catalysed by Lipase B from *Candida Antarctica*. *Adv. Synth. Catal.*, **347** (7–8), 1007–1014; (d) Cai, Y., Wu, Q., Xiao, Y.-M., Lv, D.-S., and Lin, X.-F. (2006) Hydrolase-catalyzed Michael addition of imidazoles to acrylic monomers in organic medium. *J. Biotechnol.*, **121** (3), 330–337; (e) Svedendahl, M., Hult, K., and Berglund, P. (2005) Fast carbon–carbon bond formation by a promiscuous lipase. *J. Am. Chem. Soc.*, **127** (51), 17988–17989; (f) Qian, C., Xu, J.-M., Wu, Q., Lv, D.-S., and Lin, X.-F. (2007) Promiscuous acylase-catalyzed aza-Michael additions of aromatic N-heterocycles in organic solvent. *Tetrahedron Lett.*, **48** (35), 6100–6104; (g) Strohmeier, G.A., Sović, T., Steinkellner, G., Hartner, F.S., Andryushkova, A., Purkarthofer, T., Glieder, A., Gruber, K., and Griengl, H. (2009) Investigation of lipase-catalyzed Michael-type carbon–carbon bond formations. *Tetrahedron*, **65** (29–30), 5663–5668; (h) Jiang, L., Wang, B., Li, R.-R., Shen, S., Yu, H.-W., and Ye, L.-D. (2014) Catalytic promiscuity of *Escherichia coli* BioH esterase: application in the synthesis of 3,4-dihydropyran derivatives. *Process Biochem.*, **49** (7), 1135–1138.
15. Svedendahl, M., Carlqvist, P., Branneby, C., Allnér, O., Frise, A., Hult, K., Berglund, P., and Brinck, T. (2008) Direct epoxidation in *Candida antarctica* lipase B studied by experiment and theory. *ChemBioChem*, **9** (15), 2443–2451.
16. Li, K., He, T., Li, C., Feng, X.-W., Wang, N., and Yu, X.-Q. (2009) Lipase-catalysed direct Mannich reaction in water: utilization of biocatalytic promiscuity for C–C bond formation in a "one-pot" synthesis. *Green Chem.*, **11** (6), 777–779.
17. Lou, F.-W., Liu, B.-K., Wu, Q., Lv, D.-S., and Lin, X.-F. (2008) *Candida antarctica* lipase B (Cal-B)-catalyzed carbon–sulfur bond addition and controllable selectivity in organic media. *Adv. Synth. Catal.*, **350** (13), 1959–1962.
18. (a) Reetz, M.T., Mondière, R., and Carballeira, J.D. (2007) Enzyme promiscuity: first protein-catalyzed Morita–Baylis–Hillman reaction. *Tetrahedron Lett.*, **48** (10), 1679–1681; (b) Jiang, L. and Yu, H.-W. (2014) An example of enzymatic promiscuity: the Baylis–Hillman reaction catalyzed by a biotin esterase (BIOH) from *Escherichia coli*. *Biotechnol. Lett.*, **36** (1), 99–103.
19. Humble, M.S. and Berglund, P. (2011) Biocatalytic promiscuity. *Eur. J. Org. Chem.*, **2011** (19), 3391–3401.
20. Wilson, M.E. and Whitesides, G.M. (1978) Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphine-rhodium(I) moiety. *J. Am. Chem. Soc.*, **100** (1), 306–307.
21. (a) Reetz, M.T., Peyralans, J.J.P., Maichele, A., Fu, Y., and Maywald, M. (2006) Directed evolution of hybrid enzymes: evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chem. Commun.*, (41), 4318–4320; (b) Reetz, M.T. (2012) Artificial metalloenzymes as catalysts in stereoselective Diels–Alder reactions. *Chem. Record*, **12** (4), 391–406; (c) Reetz, M. (2009) in *Bio-inspired Catalysts*, vol. 25 (ed. T. Ward), Springer, Berlin and Heidelberg, pp. 63–92.
22. (a) Thomas, C.M. and Ward, T.R. (2005) Artificial metalloenzymes: proteins as hosts for enantioselective catalysis. *Chem. Soc. Rev.*, **34** (4), 337–346; (b) Dürrenberger, M. and Ward, T.R. (2014) Recent achievements in the design and engineering of artificial metalloenzymes. *Curr. Opin. Chem. Biol.*, **19**, 99–106;

- (c) Dundas, C., Demonte, D., and Park, S. (2013) Streptavidin–biotin technology: improvements and innovations in chemical and biological applications. *Appl. Microbiol. Biotechnol.*, **97** (21), 9343–9353.
23. (a) Reetz, M.T. (2001) Optimisation of synthetic catalysts by means of directed evolution. Patent WO 2002103009A2. (priority 9.6.2001) (b) Reetz, M.T., Rentzsch, M., Pletsch, A., and Maywald, M. (2002) Towards the directed evolution of hybrid catalysts. *Chimia*, **56** (12), 721–723; (c) Reetz, M.T. (2002) Directed evolution of selective enzymes and hybrid catalysts. *Tetrahedron*, **58** (32), 6595–6602.
 24. Early studies of artificial metalloenzymes and other hybrid catalysts: (a) Qi, D., Tann, C.-M., Haring, D., and Distefano, M.D. (2001) Generation of new enzymes via covalent modification of existing proteins. *Chem. Rev.*, **101** (10), 3081–3112; (b) Polgar, L. and Bender, M.L. (1966) A new enzyme containing a synthetically formed active site. Thiol-subtilisin. *J. Am. Chem. Soc.*, **88** (13), 3153–3154; (c) Schultz, P.G. (1988) The interplay between chemistry and biology in the design of enzymatic catalysts. *Science*, **240**, 426–433; (d) Khumtaveeporn, K., DeSantis, G., and Jones, J.B. (1999) Expanded structural and stereospecificity in peptide synthesis with chemically modified mutants of subtilisin. *Tetrahedron: Asymmetry*, **10** (13), 2563–2572; (e) Smith, H.B. and Hartman, F.C. (1988) Restoration of activity to catalytically deficient mutants of ribulosebiphosphate carboxylase/oxygenase by aminoethylation. *J. Biol. Chem.*, **263** (10), 4921–4925; (f) Nicholas, K.M., Wentworth, P., Harwig, C.W., Wentworth, A.D., Shafon, A., and Janda, K.D. (2002) A cofactor approach to copper-dependent catalytic antibodies. *Proc. Natl. Acad. Sci. U.S.A.*, **99** (5), 2648–2653; (g) Hamachi, I. and Shinkai, S. (1999) Chemical modification of the structures and functions of proteins by the cofactor reconstitution method. *Eur. J. Org. Chem.*, **1999** (3), 539–549; (h) Lu, Y. and Valentine, J.S. (1997) Engineering metal-binding sites in proteins. *Curr. Opin. Struct. Biol.*, **7** (4), 495–500; (i) Lu, Y., Berry, S.M., and Pfister, T.D. (2001) Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. *Chem. Rev.*, **101** (10), 3047–3080; (j) Kaiser, E.T. (1988) Catalytic activity of enzymes altered at their active sites. *Angew. Chem., Int. Ed. Engl.*, **27** (7), 913–922; (k) Choma, C.T., Lear, J.D., Nelson, M.J., Dutton, P.L., Robertson, D.E., and DeGrado, W.F. (1994) Design of a heme-binding four-helix bundle. *J. Am. Chem. Soc.*, **116** (3), 856–865.
 25. Reetz, M.T., Rentzsch, M., Pletsch, A., Taglieber, A., Hollmann, F., Mondière, R.J.G., Dickmann, N., Höcker, B., Cerrone, S., Haeger, M.C., and Sterner, R. (2008) A robust protein host for anchoring chelating ligands and organocatalysts. *ChemBioChem*, **9** (4), 552–564.
 26. Mahammed, A. and Gross, Z. (2005) Albumin-conjugated corrole metal complexes: extremely simple yet very efficient biomimetic oxidation systems. *J. Am. Chem. Soc.*, **127** (9), 2883–2887.
 27. Mahammed, A., Gray, H.B., Weaver, J.J., Sorasaene, K., and Gross, Z. (2004) Amphiphilic corroles bind tightly to human serum albumin. *Bioconjugate Chem.*, **15** (4), 738–746.
 28. Zunszain, P.A., Ghuman, J., Komatsu, T., Tsuchida, E., and Curry, S. (2003) Crystal structural analysis of human serum albumin complexed with heme and fatty acid. *BMC Struct. Biol.*, **3** (1), 1–9.
 29. Ohashi, M., Koshiyama, T., Ueno, T., Yanase, M., Fujii, H., and Watanabe, Y. (2003) Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chem. Int. Ed.*, **42** (9), 1005–1008.
 30. Reetz, M.T. and Jiao, N. (2006) Copper–phthalocyanine conjugates of serum albumins as enantioselective catalysts in Diels–Alder reactions. *Angew. Chem. Int. Ed.*, **45** (15), 2416–2419.
 31. (a) Carey, J.R., Ma, S.K., Pfister, T.D., Garner, D.K., Kim, H.K., Abramite, J.A., Wang, Z., Guo, Z., and Lu, Y. (2004) A site-selective dual anchoring strategy for artificial metalloprotein design. *J.*

- Am. Chem. Soc.*, **126** (35), 10812–10813;
- (b) Lu, Y., Yeung, N., Sieracki, N., and Marshall, N.M. (2009) Design of functional metalloproteins. *Nature*, **460** (7257), 855–862.
32. Cornils, B. and Herrmann, W.A. (2006) *Aqueous-Phase Organometallic Catalysis*, Wiley-VCH Verlag GmbH, Weinheim.
33. (a) Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M.C., Sterner, R., and Wilmanns, M. (2002) Structural evidence for ammonia tunneling across the (β α)(8) barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure*, **10** (2), 185–193;
- (b) Beismann-Driemeyer, S. and Sterner, R. (2001) Imidazole glycerol phosphate synthase from *thermotoga maritima*: quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. *J. Biol. Chem.*, **276** (23), 20387–20396.
34. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science*, **289** (5484), 1546–1550.
35. Podtetenieff, J., Taglieber, A., Bill, E., Reijerse, E.J., and Reetz, M.T. (2010) An artificial metalloenzyme: creation of a designed copper binding site in a thermostable protein. *Angew. Chem. Int. Ed.*, **49** (30), 5151–5155.
36. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., and Houk, K.N. (2013) Computational enzyme design. *Angew. Chem. Int. Ed.*, **52** (22), 5700–5725.
37. Bjelic, S., Nivón, L.G., Çelebi-Ölçüm, N., Kiss, G., Rosewall, C.F., Lovick, H.M., Ingalls, E.L., Gallaher, J.L., Seetharaman, J., Lew, S., Montelione, G.T., Hunt, J.F., Michael, F.E., Houk, K.N., and Baker, D. (2013) Computational design of enone-binding proteins with catalytic activity for the Morita–Baylis–Hillman reaction. *ACS Chem. Biol.*, **8** (4), 749–757.
38. Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St. Clair, J.L., Gallaher, J.L., Hilvert, D., Gelb, M.H., Stoddard, B.L., Houk, K.N., Michael, F.E., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science*, **329** (5989), 309–313.
39. Colin, P.-Y., Kintses, B., Gielen, F., Miton, C.M., Fischer, G., Mohamed, M.F., Hyvonen, M., Morgavi, D.P., Janssen, D.B., and Hollfelder, F. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.*, **6**, 10008.
40. 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.
41. Letondor, C., Pordea, A., Humbert, N., Ivanova, A., Mazurek, S., Novic, M., and Ward, T.R. (2006) Artificial transfer hydrogenases based on the biotin – (strept)avidin technology: fine tuning the selectivity by saturation mutagenesis of the host protein. *J. Am. Chem. Soc.*, **128** (25), 8320–8328.
42. Creus, M., Pordea, A., Rossel, T., Sardo, A., Letondor, C., Ivanova, A., LeTrong, I., Stenkamp, R.E., and Ward, T.R. (2008) X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angew. Chem. Int. Ed.*, **47** (8), 1400–1404.
43. Hyster, T.K., Knörr, L., Ward, T.R., and Rovis, T. (2012) Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C–H activation. *Science*, **338** (6106), 500–503.
44. Jensen, C.N., Ali, S.T., Allen, M.J., and Grogan, G. (2013) Mutations of an NADP(H)-dependent flavoprotein monooxygenase that influence cofactor promiscuity and enantioselectivity. *FEBS Open Biol.*, **3**, 473–478.
45. (a) Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453** (7192), 190–195; (b) Khersonsky, O., Röthlisberger, D., Dym, O., Albeck, S., Jackson, C.J., Baker, D., and Tawfik, D.S. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.*, **396** (4), 1025–1042;

- (c) Khersonsky, O., Röthlisberger, D., Wollacott, A.M., Murphy, P., Dym, O., Albeck, S., Kiss, G., Houk, K.N., Baker, D., and Tawfik, D.S. (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.*, **407** (3), 391–412.
46. Casey, M.L., Kemp, D.S., Paul, K.G., and Cox, D.D. (1973) Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.*, **38** (13), 2294–2301.
47. Korendovych, I.V. and DeGrado, W.F. (2014) Catalytic efficiency of designed catalytic proteins. *Curr. Opin. Struct. Biol.*, **27**, 113–121.
48. Jencks, W.P. (1975) Binding energy, specificity, and enzyme catalysis: the Circe effect. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **43**, 219–410.
49. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.*, **273** (42), 27035–27038.
50. (a) Korendovych, I.V., Kulp, D.W., Wu, Y., Cheng, H., Roder, H., and DeGrado, W.F. (2011) Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U.S.A.*, **108** (17), 6823–6827; (b) Moroz, O.V., Moroz, Y.S., Wu, Y., Olsen, A.B., Cheng, H., Mack, K.L., McLaughlin, J.M., Raymond, E.A., Zhezherya, K., Roder, H., and Korendovych, I.V. (2013) A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew. Chem. Int. Ed.*, **52** (24), 6246–6249.
51. Blomberg, R., Kries, H., Pinkas, D.M., Mittl, P.R.E., Grütter, M.G., Privett, H.K., Majo, S.L., and Hilvert, D. (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*, **503**, 418–421.
52. (a) Levin, M., Amar, D., and Aharoni, A. (2013) Employing directed evolution for the functional analysis of multi-specific proteins. *Bioorg. Med. Chem.*, **21** (12), 3511–3516; (b) Moretti, R., Chang, A., Peltier-Pain, P., Bingman, C.A., Phillips, G.N., and Thorson, J.S. (2011) Expanding the nucleotide and sugar 1-phosphate promiscuity of nucleotidyltransferase RmLa via directed evolution. *J. Biol. Chem.*, **286** (15), 13235–13243.
53. Svastits, E.W., Dawson, J.H., Breslow, R., and Gellman, S.H. (1985) Functionalized nitrogen atom transfer catalyzed by cytochrome P-450. *J. Am. Chem. Soc.*, **107** (22), 6427–6428.
54. Arnold, F.H. (2015) The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.*, **48**, 404–410.
55. Wolf, J.R., Hamaker, C.G., Djukic, J.-P., Kodadek, T., and Woo, L.K. (1995) Shape and stereoselective cyclopropanation of alkenes catalyzed by iron porphyrins. *J. Am. Chem. Soc.*, **117** (36), 9194–9199.
56. Coelho, P.S., Brustad, E.M., Kannan, A., and Arnold, F.H. (2013) Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science*, **339** (6117), 307–310.
57. (a) Philpot, R.M. and Hodgson, E. (1972) The effect of piperonyl butoxide concentration on the formation of cytochrome P-450 difference spectra in hepatic microsomes from mice. *Mol. Pharmacol.*, **8** (2), 204–214; (b) Mansuy, D. (1980) New iron-porphyrin complexes with metal-carbon bond – biological implications. *Pure Appl. Chem.*, **52**, 681–690.
58. Roiban, G.-D. and Reetz, M.T. (2013) Enzyme promiscuity: using a P450 enzyme as a carbene transfer catalyst. *Angew. Chem. Int. Ed.*, **52** (21), 5439–5440.
59. Coelho, P.S., Wang, Z.J., Ener, M.E., Baril, S.A., Kannan, A., Arnold, F.H., and Brustad, E.M. (2013) A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.*, **9** (8), 485–487.
60. Wang, Z.J., Renata, H., Peck, N.E., Farwell, C.C., Coelho, P.S., and Arnold, F.H. (2014) Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables enantioselective formal synthesis of levomilnacipran. *Angew. Chem., Int. Ed. Engl.*, **53** (26), 6810–6813.
61. McIntosh, J.A., Coelho, P.S., Farwell, C.C., Wang, Z.J., Lewis, J.C., Brown,

- T.R., and Arnold, F.H. (2013) Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vivo and in vitro. *Angew. Chem. Int. Ed.*, **52** (35), 9309–9312.
62. Singh, R., Bordeaux, M., and Fasan, R. (2014) P450-catalyzed intramolecular sp³ C-H amination with arylsulfonyl azide substrates. *ACS Catal.*, **4** (2), 546–552.
63. Hyster, T.K., Farwell, C.C., Buller, A.R., McIntosh, J.A., and Arnold, F.H. (2014) Enzyme-controlled nitrogen-atom transfer enables regiodivergent C–H amination. *J. Am. Chem. Soc.*, **136** (44), 15505–15508.
64. (a) Sauer, D.F., Himiyama, T., Tachikawa, K., Fukumoto, K., Onoda, A., Mizohata, E., Inoue, T., Bocola, M., Schwaneberg, U., Hayashi, T., and Okuda, J. (2015) A highly active biohybrid catalyst for olefin metathesis in water: impact of a hydrophobic cavity in a β -barrel protein. *ACS Catal.*, **5**, 7519–7522; (b) Al Toma, R.S., Kuthning, A., Exner, M.P., Denisiuk, A., Ziegler, J., Budisa, N., and Süssmuth, R.D. (2015) Site-directed and global incorporation of orthogonal and isostructural noncanonical amino acids into the ribosomal lasso peptide capistruin. *ChemBioChem*, **16**, 503–509.

8

Learning from Directed Evolution

8.1

Background Information

Different types of lessons can be learned upon applying the technique of directed evolution, depending upon the specific goal of a research project. In analogy to the traditional use of rational site-specific mutagenesis in the quest to clarify the basic mechanism of an enzyme, randomization techniques also generate insights, provided the respective mutations are analyzed [1]. In these kinds of studies, the introduction of structural diversity ensures an expanded set of data not readily accessible by rational design. The experimental results of altered catalytic profiles allow for more detailed mechanistic interpretations and conclusions. When the sole goal is the clarification of an enzyme's mechanism, efficiency in terms of reducing the screening effort by generating focused smart libraries plays no significant role. A different situation arises when applying directed evolution in order to improve or alter the catalytic profile of an enzyme for practical purposes, for example, upon expanding or shifting substrate scope or manipulating stereo- and/or regioselectivity (see Chapters 4 and 5). In such cases, it is rewarding to invest further research efforts in order to unveil the origin of mutational effects on the evolved catalytic profile at a molecular level. This approach deepens our knowledge of the details of catalytic enzyme machineries.

A different kind of lesson can be learned from extensive deconvolution experiments of variants characterized by several point mutations because unique and valuable data emerges useful for the following:

- Studying the interaction of two or more point mutations with regard to additive or non-additive effects.
- Constructing fitness pathway landscapes that reveal the existence or absence of local minima.
- Drawing sound mechanistic conclusions on a molecular level when flanked by molecular dynamics (MD)/docking computations and X-ray data.

Hybrid quantum mechanics/molecular mechanics (QM/MM) methods continue to be refined with the aim of deepening our understanding of enzymes [2]. It is therefore logical to apply these computational methods to variants generated by directed evolution. Since this requires a great deal of computational effort, researchers of directed evolution usually analyze their results on the basis of MD/docking computations without resorting to QM. On the other extreme, in some publications only the crude manual placement of the substrate in the enzyme's binding pocket is reported using the wild-type (WT) crystal structure, with the respective point mutations simply being "added." This is of limited value, since conformational changes are not considered, and the chosen docking pose may have little to do with the real structure of the variant. Therefore, MD/docking experiments are indispensable, especially in the absence of X-ray structures of evolved variants. QM calculations needed in estimating $\Delta\Delta G^\ddagger$ of reaction in the (*R*)- and (*S*)-enantiomeric regimes, respectively, would be of even greater value, but such studies are rare due to significant time investment [3]. In view of the inherent uncertainty in estimating free energy values, even high-level QM/MM studies offer "only" models for understanding the catalytic profile of a mutant.

Stereoselectivity constitutes a sensitive probe when aiming to define an enzyme's mechanism in maximal detail. If the WT of an enzyme is highly enantioselective in a model reaction involving the natural or unnatural substrate, and its crystal structure has been solved, then details of the mechanism become accessible. Mechanistic and structural intricacies are also uncovered if a notable increase or reversal of stereo- and/or regioselectivity of an unselective enzyme has been achieved by directed evolution, provided a sound analysis by MD/docking computations follows, possibly flanked by X-ray structure determination(s), kinetic characterization, and inhibition experiments. Ideally, QM/MM should be applied to both evolved mutants and WT so that comparisons can be made. Most of the current and past directed evolution studies aimed at manipulating stereo- and/or regioselectivity, activity, or substrate scope do not include crystal structures of the evolved variants or QM calculations. The models that are proposed, if correct, deepen our understanding of (mutant) enzymes, but X-ray structures and QM studies would put them at a higher level of credibility.

For illustrative purposes, selected studies with and without crystal structures of evolved variants are introduced and analyzed in Section 8.2. In all cases, the focus is on stereo- and/or regioselectivity. In the section that follows thereafter (Section 8.3), studies are featured which describe deconvolution experiments and the use of such data in making conclusions concerning additive versus non-additive mutational effects. Both types of lessons are of fundamental importance in understanding and improving the efficacy of laboratory evolution.

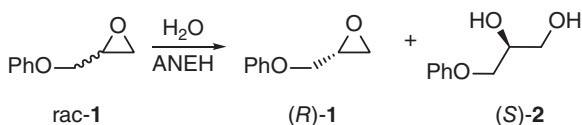
8.2

Case Studies Featuring Mechanistic, Structural, and/or Computational Analyses of the Source of Evolved Stereo- and/or Regioselectivity

8.2.1

Epoxide Hydrolase

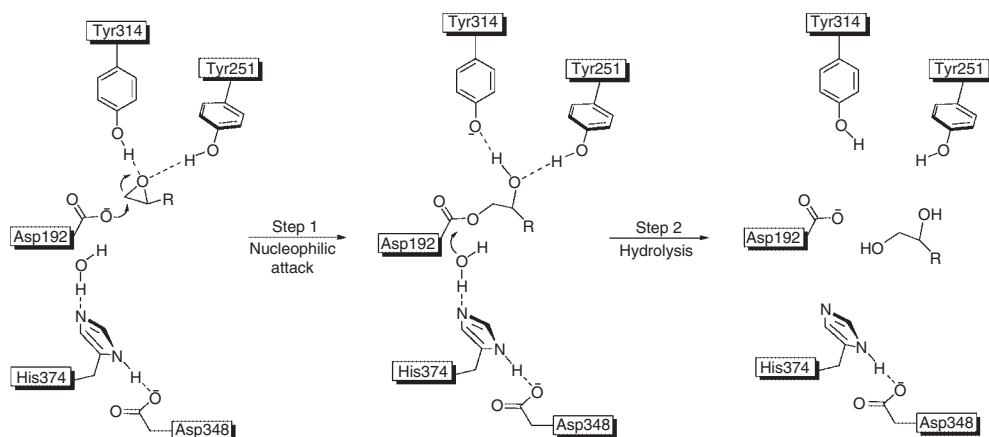
An informative mechanistic and structural study that includes the crystal structure of a stereoselective variant produced earlier by directed evolution concerns the epoxide hydrolase from *Aspergillus niger* (ANEH) as the catalyst in the hydrolytic kinetic resolution of *rac*-**1** with preferential formation of (*S*)-**2** (Scheme 8.1) [4]. WT ANEH is characterized by poor (*S*)-selectivity ($E = 4.6$). In the original study, six CAST sites were chosen for iterative saturation mutagenesis (ISM), A (comprising amino acid positions 193/195/196), B (215/217/219), C (329/330), D (349/350), E (317/318), and F (244/245/249). An arbitrarily chosen pathway $B \rightarrow C \rightarrow D \rightarrow F \rightarrow E$ provided the best variant LW202 showing a selectivity factor of $E = 115$ in favor of (*S*)-**1** [5]. Due to the very high degree of enantioselectivity, the upward climb was terminated without visiting site A (see also Chapter 3) [5]. This variant has nine point mutations L215F/A217N/R219S/L249Y/T317W/T318V/M329P/L330Y/C350V, which accumulated along the ISM pathway $WT \rightarrow B(\text{variant LW081}) \rightarrow C(\text{LW086}) \rightarrow D(\text{LW123}) \rightarrow F(\text{LW44}) \rightarrow E(\text{LW202})$.



Scheme 8.1 Hydrolytic kinetic resolution of *rac*-**1** catalyzed by ANEH [4, 5].

Prior to these studies, the crystal structure of WT ANEH had been solved [6] and the basic mechanism elucidated [7]. It involves binding of the substrate and activation by H-bonds to the epoxide O-atom originating from Tyr251 and Tyr314, catalytically active Asp192 then inducing an S_N2 reaction in the rate determining initial step followed by fast hydrolysis of the short-lived acyl-enzyme intermediate (Scheme 8.2).

As a first step in unraveling the source of enhanced stereoselectivity of the best variant LW202, kinetic analyses based on the Michaelis–Menten equation were carried out using in separate experiments enantiomerically pure (*R*)- and (*S*)-**1**, respectively. The results demonstrate a nearly ideal behavior of a kinetic resolution in that the reaction of the disfavored (*R*)-enantiomer is essentially shut down (Figure 8.1). The data also allows for a more exact determination of



Scheme 8.2 Mechanism of ANEH [4, 6, 7].

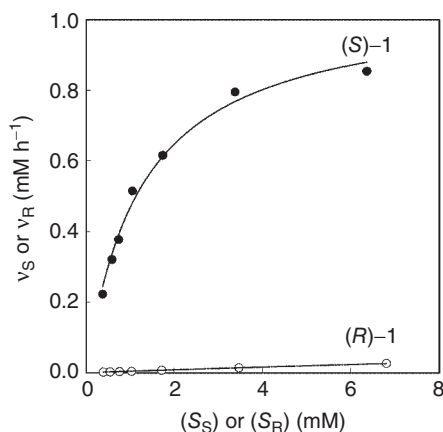


Figure 8.1 Kinetic analysis of variant LW202 as catalyst in separate reactions of (*R*)- and (*S*)-1, where v_R and v_S are the initial rates of hydrolysis of (*R*)- and (*S*)-1 at different

substrate concentrations $[S_R]$ or $[S_S]$ [4]. (Reetz *et al.* [4]. Reproduced with permission of American Chemical Society.)

the selectivity factor, which is even higher ($E = 195$) than the original estimation based on the standard Sih-equation [4]. Moreover, the relative values of k_{cat}/K_m for the two enantiomers also reflect pronounced (*S*)-selectivity. The K_m -value of the reaction of (*R*)-1 is considerably higher than that of (*S*)-1, while the situation reverses when comparing the k_{cat} -values (factor of 4.7 in favor of (*S*)-1) [4, 5].

In order to identify the factors that lead to enhanced (*S*)-enantioselectivity at every stage of the five-step evolutionary process, WT $\rightarrow$ LW081 $\rightarrow$ LW086 $\rightarrow$ LW123 $\rightarrow$ LW44 $\rightarrow$ LW202, extensive MD simulations were performed using (*R*)- and (*S*)-1 as substrates separately [4]. The distance, d , between the attacking O-atom of Asp192 and the epoxide C-atom undergoing S_N2 reaction was defined

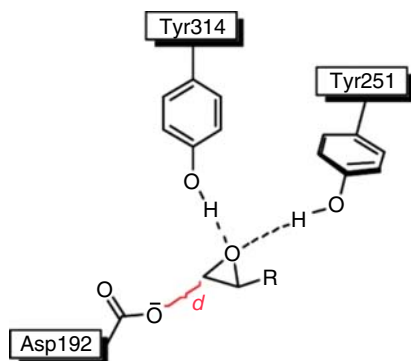


Figure 8.2 Definition of the distance d in the rate-determining step of the ANEH-catalyzed reaction of *rac*-1 [4]. (Reetz *et al.* [4]. Reproduced with permission of American Chemical Society.)

Table 8.1 Results of MD calculations [4].

Mutant	d_R	d_S	Δd_{R-S}	E (exp)
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

as the crucial parameter (Figure 8.2) [4]. It was assumed that a sufficiently small d -value would correspond fairly well to a near-attack pose, a concept proposed for many enzyme-catalyzed reactions [8]. Postulated more generally, productive binding can be expected if d is relatively short in the range of ≈ 3.5 Å. Large values were expected in the reaction of the disfavored enantiomer (*R*)-1. Indeed, a striking correlation ($R^2 = 0.86$) was observed between the experimental E -values and the differences in the computed distance, Δd_{R-S} , for the two enantiomeric substrates (Table 8.1) [4]. It can be seen that this difference increases as the evolutionary process proceeds. In the final variant LW202, d_R amounts to 5.4 Å, a clear indication that (*R*)-1 is disfavored. This model is somewhat crude, but in full agreement with the kinetics (Figure 8.2). It means that LW202 binds (*R*)-1 in an unproductive mode, thereby essentially shutting down the reaction, quite different from complexed (*S*)-1. In contrast, in the binding pocket of WT ANEH, both enantiomers are bound in a productive pose. The reasons for the different binding modes in LW202 were elucidated by the MD and docking computations. Disfavored (*R*)-1 is bound in an unproductive pose in which the C-atom of the epoxide undergoing S_N2 reaction is pointing “away” from the nucleophilic Asp192. Moreover, the MD computations indicate differences in flexibility in side-chain conformers as factors that contribute to differences in binding modes [4].

A significant asset of this study was the determination of two crystal structures, that of WT ANEH harboring the inhibitor valpromide (2-propyl-pentanoic acid amide) and that of apo (unbound) variant LW202 [4]. A comparison with apo

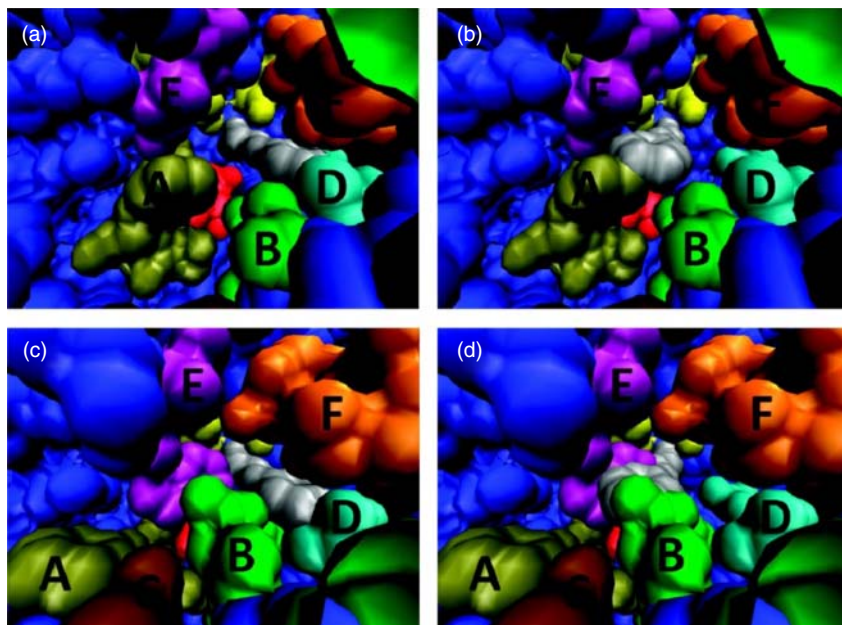


Figure 8.3 Interpretation of crystal structures of WT ANEH and evolved variants by manually docking (*R*)- and (*S*)-1 into binding pockets, A, B, C, D, E, and F representing the originally designed randomization sites in the ISM process. (a) Favored (*S*)-1 in WT ANEH binding pocket; (b) disfavored (*R*)-1 in binding pocket of WT ANEH; (c) favored (*S*)-1 in variant LW202; and (d) disfavored (*R*)-1 in variant LW201 [4]. (Reetz *et al.* [4]. Reproduced with permission of American Chemical Society.)

WT ANEH [6] was also made. While the gross features of all structures are almost identical (essentially same fold) and the positions of the amino acids participating in the catalytic machinery have not been perturbed, clear differences in the shape of the binding pocket of LW202 relative to apo or bound WT became visible. The structures were used for manually docking the favored (*S*)-1 and disfavored (*R*)-1 in the respective binding pockets in a manner that ensures smooth attack by nucleophilic Asp192 (Figure 8.3). The preferred (*S*)-enantiomer fits well into the WT (Figure 8.3a) and into LW202 (Figure 8.3c) without any steric clashes while maintaining activation by Tyr251/Tyr314 as well as optimal positioning for nucleophilic attack by Asp192. In sharp contrast, the disfavored (*R*)-enantiomer fits well into the WT binding pocket (Figure 8.3b), but not into LW202 because in this “forced” pose severe steric clashes occur between the substrate and the sidechains of mutated residues, especially at sites B and E (Figure 8.3d). Thus, productive binding is strongly prevented. This interpretation of the crystal structures is in line with the original MD computations performed in the absence of the LW202 X-ray structure and with the results of the kinetic study. Inhibition experiments proved to be in accord with the proposed model, which also explained the substrate profile of LW202 as catalyst in the stereoselective hydrolytic kinetic resolution of seven other substrates [4].

In conclusion, important mechanistic lessons can be learned from directed evolution, provided sufficient efforts are invested in interpreting altered catalytic profiles. This study uncovered the source of enhanced enantioselectivity on a molecular level. It also contributed to a deeper understanding of the mechanistic intricacies of this enzyme, and probably of other structurally and mechanistically related epoxide hydrolases [4]. If one views the docked substrates in the binding pockets (Figure 8.3) in greater detail, it seems that the angle of attack is not likely to be 180° as in traditional trajectories $\text{Nu}-\text{C}-\text{X}$ (Nu = nucleophile; X = leaving group) [9]; rather it should be smaller. Indeed, it has been reported that in (nonenzymatic) $\text{S}_{\text{N}}2$ reactions of epoxides the situation is different from reactions such as methyl iodide undergoing nucleophilic substitution [10]. QM calculations for certain epoxides and nucleophiles predict trajectories of $105-114^\circ$ [11]. This is in accord with a QM/MM study of limonene epoxide hydrolase (LEH) in which activated water functions as the nucleophile [12], and with MD computations of an evolved LEH mutant [13].

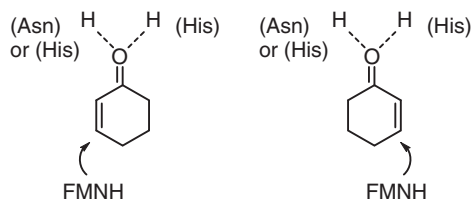
In a directed evolution study of LEH as catalyst in the desymmetrization of cyclohexene oxide, saturation mutagenesis at a 10-residue CAST site using a single amino acid as building block provided (*R,R*)- and (*S,S*)-selective variants. X-ray structures of the respective apo and product bound forms flanked by MD/docking computations provided sound models for explaining the origin of enhanced and inverted enantioselectivity [13]. The observation that a highly stereoselective LEH mutant, evolved for the desymmetrization of cyclohexene oxide, fails to accept the homologous substrate cyclopentene oxide, was puzzling. The crystals of this variant were soaked with the cyclopentene oxide, hoping to obtain crystals that show where and how this inert substrate is bound in the binding pocket of the LEH mutant. This experiment proved to be successful, the bound non-reacted substrate and the activated nucleophilic water occupying poses that are not conducive to smooth $\text{S}_{\text{N}}2$ reaction [13].

8.2.2

Ene-Reductase of the Old Yellow Enzyme (OYE)

Ene-reductases are excellent biocatalysts for the enantioselective reduction of a wide range of olefins bearing electron-withdrawing groups, for example, α,β -unsaturated ketones, esters, nitriles, and so on [14]. They generally belong to the class of Old Yellow Enzymes (OYEs). As in the case of other enzyme types, many substrates of interest to the synthetic organic chemist show poor or the wrong enantioselectivity, in which case directed evolution has been applied successfully [15] (see also Table 5.1).

The mechanism of OYEs has been shown to involve the traditional binding mode in which an asparagine and a histidine (or two histidines) form H-bonds to the carbonyl moiety of the substrate, the reduced flavin FMNH then delivering a hydride to one π -face of the activated β -position, and tyrosine spending a proton on the opposite π -face in an overall trans-specific manner (Scheme 8.3a). In the case of a prochiral substrate such as 3-substituted cyclohexenes, the addition



Scheme 8.3 Binding modes in the active site of ene-reductases. (left) Traditional (normal) binding mode and (right) flipped binding mode.

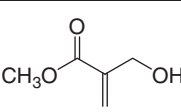
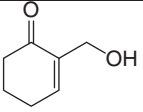
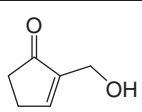
of two hydrogens leads to an enantiomer of defined absolute configuration. In the so-called flipped binding mode (Scheme 8.3b), the enantiomer of opposite absolute configuration results, which has been achieved by directed evolution [15a,c, 16]. When performing OYE-catalyzed reductions, it is necessary to employ an NADPH regeneration system such as glucose dehydrogenase/glucose.

In a series of illuminating protein engineering studies of the prototypical ene-reductase OYE1 from *Saccharomyces pastorianus* as a catalyst in the reduction of structurally different prochiral substrates, crystal structures of the evolved variants coupled with computational analyses uncovered the source of altered stereoselectivity [16]. In a preliminary investigation, it was postulated on the basis of the X-ray structure of WT OYE1 [17] that substitutions at position Trp116 could strongly influence the degree and direction of enantioselectivity. Indeed, upon screening an NNK saturation mutagenesis library at this position, improved variants were found for the reduction of such substrates as 2-methylcyclohexenone and (*R*)- and (*S*)-carvone [18]. Later, this library was screened for a wide variety of other substrates, the mutants being characterized by crystal structures [16].

In the initial study, the OYE1 catalyzed reduction of Baylis–Hillman adducts was investigated; it was shown, inter alia, that in the reaction of 2-hydroxymethylcyclopentenone variant Trp116Ile causes reversal of enantioselectivity (91% ee (*S*) versus 60% ee (*R*) of WT) (Table 8.2) [16a].

Variant Trp116Ile was soaked with 2-hydroxymethylcyclopentenone, leading to crystals harboring the substrate, which diffracted at 1.7–1.4 Å. Surprisingly, two different binding poses were identified in the crystal, mode 1 with the re-face of the π -system facing the FMNH₂ hydride source and leading to the minor (*R*)-product, and mode 2 in the flipped si-orientation leading to the observed (*S*)-product (Figure 8.4). Analysis of the two structures revealed different H-bond interactions of the hydroxymethyl group of the substrate with the respective environment in the two modes. Moreover, mode 1 was identified as an inferior Michaelis complex for catalysis, because the angle formed by FMN N₁₀–N₅–substrate β -carbon is 78°, outside of the usual 96–117° range observed previously in smooth hydride transfer [19]. This led to a model in which mode 2 is the productive pose. The results teach us that crystal structures alone are insufficient for understanding the source of enantioselectivity, and that careful interpretations are necessary. MD/docking computations were not reported in this study.

Table 8.2 Stereoselective reduction of Baylis–Hillman adducts catalyzed by variants of OYE1 produced by saturation mutagenesis at position 116 [16a].

Protein	Substrate					
						
	% conv	% ee	% conv	% ee	% conv	% ee
<i>Saccharomyces pastorianus</i>						
OYE1 mutants (116 residue)						
Trp (wt)	19	>98 (<i>R</i>)	≤5	N.D.	51	60 (<i>R</i>)
Ala	9	90 (<i>R</i>)	84	>98 (<i>S</i>)	>98	72 (<i>S</i>)
Val	52	86 (<i>R</i>)	84	>98 (<i>S</i>)	97	92 (<i>S</i>)
Tyr	68	76 (<i>R</i>)	>98	>98 (<i>S</i>)	>98	87 (<i>S</i>)
Phe	37	70 (<i>R</i>)	>98	>98 (<i>S</i>)	98	>98 (<i>S</i>)
Ser	13	46 (<i>R</i>)	84	>98 (<i>S</i>)	87	>98 (<i>S</i>)
Ile	50	9 (<i>R</i>)	>98	>98 (<i>S</i>)	>98	91 (<i>S</i>)
Arg	≤5	N.D.	≤5	N.D.	≤5	N.D.
Pro	≤5	N.D.	14	>98 (<i>S</i>)	16	77 (<i>S</i>)
Thr	≤5	N.D.	28	>98 (<i>S</i>)	44	>98 (<i>S</i>)
Cys	≤5	N.D.	31	>98 (<i>S</i>)	47	77 (<i>S</i>)
Lys	≤5	N.D.	60	>98 (<i>S</i>)	75	76 (<i>S</i>)
Glu	≤5	N.D.	93	90 (<i>S</i>)	96	88 (<i>S</i>)
Asp	≤5	N.D.	>98	91 (<i>S</i>)	95	77 (<i>S</i>)
Gly	14	16 (<i>S</i>)	98	>98 (<i>S</i>)	>98	86 (<i>S</i>)
Leu	>98	20 (<i>S</i>)	>98	>98 (<i>S</i>)	>98	57 (<i>S</i>)
Asn	>98	41 (<i>S</i>)	>98	>98 (<i>S</i>)	>98	89 (<i>S</i>)
Met	15	64 (<i>S</i>)	>98	>98 (<i>S</i>)	>98	86 (<i>S</i>)
His	67	97 (<i>S</i>)	>98	>98 (<i>S</i>)	>98	77 (<i>S</i>)
Gln	78	>98 (<i>S</i>)	>98	>98 (<i>S</i>)	>98	89 (<i>S</i>)
<i>P. stipitis</i> OYE 2.6	>98	>98 (<i>S</i>)	>98	>99 (<i>S</i>)	>98	76 (<i>S</i>)

Source: Walton *et al.* [16a]. Reproduced with permission of American Chemical Society.

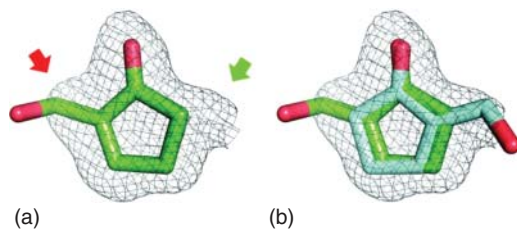


Figure 8.4 Location of substrate 2-hydroxymethyl cyclopentenone in OYE1 mutant Trp116Ile within the observed electron density (0.4σ contour level). (a) Attempted poor fit by a single substrate orientation. Red and green arrows indicate regions of negative and positive electron

density peaks, respectively, in the difference map (not shown). (b) Successful fit by two substrate populations. C-atoms in binding mode 1 are pictured in green, those in binding mode 2 are shown in light blue [16a]. (Walton *et al.* [16a]. Reproduced with permission of American Chemical Society.)

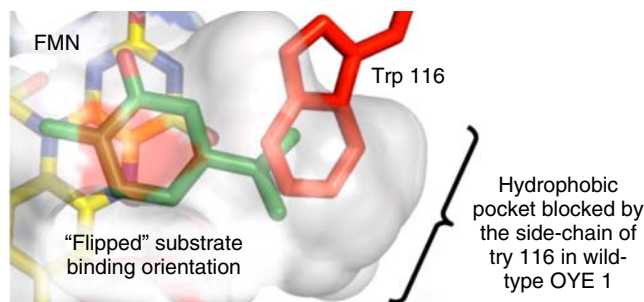


Figure 8.5 Schematic representation of the role of OYE1 variants characterized by mutations at position 116 [16b]. (Pompeu *et al.* [16b]. Reproduced with permission of American Chemical Society.)

In the second study, the source of enhanced and reversed diastereoselectivity of OYE1-catalyzed reduction of (*R*)- and (*S*)-carvone observed earlier [18] was elucidated [16b]. The X-ray crystal structures of several key Trp116 variants and of WT OYE1 harboring (*R*)- or (*S*)-carvone revealed subtle changes that appear to control the orientation of substrate binding and thus determine diastereoselectivity. In line with the hydrophobic character of carvone (and of most other compounds), substrate binding appears to be controlled largely by steric factors, and indeed the binding pocket of the ene-reductase is defined primarily by hydrophobic amino acids (Thr37, Met39, Phe74, Tyr82, Ala85, and Leu118), where substrates having large substituents can be accommodated, but access is blocked by tryptophan at position 116 (Figure 8.5). Substitutions at this position with incorporation of polar amino acids generally reduce activity sharply, with few exceptions. In the reaction of (*S*)-carvone, appropriate mutants allow the substrate to enter and also to determine either normal (e.g., Trp116Leu) or flipped binding modes (e.g., Trp116Ala). In the case of (*R*)-carvone, all variants lead to the normal binding mode, with the exception of Trp116Ala and vTrp116Val. This is surprising, since variants having leucine or isoleucine at position 116 bind (*R*)-carvone only in the normal orientation. Subtle effects play a crucial role, but these are not easy to identify. Nevertheless, reasonable models based on the stereochemical and X-ray structural data were proposed.

The third study focuses on the origin of enhanced and reversed enantioselectivity in the bioreduction of (*Z*)- β -aryl- β -cyanoacrylates catalyzed by OYE1 mutants [16c]. The original saturation mutagenesis library at residue Trp116 was used once more. Both (*R*)- and (*S*)-products were observed, depending upon the amino acid substitution at position 116 of the variant. The degree of enantioselectivity varied considerably, depending upon the substitution pattern of the aryl-substituent, and upon the particular mutation. Normal but also flipped binding modes were deduced. Soaking experiments with the aim of visualizing the binding mode of this class of substrates in the OYE1 variants were successful only in the case of the Trp116Ala variant. Two crystal structures were analyzed harboring (*Z*)- β -phenyl- β -cyano methyl acrylate and the *p*-fluoro-analog,

respectively. Both showed a non-productive pose that would be expected to provide the disfavored enantiomeric products. Even though such intermediates are not involved in the formation of the observed enantiomers, the structural data is nevertheless useful in interpreting the stereochemical results. It was concluded that small amino acids at position 116 induce the substrate to adopt a classical orientation with formation of (*S*)-configured products, while larger amino acids such as leucine result in a flipped pose leading to (*R*)-products [16c].

In conclusion, the combination of stereochemical results, X-ray structural data of OYE1 variants harboring prochiral substrates, and detailed analyses of the data have resulted in sound models that explain the observed activity and enantioselectivity. Pinpointing the factors that are necessary for high activity and enantioselectivity in the case of structurally different unnatural substrates also throws light on the fundamental efficacy of OYEs in the reduction of natural substrates. Further refinement of the models may be possible by applying QM/MM. Moreover, the X-ray data suggests that in addition to position 116, other CAST sites should also be considered for saturation mutagenesis.

In a different approach, laboratory evolution of the ene-reductase from *Candida macedoniensis* AKU4588 (CmOYE) was performed by introducing mutations in the substrate-recognition loop near the catalytically active site [20]. The results show that such a focus could also be successful in protein engineering of other OYEs. When comparing the crystal structure of WT CmOYE (apo) with the WT harboring the inhibitor *p*-hydroxybenzaldehyde (*p*-HBA), the high flexibility of loop 6 became apparent (Figure 8.6), which is crucial in determining substrate acceptance and stereoselectivity [20]. In the apo form, the loop acts as a lid that closes the active site

The WT enzyme was used in the reduction of ketoisophorone, which produced solely the desired (*R*)-product, but unfortunately the reaction proved to be too slow for industrial applications (Scheme 8.4) [20]. Structure-guided mutations

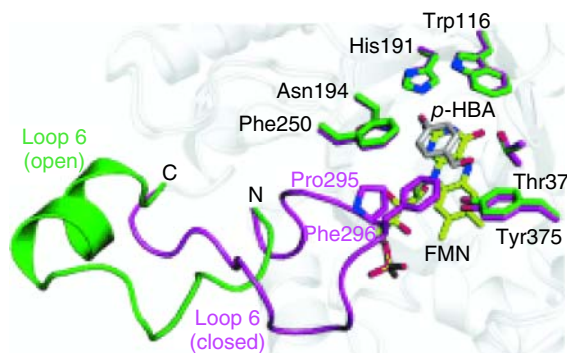
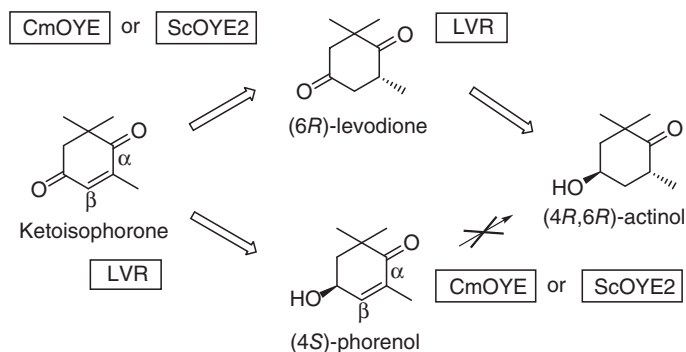


Figure 8.6 Superposition of CmOYE structures in the absence (green) and presence (magenta) of *p*-HBA in the catalytic pockets. The structures shown in green and magenta represent open and closed forms

of CmOYE (loop 6), respectively. Amino acid residues in the catalytic sites, FMN (yellow), and *p*-HBA (gray), are shown as stick models [20]. (Horita *et al.* [20]. Reproduced with permission of John Wiley & Sons.)



Scheme 8.4 Two-step biocatalytic conversion of ketoisophorone to (4R,6R)-actinol. Biocatalytic synthesis of (4R,6R)-actinol from ketoisophorone is performed by CmOYE (or ScOYE2) and LVR. CmOYE and ScOYE2 show

less catalytic activity in the reduction of (4S)-phorenol than in the other reactions [20]. (Horita *et al.* [20]. Reproduced with permission of John Wiley & Sons.)

in loop 6 led to a twofold increase in activity while maintaining complete enantioselectivity, single mutant Pro295Gly being a prominent example. This variant was also effective in the stereoselective reduction of (4S)-phorenol with formation of the industrially desired (4R,6R)-actinol (Scheme 8.4), a 12-fold increase in activity being observed. Although X-ray data of the variants were not reported, it was possible to propose a reasonable model for gaining insight into the origin of the altered catalytic profile. It was postulated that Pro295 collides sterically with the dimethyl group of ketoisophorone and (4S)-phorenol, which means that it acts as substrate gate keeper [20]. Mutation to Pro295Gly leads, *inter alia*, to greater flexibility and provides more space for substrate binding, an intriguing proposition that appears to be in line with the data. In order to corroborate this conclusion, MD/docking computations would need to be performed.

Recently, the first QM/MM study of an OYE (wildtype YqjM from *Bacillus subtilis*) appeared that extends the mechanistic picture of this class of enzymes [21]. This OYE has been used extensively in stereoselective reductions and is believed to follow a mechanism similar to other members of this class of enzymes except that two histidines (His164 and His167) bind and activate the substrate (instead of an Asn/His pair) [22]. Using 2-cyclohexenone as the substrate, two-dimensional QM reaction pathways at the B3LYP-D/OPLS2005 level suggest that the hydride adds in a distinct rate-determining step followed by protonation of the stabilized enolate at the C-atom (not O-atom) in the second step (Figure 8.7). Thus, hydride and proton motions are not concerted as sometimes proposed.

The computed activation barriers ($\Delta E^\ddagger$) and reaction energies (ΔE) for hydride transfer from N5 of the reduced flavin FMNH to the β -C-atom of 2-cyclohexenone are summarized in Table 8.3. The identification of weak substrate binding, polarization, and activation of the carbonyl O-atom as well as the structural information of the transition state provide detailed models of the catalytic machinery of OYEs in general. Details of the origin of the proton

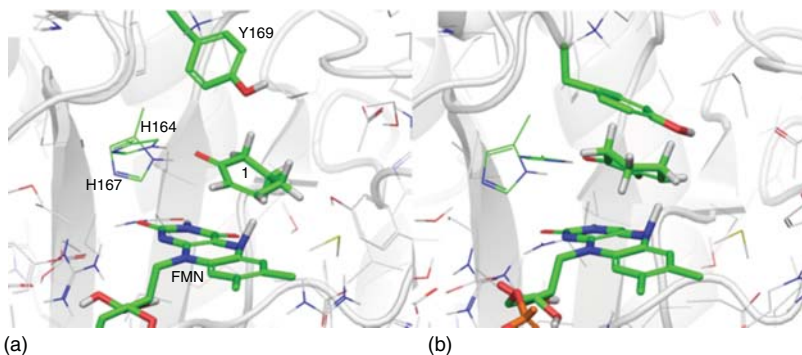


Figure 8.7 Computed transition state geometries of the lowest energy pathways for hydride transfer: (a) in the normal pose and (b) in the flipped pose [21]. (Lonsdale and Reetz [21]. Reproduced with permission of American Chemical Society.)

Table 8.3 Computed activation barriers and reaction energies for rate-determining hydride transfer from FMNH to 2-cyclohexenone in kcal mol⁻¹ [21].

Substrate orientation	Profile	$\Delta E^\ddagger$	ΔE
Normal	1	15.7	7.3
	2	16.9	7.0
	3	18.0	9.3
Flipped	1 ^a	32.5	22.6
	2 ^a	32.9	24.5
	3 ^a	33.2	19.1
	4 ^b	22.5	16.4
	5 ^b	25.5	20.2

a) Denotes starting structures that were obtained from the MD simulation prior to the observed conformation rearrangement;

b) Denotes starting structures that were obtained from the MD simulation following the observed conformation rearrangement.

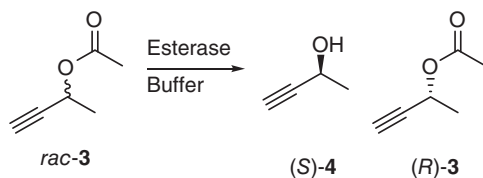
Source: Lonsdale and Reetz [21]. Reproduced with permission of American Chemical Society.

adding to the α -C-atom of the reacting cyclohexanone are also included; it does not appear to be Tyr169 as traditionally assumed. The insights generated by this theoretical study may prove to be useful in designing optimal mutagenesis experiments in future directed evolution studies.

8.2.3

Esterase

An unusual lesson was learned in a directed evolution project involving the esterase from *Pseudomonas fluorescens* (PFE) [23]. The enantioselectivity



Scheme 8.5 Model hydrolytic kinetic resolution of *rac*-3 catalyzed by PFE [23]. (Schmidt *et al.* [23]. Reproduced with permission of American Chemical Society.)

of the PFE-catalyzed hydrolytic kinetic resolution of the propargyl alcohol *rac*-3 favoring the formation of (*S*)-4 is very low ($E=3$) (Scheme 8.5). A 7000-membered error-prone polymerase chain reaction (epPCR) library was created at a mutation frequency of 1–2 mutations per gene, which was screened for stereoselective variants. Several hits were identified, the best one (V2A) showing a selectivity factor of $E=89$ at 54% conversion, although the reaction proved to be very slow.

Sequencing showed that V2A is a triple mutant Ile76Val/Gly98Ala/Val175Ala. Unfortunately, in *Escherichia coli* it was expressed as inclusion bodies with only a minor portion being soluble. In contrast, WT PFE posed no problems in this respect. It was initially assumed that the Gly98Ala mutation is spatially close to the catalytic triad (Ser94, His251, and Asp222), and that the remote mutations Ile76Val and Val175Ala are responsible for inclusion body formation [23]. In an attempt to test this hypothesis, several experiments and computations were performed. Using the known 3D structure of PFE [24], a homology model displaying the three point mutations was created (Figure 8.8). It can be seen that mutation Gly98Ala is indeed fairly close to catalytically active Ser94.

In order to develop some understanding of the structural reasons for the experimental results, all single and double mutants corresponding to variant V2A were generated by site-specific mutagenesis and tested as catalysts in the model reaction (Table 8.4). An additional mutation at position 99 (Asp99Glu) was included in variant VEA₂. Four variants that lacked the Gly98Ala mutations were expressed nicely without forming inclusion bodies and showed specific activities similar to WT PFE. It was, therefore, concluded that mutation Gly98Ala in the variants is responsible for inclusion body formation, a surprising result since the mutation site is near the catalytically active triad [23]. Indeed, single mutant Gly98Ala also forms IBs, and this effect seems to be enhanced by the other mutations in the respective double and triple mutants.

MD simulations using WT PFE and triple mutant V2A led to novel insights. The triple mutant V2A is characterized by a helix that is extended by one loop at position 98 (Figure 8.9). The helix breaker glycine in WT PFE was replaced in the mutant by helix supporter alanine at the end of the α -helix. This leads to a new conformation that destroys the original tertiary structure. It was postulated that this is probably the cause of IB-formation. The origin of enhanced enantioselectivity was not analyzed in this report, but the lesson learned is nevertheless



Figure 8.8 3D homology model of PFE. The catalytic triad is shown in gray (Ser94, His 251, Asp222), mutation sites are highlighted in black (Val76, Ala98, and Ala175). The model was created by using PyMOL

and amino-acid exchanges were introduced with the “Wizard/Mutagenesis” feature [23]. (Schmidt *et al.* [23]. Reproduced with permission of American Chemical Society.)

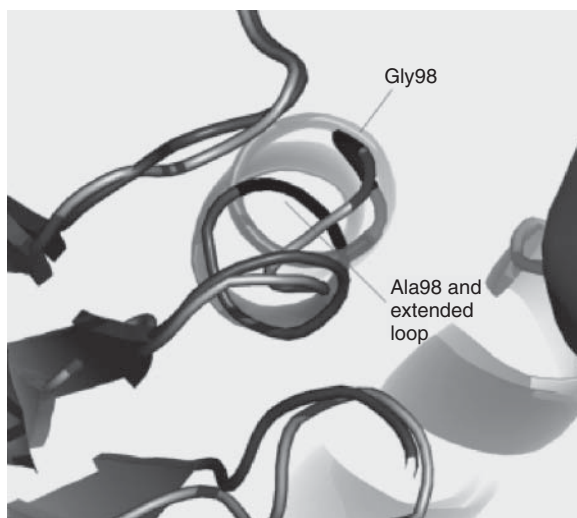


Figure 8.9 Alignment of WT PFE (light gray, Gly98 labeled) and variant V2A (dark gray, Ala98 labeled). The extended loop of the helix is highlighted [23]. (Schmidt *et al.* [23]. Reproduced with permission of American Chemical Society.)

Table 8.4 Catalytic profiles of WT PFE, variant V2A generated by epPCR and variants resulting from deconvolution of the latter.

Name	PFE variant	Activity (U per mg protein)		$E \sim 50$ (%) ^{a)}	t (min)	$E_{\max}$ ^{b)c)}	t (min)
		Lyophilisate	IB ^{c)}				
WT	Wild-type	77	–	63	5	3 (96 %)	1440
V2A	Ile76Val/Gly98Ala/Val175Ala	0.006	+	89	5700	89 (54 %)	5700
2A	Gly98Ala/Val175Ala	0.2	+	>100 ^{d)}	180	>100 (40 %)	180
VA ₁	Ile76Val/Gly98Ala	0.6	+	80 ^{e)}	10	80 (25 %)	10
A ₁	Gly98Ala	9	+	>100	5	90 (57 %)	1500
VEA ₂	Ile76Val/Asp99Glu/Val175Ala	37	–	92	420	92 (53 %)	420
V	Ile76Val	49	–	>100	1	16 (83 %)	1500
VA ₂	Ile76Val/Val175Ala	57	–	96	20	96 (53 %)	20
A ₂	Val175Ala	67	–	>100	1	26 (74 %)	1500

a) Calculated at 50% conversion.

b) Calculated at maximal conversion given in brackets (%).

c) IB, inclusion body.

d) Calculated at 40% conversion.

e) Calculated at 25% conversion.

Specific activities were measured by monitoring the hydrolysis of *p*-nitrophenyl acetate (*p*NPA). Enantioselectivity was ascertained by measuring the selectivity factor E in the hydrolytic kinetic resolution of *rac*-3 [23].

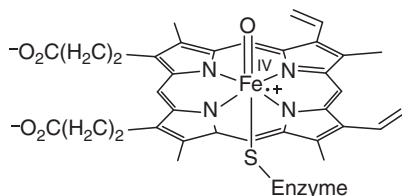
Source: Schmidt *et al.* [23]. Reproduced with permission of American Chemical Society.

intriguing, namely that inclusion bodies may well arise when mutations occur near the active site.

8.2.4

Cytochrome P450 Monooxygenase

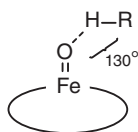
Cytochrome P450 (CYP) monooxygenases are heme-Fe dependent enzymes that catalyze the oxidative hydroxylation $R-H \rightarrow R-OH$ of a wide range of natural and unnatural organic compounds [25]. Olefin epoxidation has also been observed. The mechanism of oxidative hydroxylation involves the intermediacy of a catalytically active high-spin heme-Fe=O intermediate (so-called “Compound I”) as shown in Scheme 8.6, which induces in the rate-determining step the abstraction of an H-atom from the substrate $R-H$ with formation of the short-lived radical $\bullet R$ followed by rapid C–O bond formation. It has been estimated on the basis of theoretical computations that the ideal substrate pose for smooth C–H activation involves an O–H–C angle of about 130° defined by the oxygen atom of Fe=O and the C–H entity of the substrate [26]. This high-energy process has been exploited in industrially relevant regio- and stereoselective transformations for a long time [27]. During the past decades, several other highly selective transformations have been reported, but unfortunately the vast majority of tested substrates have failed to react with high regio- and stereoselectivity. Moreover, many alkanes are not even accepted for various reasons. Due to the very large binding pockets



Scheme 8.6 Compound I as the catalytically active high-spin intermediate in CYP-catalyzed oxidative hydroxylation.

of CYPs, particularly small substrates such as alkanes of the type propane, ethane, or methane are not oxidized because they appear in poses which, on average, are too far away from the catalytically active high-spin heme-Fe=O intermediate (Scheme 8.6) [28]. Many, but not all of these problems have been solved by directed evolution; for reviews of protein engineering of CYPs, see [29].

It is clear that CYPs are quite different from most other enzyme types, which means that different interpretations for explaining catalytic profiles are necessary. Emil Fischer's lock-and-key postulate emphasizing shape complementarity was extended by Linus Pauling's general enzyme theory based on the hypothesis that the transition states of enzyme-catalyzed transformations are stabilized by intimate interactions with the protein environment. CYP-catalyzed oxidative hydroxylations appear to be an exception to the Pauling hypothesis, because it is unlikely that the energy of a radical process can be reduced to a notable extent by such interactions, although some electrostatic stabilization may occur [25e]. Therefore, other explanations for substrate acceptance and regio- and stereoselectivity had to be developed, directed evolution playing an important role in this endeavor. As a result of many protein studies flanked by theoretical analyses, it has become clear that the protein environment in the binding pocket "holds" the substrate above the catalytically active high-spin Compound I (Scheme 8.7) in a pose in which at least one of the CH-entities points toward the O-atom of heme-Fe=O. In the case of enantio- or diastereoselectivity, the hydrogen of the substrate that spends the largest amount of time in a sufficiently close position reacts preferentially and therefore defines the stereochemical outcome. This may well mean that the weakest C-H bond in a molecule is not oxidized. In a series of studies, MD/docking calculations were performed, which provide some insight regarding the source of regio- and stereoselectivity [30].



Scheme 8.7 Ideal pose of a substrate for smooth oxidative hydroxylation initiated by H-atom abstraction and formation of an intermediate short-lived radical $\bullet R$ that undergoes rapid C-O bond formation. The ideal O-H-C angle has been computed to be about 130° [26]. (Lonsdale *et al.* [26]. Reproduced with permission of American Chemical Society.)

P450-BM3 is a self-sufficient CYP from *Bacillus megaterium* consisting of a heme-Fe domain and a diflavin reductase domain [31]. Fatty acids are the natural substrates. It has been characterized by X-ray crystallography [32] and used in many directed evolution studies [28–30]. In one report, it was noted that the C12 cycloalkane, cyclododecane, is not accepted by WT P450-BM3 [33]. A small library of 24 mutants was tested, generated by utilizing five hydrophobic amino acids (Ala, Val, Phe, Leu, and Ile) at positions 87 and 328 directly above the heme-group. The double mutant Phe87Ala/Ala328Val proved to be active. It had been known for some time that mutation Phe87Ala is instrumental in expanding the substrate scope of P450-BM3, because Phe87 appears to shield heme-Fe=O to some extent. In order to learn more about the effect of the double mutant as the catalyst in cyclododecane oxidation, MD/docking computations were performed (Figure 8.10) [33].

It was discovered that in the case of the double mutant Phe87Ala/Ala328Val, the pose of the substrate remains stable, the compound being in close proximity to the catalytically active heme-Fe=O (Compound I) with the C–H O–Fe distance of 2.95 Å, as indicated computationally. It can also be seen that the small sidechain of Ala in the double mutant creates a space, part of which is occupied by cyclododecane (Figure 8.10a). The situation in the case of inactive WT P450-BM3 is quite different. The bulky sidechain of Phe87 occupies this space, thereby blocking an optimal substrate pose (Figure 8.10b). However, this effect alone does not explain high activity toward the substrate, because single mutant Phe87Ala leads to a mere 4% conversion under the same experimental conditions. The other single mutant Ala328Val fails completely, which shows that the combined

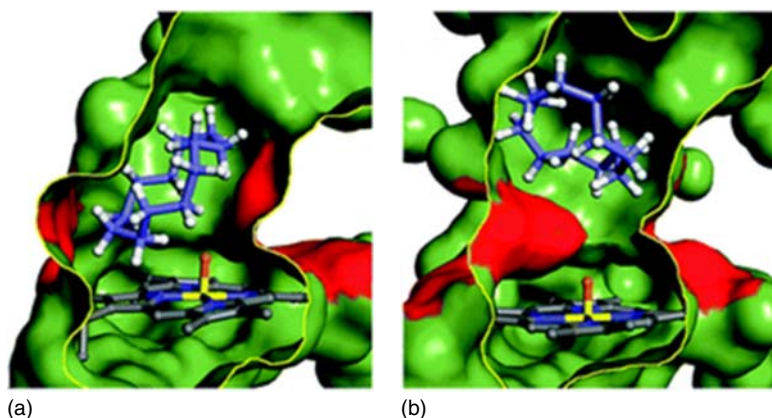


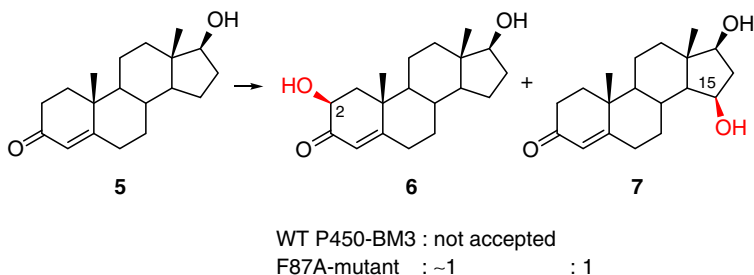
Figure 8.10 The substrate-binding cavity of P450 BM3 F87A/A328V (a) and P450 BM3 wild type (b) in complex with cyclododecane after 3 ns of unrestraint MD simulation. The mutated positions are depicted in red. Positions 87 (left) and 328 (right)

stabilize the substrate in the active site cavity. The activated oxygen of the heme is shown in orange [33]. (Weber *et al.* [33]. Reproduced with permission of Royal Society of Chemistry.)

action of the two mutations is necessary for smooth oxidation (Figure 8.10a) [33]. MD/docking data of the single mutant Ala328Val was not reported, but would provide additional insight into the details of substrate acceptance.

P450-BM3 has also been used in directed evolution for manipulating regio- and stereoselectivity in one and the same optimization process, as, for example, in the oxidative hydroxylation of steroids [30a], of 1-cyclohexene carboxylic acid ester with formation of the (*R*)- and (*S*)-enantiomers [30b], of ketones affording chiral acyloins [30c], and of 1-tetralones leading to reaction at the 4-position [30d], of appropriate substrates with the creation of two chirality centers [30e], and of caged substrates in bioorthogonal processes [30f]. In all cases, ISM at CAST sites lining the binding pocket was applied. MD/docking calculations showed that in the binding pockets of the P450-BM3 mutants the substrates occupy an energetically preferred pose in which one of the two stereotopic C–H entities of a methylene group points to the high-spin heme-Fe=O. In contrast, when performing such computations using WT P450-BM3 or the “standard” mutant Phe87Ala, poor regio- as well as diastereo- and enantioselectivity are indicated.

One of many examples of this kind of analysis was reported in the ISM-based study of P450-BM3 as a catalyst in the oxidative hydroxylation of testosterone (5) and other steroids (Scheme 8.8) [30a]. WT P450-BM3 fails to accept this substrate, while variant Phe87Ala is active but delivers a 50 : 50 mixture of regioisomers 6 and 7 in addition to small amounts of other alcohols. The evolved triple mutant Arg47Ile/Thr49Ile/Phe87Ala leads to 94% overall selectivity in favor of the 2 β -product 6, while variant Arg47Tyr/Thr49Phe/Val78Leu/Ala82Met/Phe87Ala is 96% selective for the 15 β -product 7.



Scheme 8.8 P450-BM3 catalyzed oxidative hydroxylation of testosterone [30a]. (Kille *et al.* [30a]. Reproduced with permission of Nature Publishing Group.)

Extensive MD/docking simulations using substrate 5 and mutant Phe87Ala point to two poses of essentially equal energy, one predicting 2 β -selectivity and the other 15 β -selectivity, in line with the experimental results [30a]. In sharp contrast, the analogous computations using the two ISM-evolved variants indicate only one reactive pose in each case. Figure 8.11 shows that in the 2 β -selective variant the pro- β H-atom of the methylene group at the 2C-position

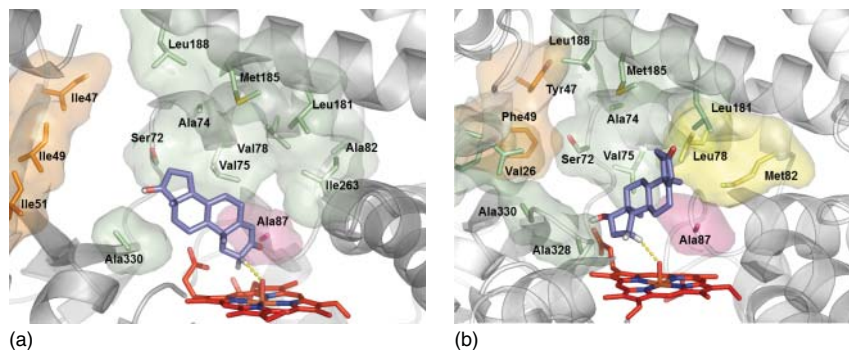


Figure 8.11 Computed pose of testosterone (5) explaining 2 β -selectivity (mutant R47I/T49I/F87A) and the respective pose leading to 15 β -selectivity (mutant

R47Y/T49F/V78L/A82M/F87A) [29a, 30a]. (Roiban and Reetz [29a]. Reproduced with permission of Royal Society of Chemistry.)

points toward the heme-Fe=O, while in the case of the 15 β -variant it is the pro- β H-atom at the 15C-position, which is closest to the catalytic center [30a]. Similar analyses have been performed in the successful attempt to unveil the source of regio- and stereoselectivity in reactions of smaller unnatural substrates leading to (*R*)- and (*S*)-enantiomers [30b–f].

In summary, the directed evolution of P450-BM3 and other CYPs, flanked by MD/docking computations, have thrown light on the mechanism of these enzymes [29]. The function of the immediate protein environment at the active site is not so much to stabilize the rate-determining transition state, but to position the substrate in a pose, which leads to the observed regio- and stereoselectivity of mutants generated by directed evolution. Standard MD/docking software is generally successful [29, 30], but many studies do not include such theoretical analyses. In the case of directed evolution of P450-BM3 as the catalyst in the oxidative hydroxylation of methylcyclohexane, a variant was produced leading to the simultaneous creation of two centers of chirality with high regio- as well as diastereo- and enantioselectivity [30e]. Amazingly, this substrate contains no functional groups that could participate in binding, and still all three types of selectivity are high. Interestingly, standard MD/docking computations were not successful in predicting the experimental outcome. This shows that more refined theoretical approaches need to be tested (or developed!), a challenging conclusion. Perhaps more than one molecule of such small hydrophobic substrates needs to be considered in the binding pocket of CYPs. Following the above publications, a revealing study of WT P450-BM3 as the catalyst in fatty acid hydroxylation appeared, in which MD and QM/MM computations were shown to predict substrate-induced gating as well as regio- and enantioselectivity [31c].

8.3

Additive versus Non-additive Mutational Effects in Fitness Landscapes

Biochemists and molecular biologists have applied site-specific mutagenesis as introduced by Smith [34] for a long time, specifically in the quest to identify the catalytic machinery [34, 35]. In many early studies, two point mutations were introduced separately for mechanistic purposes, their individual contribution to the catalytic profile (e.g., activity) being measured and subsequently compared to the mutational effect by combining them into the respective double mutant [36]. In most cases at the time, additive effects regarding mutation X and mutation Y were found according to the Fersht equation (Eq. 8.1) in which the “additional” free energy $\Delta\Delta G_I$ is zero.

$$\Delta\Delta G_{(X,Y)} = \Delta\Delta G_{(X)} + \Delta\Delta G_{(Y)} + \Delta\Delta G_{(I)} \quad (8.1)$$

This means that the two mutations, being independent, do not interact with one another, as shown by a survey of a number of studies at the time [36c] (Figure 8.12). Further examples of classical mutational additivity were reported later [37].

However, exceptions were noted early on and subsequently thereafter. It was suggested that such rare non-additivity is induced when the side-chains of the two amino acids in the double mutant are in close spatial vicinity to one another [36]. In an investigation of mutational effects in dihydrofolate, a visionary conclusion was made: “...with an enhanced knowledge of the molecular origin of non-additive effects, it may be possible to optimize an approach to improve

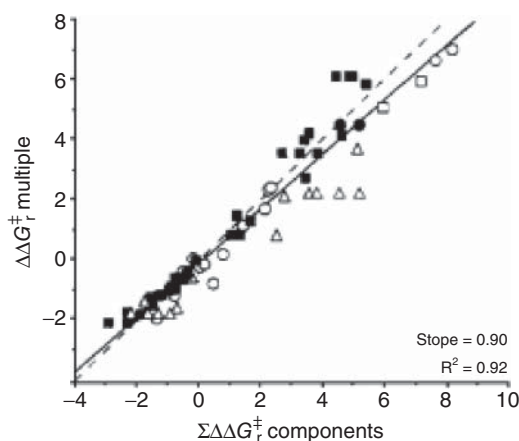


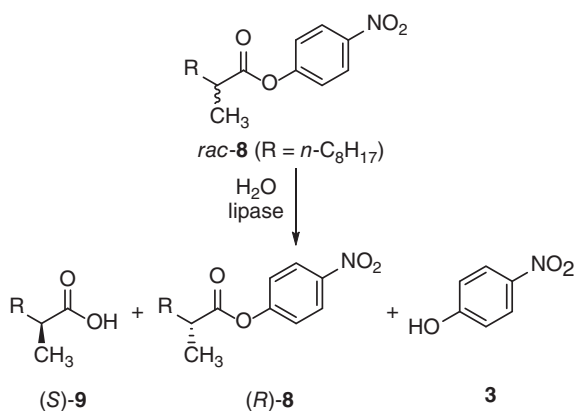
Figure 8.12 Changes in transition-state stabilization energies for the multiple mutant versus the sum of the component mutants [36c]. The data represent mutants from subtilisin, tyrosyl-tRNA synthetase, trypsin, DHFR, and glutathione

reductase. The dashed line has a slope of 1 representing perfect additivity, and the solid line corresponds to the best fit of the data. (Wells [36c]. Reproduced with permission of American Chemical Society.)

the enzyme's efficiency by coupling mutations" [36j,k]. As shown in the following discourse, this challenging goal has been reached.

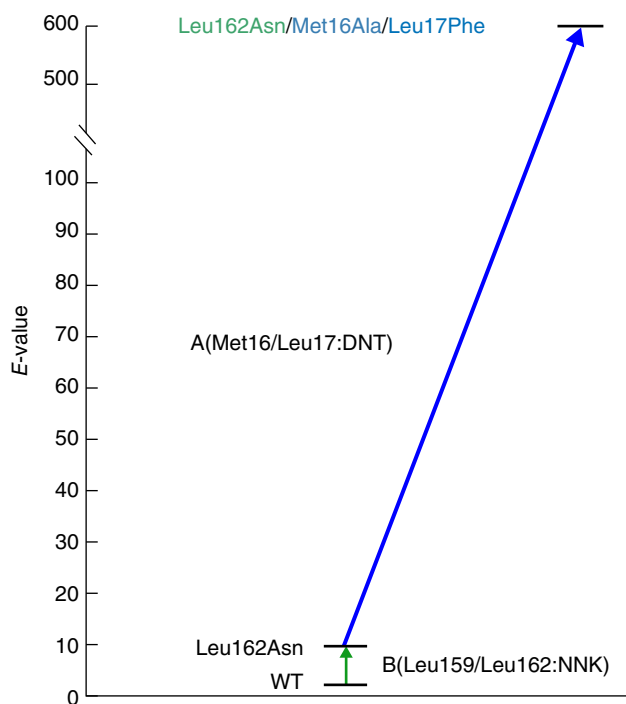
It should be noted that non-additivity may imply either a positive cooperative mutational effect, or a deleterious one in which the positive influence of each point mutation is partially or fully destroyed when combining them. With the current level of theory, it is difficult to predict which type of epistasis will result in such experiments, if any. In contrast to the classical double mutant free-energy cycles [36], directed evolution usually involves multiple rounds of mutagenesis/expression/screening with the accumulation of several point mutations. Predicting additivity or non-additivity between individual mutations or between sets of mutations is even more difficult, if not impossible. In fact, in the absence of deconvolution experiments, only the catalytic effect of the first mutation(s) is accessible experimentally in addition to the effect of all mutations acting in concert in the final variant.

Partial or ideally full deconvolution of a multi-mutational variant constitutes a unique way to learn from directed evolution. An eye-opening example concerns the hydrolytic kinetic resolution of *rac*-**8** with preferential formation of (*S*)-**9** catalyzed by the lipase from *Pseudomonas aeruginosa* (PAL) (Scheme 8.9) [38, 39], this transformation being the model reaction of a series of studies in which different mutagenesis strategies were compared (see Chapter 4). At an intermediate stage of methodology development, a variant characterized by six point mutations, introduced in several cycles by epPCR, saturation mutagenesis, and DNA shuffling, led to a selectivity factor of $E=51$ [39a]. Subsequently, a QM/MM study predicted that four of the six point mutations are superfluous, and indeed the respective double mutant proved to be even better ($E=63$), for which a reasonable model was proposed on the basis of a relay effect originating from a remote position and extending to a residue next to the active site [39b].



Scheme 8.9 Model reaction catalyzed by the lipase PAL [38, 39]. (Reetz *et al.* [38]. Reproduced with permission of American Chemical Society.)

In a study that included deconvolution experiments, ISM was applied to the same model system, leading to a dramatically improved variant, while screening considerably less transformants (for details, see Chapter 4) [38]. A 3-site ISM scheme composed of two-residue sites A, B, and C was designed, the best pathway B $\rightarrow$ A leading to the final triple mutant composed of Leu162Asn obtained by saturation mutagenesis at site B (Leu159/Leu162) and of Met16Ala/Leu17Phe by subsequent saturation mutagenesis at site A (Met16/Leu17) (Scheme 8.10). The final best variant 1B2 (Leu162Asn/Met16Ala/Leu17Phe) showed a selectivity factor of $E = 594$ in favor of (*S*)-9 [38]. In addition to enhanced stereoselectivity, the reaction rate of the preferred enantiomer (*S*)-8 and, therefore, of the overall kinetic resolution was increased notably: WT PAL ($k_{\text{cat}} = 37 \times 10^{-3} \text{ s}^{-1}$; $k_{\text{cat}}/K_{\text{m}} = 43.5 \text{ s}^{-1} \text{ M}^{-1}$) versus variant 1B2 ($k_{\text{cat}} = 1374 \times 10^{-3} \text{ s}^{-1}$; $k_{\text{cat}}/K_{\text{m}} = 4041 \text{ s}^{-1} \text{ M}^{-1}$). In this case study, higher activity clearly correlates with higher enantioselectivity.



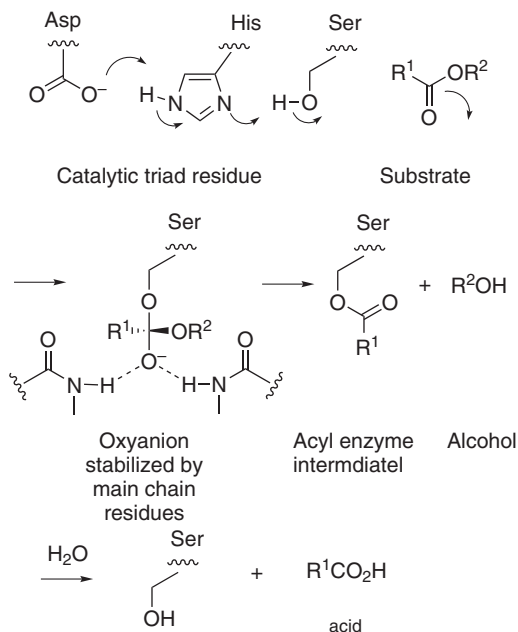
Scheme 8.10 Best ISM pathway B $\rightarrow$ A leading to the triple mutant 1B2 (Leu162Asn/Met16Ala/Leu17Phe) displaying a selectivity factor of $E = 594$ in the hydrolytic

kinetic resolution of *rac*-8 with preferential formation of (*S*)-9 [38]. (Reetz *et al.* [38]. Reproduced with permission of American Chemical Society.)

It is clear that the effect of the first mutational change (Leu162Asn) alone enhances enantioselectivity slightly to $E = 8$ (*S*), but the influence of the second set of mutations alone is not accessible from the original data unless

deconvolution is performed. Upon preparing and testing the respective double mutant Met16Ala/Leu17Phe, it was discovered that this variant by itself hardly improves enantioselectivity relative to WT PAL, the selectivity factor amounting to only $E = 2.6$ (S). If additivity were to be operating, then the selectivity factor would be expected to be $E \approx 22$. Since it is much larger ($E = 594$), a notable cooperative non-additive effect is involved amounting to an energy contribution of $\approx 2 \text{ kcal mol}^{-1}$, which is significant [38]. Relative to WT PAL, the calculated difference in stabilization energy of the two enantiomers amounts to about 3 kcal mol^{-1} . Complete deconvolution by generating and testing separately the two single mutants Met16Ala and Leu17Phe was not investigated in this study.

Upon deconvoluting the final mutants in other CAST- or ISM-based directed evolution studies, positive epistatic effects were also uncovered [40]. A short review of this emerging phenomenon in directed evolution has appeared [41]. The occurrence of strongly cooperative mutational effects in ISM seems to be the underlying factor responsible for the efficacy of this approach to protein engineering. In the present case (Scheme 8.10), a second lesson was learned upon unveiling the reason for the synergistic effect (more than additivity) on a molecular level. The mechanism of PAL involves the catalytic triad Asp229/His251/Ser82, which enables rateselectivity- and stereoselectivity-determining nucleophilic addition of activated Ser82 to the carbonyl function of esters with formation of short-lived oxyanion intermediates, followed by rapid product formation (Scheme 8.11). It is the typical lipase mechanism [42].



Scheme 8.11 Mechanism of lipase-catalyzed hydrolysis of esters.

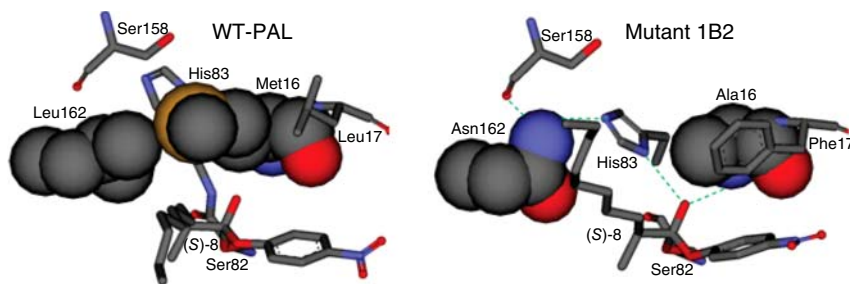


Figure 8.13 Comparison of the oxyanions with bound (*S*)-substrate at the catalytically active Ser82 of WT PAL (a) versus best variant 1B2 (b) [38]. (Reetz *et al.* [38]. Reproduced with permission of American Chemical Society.)

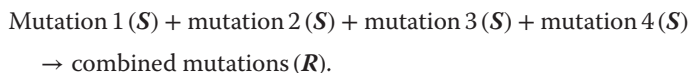
MD and docking computations were performed using the crystal structure of WT PAL [43] as the starting structure, with the three point mutations then being introduced by a docking program [38]. Substrates (*R*)- and (*S*)-8 were then introduced in the PAL-binding pocket as the respective oxyanions covalently bound to Ser82. Figure 8.13 shows the case of the favored (*S*)-substrate bound in WT PAL and in mutant 1B2.

Upon comparing the two models, several important features stand out that provide significant insight. Firstly, in WT PAL, the bulky side-chain of Leu162 clashes with the *n*-octyl moiety of the ester, in contrast to the situation in variant 1B2 in which the position of Asn162 provides sufficient space for the long alkyl chain of (*S*)-8. It can be argued that the sidechain of asparagine is not that much smaller than that of leucine, but in the present case another factor is involved. In 1B2, Asn162 forms a hydrogen bond to Ser158, thereby positioning the asparagine sidechain farther away from the bound substrate, which avoids steric clashes. This also explains the increase in activity when going from WT to variant 1B2.

Secondly, the substitution Met16 → Ala16 provides more space in a sterically congested part of WT PAL, which allows more sidechain flexibility of His83. The new pose enables additional stabilization of the oxyanion by way of hydrogen bond formation (Figure 8.13b). It should be noted that Ser158 and His83 are not part of the catalytically active triad Asp229/His251/Ser82, which means that in WT PAL they are “innocent” residues. In variant 1B2, however, they participate in a H-bond network that involves Asn162, Ser158, His83 and the O-atom of the oxyanion (Figure 8.13b) [38].

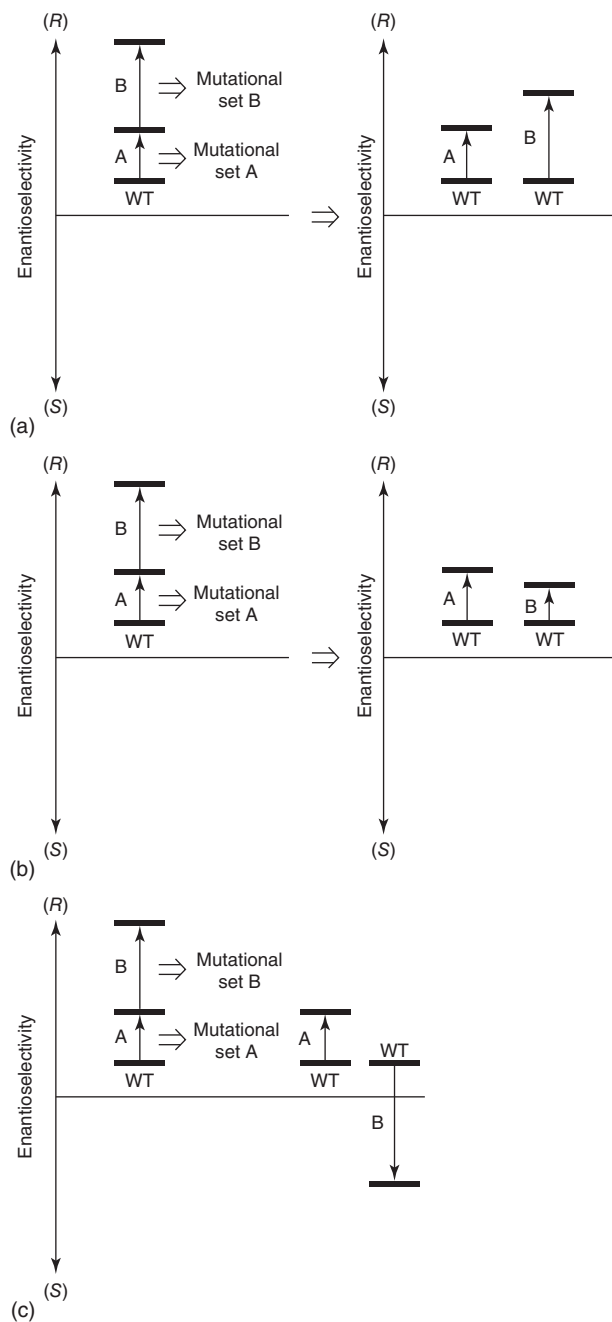
Thirdly, the substitution Leu17 → Phe17 enables π -stacking between the phenyl sidechain of phenylalanine and the *p*-nitrophenyl moiety of the ester substrate, which is not possible in WT PAL (Figure 8.13). This model explains the origin of the observed strong cooperative effect brought about by the ISM mutational substitutions. Neither mutation Leu162Asn nor Met16Ala/Leu17Phe by themselves exert little influence, but in concert more than additivity results. Finally, upon modeling the disfavored substrate (*R*)-8 into variant 1B2, the respective methyl group at the stereogenic center points “upward,” thereby preventing His83 from stabilizing the oxyanion. This explains the high (*S*)-selectivity.

A limited number of other studies focusing on stereoselectivity of enzymes have appeared in which deconvolution experiments were performed, thereby uncovering unusual mutational effects [39, 40]. In one study, an ISM-evolved quadruple mutant of a Baeyer–Villiger monooxygenase obtained in two ISM steps with accumulation of two new mutations in each cycle, was used as the catalyst in asymmetric sulfoxidation of methyl tolyl thio-ether (95% ee in favor of the (*R*)-sulfoxide starting from WT showing 90% ee of opposite (*S*)-enantioselectivity) [40a]. Deconvolution with formation of the respective four single mutants led to a surprising result. All of the four single mutants showed (*S*)-selectivity, which constitutes a counter-intuitive finding:



A common technique in protein engineering is to combine separately generated point mutations in order to improve a catalytic property (Chapters 4 and 5). However, it has been observed that this procedure does not always work for reasons that are not well understood [44]. By going the reverse way, that is, deconvoluting multi-mutational variants, some insight as delineated here has been gained. Using enantioselectivity as the catalytic parameter, additive and non-additive mutational effects as revealed by deconvolution experiments can be systematized (Scheme 8.12). This illustration features the case of an initial set of mutations A followed by a second set of mutations B that accumulated in a hypothetical directed evolution experiment, irrespective of the mutagenesis method. Deconvolution of the two sets with generation of B alone can in principle signal classical additivity (Scheme 8.12a). In this case, A and B do not interact with one another, and both favor the same direction of enantioselectivity, for example,, (*R*). Several kinds of non-additivity are possible. In one type, deconvolution reveals that the contribution of B is less than expected, but the sense of enantioselectivity is the same as displayed by A (Scheme 8.12b). A second type of non-additivity is uncovered upon deconvolution with formation of B alone, which favors the opposite enantiomer (Scheme 8.12c).

The situation becomes even more complex when deconvoluting both sets of mutations A and B individually with formation of the respective single mutants, which may show different types of non-additivity or classical additivity. Complete deconvolution does not only dissect a multi-mutational variant into the respective single mutants, it also calls for the generation of all theoretically possible combinations of point mutations (double, triple mutants, etc.). When these are prepared by site-specific mutagenesis and used as catalysts in an enantioselective transformation, it is possible to construct a complete fitness pathway landscape that features the mapping of all theoretically possible pathways leading from WT to the final best mutant. This type of fitness pathway landscape was constructed in the directed evolution project of the abovementioned Baeyer–Villiger monooxygenase as catalyst in the asymmetric sulfoxidation of methyl tolyl thio-ether [40a]. The final reversed (*R*)-selective quadruple mutant was fully deconvoluted, making the construction of $4! = 24$ evolutionary pathways



Scheme 8.12 Systematization of additive and non-additive mutational effects in protein engineering, in this scheme using two sets of mutations A and B, illustrated by employing enantioselectivity as the catalytic parameter. (a) Classical additive mutational effect;

(b) non-additive mutational effect in which set B shows lower than expected enantioselectivity but in the same direction; and (c) non-additive effect in which mutational set B shows reversed enantioselectivity.

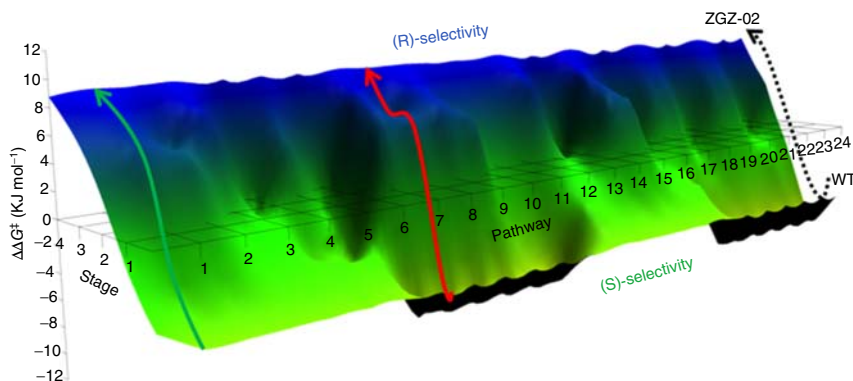


Figure 8.14 Fitness pathway landscape showing the 24 pathways leading from WT PAMO (bottom) to best (*R*)-selective variant ZGZ-2 in asymmetric sulfoxidation, a typical trajectory lacking local

minima (green pathway) and one having local minima (red) being featured [40a]. (Zhang *et al.* [40a]. Reproduced with permission of American Chemical Society.)

possible (Figure 8.14) [40a]. Out of the 24 trajectories, 6 proved to lack any local minima, while 18 displayed such local “valleys” (Figure 8.15).

In an earlier study in which the identical procedure was applied to an enantioselective epoxide hydrolase, a five-step ISM pathway was deconvoluted, which enabled the construction of the respective fitness landscape characterized by $5! = 120$ pathways [45]. A total of 55 trajectories proved to be characterized by the absence of any local minima. In the other 65 pathways, an inferior set of mutations was necessarily utilized, thereby mapping trajectories that terminate with the same final variant, as all other pathways. When comparing the fitness pathway landscapes of the Baeyer–Villiger [40a] and the epoxide hydrolase [45] studies, it can be seen that the respective number of pathways characterized by the absence of local minima differs considerably: 6 out of 24 (25%) versus 55 out of 120 (46%), respectively. There is no reason to expect similar percentages. Each enzyme system, and particularly the mutagenesis method and strategy, will lead to different results. It needs to be emphasized that this type of “constrained” fitness pathway landscape, which maps all trajectories connecting WT with a given previously evolved best mutant, is different from exploring all theoretically possible trajectories of a defined ISM system (“unconstrained” fitness pathway landscape) as delineated in Figure 4.6 [46]. To date, this is the only study in which all pathways of an ISM system have been explored experimentally.

Although both “constrained” and “unconstrained” fitness pathway landscapes were constructed and analyzed in order to learn how to increase the efficacy of directed evolution of enzymes for use in organic chemistry and biotechnology [40a,c, 41, 45, 46], the results touch on a very different research field, namely evolutionary biology. Using a mutant of a β -lactamase characterized by five point mutations evolved previously, an experimental selection-based platform was devised on which a “constrained” fitness pathway landscape featuring $5! = 120$

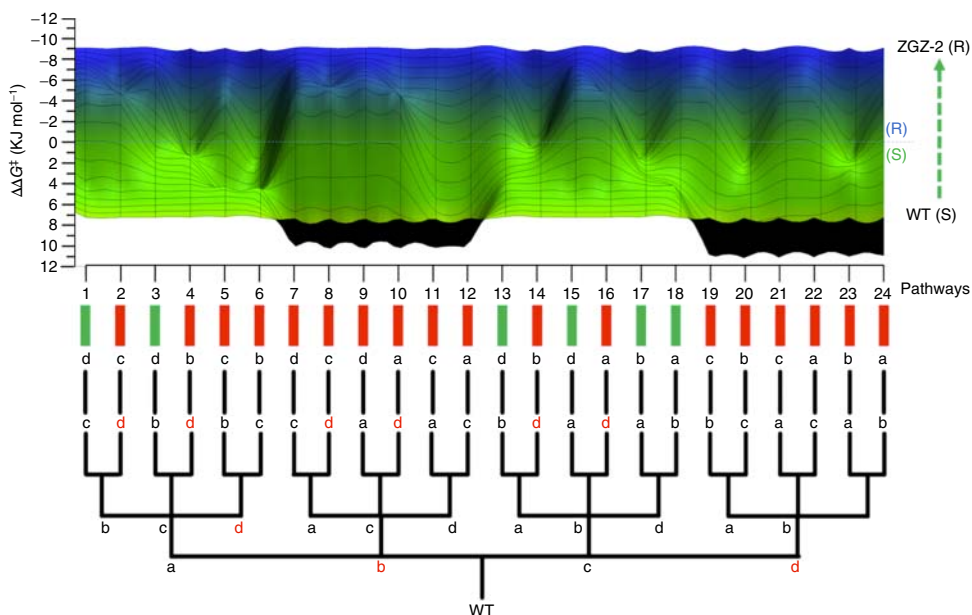


Figure 8.15 Fitness pathway landscape in the frontal view of Figure 8.4 of all 24 trajectories leading from WT PAMO to variant ZGZ-2 characterized by four point mutations [40a]. Green notations indicate energetically favored pathways, whereas red notations

represent disfavored trajectories having local minima. Letters in red in the dendrogram denote a local minimum after the introduction of this mutation. (Zhang *et al.* [40a]. Reproduced with permission of American Chemical Society.)

pathways was constructed [47a], inspired by an earlier investigation [47b]. Four of the mutations were based on Stemmer's first report on DNA shuffling [48]. The purpose of the study was to address the question of how many pathways in natural evolution lead from a starting gene to an evolved mutant gene. It turned out that of the 120 trajectories connecting WT with the quintuple mutant, 102 are characterized by local minima. This led the authors to conclude that "Darwinian evolution can follow only very few mutational pathways to fitter proteins" [47a]. They described these 102 pathways as being "inaccessible to Darwinian selection" and stated that "many of the remaining trajectories have negligible probabilities of realization" [47a]. In a somewhat different experimental setup, an alternative conclusion was reached [49].

The reader is advised to study the original publications for more details regarding the experimental platform [47]. The question arises whether the general conclusion of this study [47a] contradicts the results of the subsequent studies featured above [40a,c, 41, 45, 46]. Would the application of ISM or some other molecular biological method for evolving a β -lactamase variant with five (different) point mutations lead to a higher number of pathways to fitter proteins, and therefore to a different conclusion? It needs to be pointed

out that the experimental platforms used in the β -lactamase studies [47] and in the subsequent studies [40a,c, 41, 45, 46] are quite different and, therefore, not directly comparable, one utilizing selection based on bacterial survival and the others employing screening based on analytical assays, respectively. Nevertheless, care must be taken when exploiting the results of laboratory (directed) evolution for drawing general conclusions regarding the nature of Darwinian evolution.

References

1. Fischer, M., Kang, M., and Brindle, N.P. (2016) Using experimental evolution to probe molecular mechanisms of protein function. *Protein Sci.*, **25**, 352–359.
2. (a) Warshel, A. (2014) Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture). *Angew. Chem. Int. Ed.*, **53** (38), 10020–10031; (b) Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.*, **48** (7), 1198–1229; (c) Mulholland, A.J. (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today*, **10** (20), 1393–1402; (d) Pu, J., Gao, J., and Truhlar, D.G. (2006) Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem. Rev.*, **106** (8), 3140–3169; (e) Chung, L.W., Sameera, W.M.C., Ramozzi, R., Page, A.J., Hatanaka, M., Petrova, G.P., Harris, T.V., Li, X., Ke, Z., Liu, F., Li, H.-B., Ding, L., and Morokuma, K. (2015) The ONIOM method and its applications. *Chem. Rev.*, **115**, 5678–5796; (f) Blomberg, M.R.A., Borowski, T., Himio, F., Liao, R.-Z., and Siegbahn, P.E.M. (2014) Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.*, **114**, 3601–3658.
3. (a) Polyak, I., Reetz, M.T., and Thiel, W. (2013) Quantum mechanical/molecular mechanical study on the enantioselectivity of the enzymatic Baeyer-Villiger reaction of 4-hydroxycyclohexanone. *J. Phys. Chem. B*, **117** (17), 4993–5001; (b) Polyak, I., Reetz, M.T., and Thiel, W. (2012) Quantum mechanical/molecular mechanical study on the mechanism of the enzymatic Baeyer-Villiger reaction. *J. Am. Chem. Soc.*, **134** (5), 2732–2741.
4. Reetz, M.T., Bocola, M., Wang, L.W., Sanchis, J., Cronin, A., Arand, M., Zou, J., Archelas, A., Bottalla, A.L., Naworyta, A., and Mowbray, S.L. (2009) Directed evolution of an enantioselective epoxide hydrolase: uncovering the source of enantioselectivity at each evolutionary stage. *J. Am. Chem. Soc.*, **131** (21), 7334–7343.
5. Reetz, M.T., Wang, L.W., and Bocola, M. (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. Ed.*, **45** (8), 1236–1241.
6. Zou, J., Hallberg, B.M., Bergfors, T., Oesch, F., Arand, M., Mowbray, S.L., and Jones, T.A. (2000) Structure of *Aspergillus niger* epoxide hydrolase at 1.8 Å resolution: implications for the structure and function of the mammalian microsomal class of epoxide hydrolases. *Structure*, **8** (2), 111–122.
7. Morisseau, C., Archelas, A., Guittton, C., Faucher, D., Furstoss, R., and Baratti, J.C. (1999) Purification and characterization of a highly enantioselective epoxide hydrolase from *Aspergillus niger*. *Eur. J. Biochem.*, **263** (2), 386–395.
8. Bruice, T.C. (2002) A view at the millennium: the efficiency of enzymatic catalysis. *Acc. Chem. Res.*, **35** (3), 139–148.
9. Chandrasekhar, J., Smith, S.F., and Jorgensen, W.L. (1985) Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.*, **107** (1), 154–163.
10. Stork, G., Cama, L.D., and Coulson, D.R. (1974) Epoxynitrile cyclization – general method of ring formation. *J. Am. Chem. Soc.*, **96** (16), 5268–5270.

11. (a) Na, J., Houk, K.N., Shevlin, C.G., Janda, K.D., and Lerner, R.A. (1993) The energetic advantage of 5-exo versus 6-endo epoxide openings – a preference overwhelmed by antibody catalysis. *J. Am. Chem. Soc.*, **115** (18), 8453–8454; (b) Laitinen, T., Rouvinen, J., and Peräkylä, M. (1998) Ab initio quantum mechanical and density functional theory calculations on nucleophile- and acid-catalyzed opening of an epoxide ring: a model for the covalent binding of epoxyalkyl inhibitors to the active site of glycosidases. *J. Org. Chem.*, **63** (23), 8157–8162.
12. Lind, M.E. and Himó, F. (2013) Quantum chemistry as a tool in asymmetric biocatalysis: limonene epoxide hydrolase test case. *Angew. Chem. Int. Ed.*, **52** (17), 4563–4567.
13. Sun, Z., Lonsdale, R., Kong, X.-D., Xu, J.-H., Zhou, J., and Reetz, M.T. (2015) Reshaping an enzyme binding pocket for enhanced and inverted stereoselectivity: use of smallest amino acid alphabet in directed evolution. *Angew. Chem. Int. Ed.*, **54**, 12410–12415.
14. (a) Gatti, F.G., Parmeggiani, F., and Sacchetti, A. (2014) in *Synthetic Methods for Biologically Active Molecules* (ed. E. Brenna), Wiley-VCH Verlag GmbH, Weinheim, pp. 27–47; (b) Winkler, C.K., Tasnádi, G., Clay, D., Hall, M., and Faber, K. (2012) Asymmetric bioreduction of activated alkenes to industrially relevant optically active compounds. *J. Biotechnol.*, **162** (4), 381–389; (c) Bougioukou, D.J. and Stewart, J.D. (2012) in *Enzyme Catalysis in Organic Synthesis* (eds K. Drauz, H. Gröger, and O. May), Wiley-VCH Verlag GmbH, Weinheim, pp. 1111–1203; (d) Hall, M. and Bommaris, A.S. (2011) Enantioenriched compounds via enzyme-catalyzed redox reactions. *Chem. Rev.*, **111** (7), 4088–4110.
15. (a) Kille, S. and Reetz, M.T. (2013) in *Synthetic Methods for Biologically Active Molecules-Exploiting the Potential of Bioreductions* (ed. E. Brenna), Wiley-VCH Verlag GmbH, Weinheim, pp. 139–181; (b) Amato, E.D. and Stewart, J.D. (2015) Applications of protein engineering to members of the old yellow enzyme family. *Biotechnol. Adv.*, **33** (5), 624–631; (c) Toogood, H. and Scrutton, N.S. (2013) Enzyme engineering toolbox – a ‘catalyst’ for change. *Catal. Sci. Technol.*, **3**, 2182–2194.
16. (a) Walton, A.Z., Conerly, W.C., Pompeu, Y., Sullivan, B., and Stewart, J.D. (2011) Biocatalytic reductions of Baylis-Hillman adducts. *ACS Catal.*, **1** (9), 989–993; (b) Pompeu, Y.A., Sullivan, B., and Stewart, J.D. (2013) X-ray crystallography reveals how subtle changes control the orientation of substrate binding in an alkene reductase. *ACS Catal.*, **3** (10), 2376–2390; (c) Brenna, E., Crotti, M., Gatti, F.G., Monti, D., Parmeggiani, F., Powell, R.W., Santangelo, S., and Stewart, J.D. (2015) Opposite enantioselectivity in the bioreduction of (Z)- β -aryl- β -cyanoacrylates mediated by the tryptophan 116 mutants of old yellow enzyme 1: synthetic approach to (R)- and (S)- β -aryl- γ -lactams. *Adv. Synth. Catal.*, **357** (8), 1849–1860.
17. Fox, K.M. and Karplus, P.A. (1994) Old yellow enzyme at 2-Å resolution – overall structure, ligand-binding, and comparison with related flavoproteins. *Structure*, **2** (11), 1089–1105.
18. Padhi, S.K., Bougioukou, D.J., and Stewart, J.D. (2009) Site-saturation mutagenesis of tryptophan 116 of *Saccharomyces pastorianus* old yellow enzyme uncovers stereocomplementary variants. *J. Am. Chem. Soc.*, **131** (9), 3271–3280.
19. Fraaije, M.W. and Mattevi, A. (2000) Flavoenzymes: diverse catalysts with recurrent features. *Trends Biochem. Sci.*, **25** (3), 126–132.
20. Horita, S., Kataoka, M., Kitamura, N., Nakagawa, T., Miyakawa, T., Ohtsuka, J., Nagata, K., Shimizu, S., and Tanokura, M. (2015) An engineered old yellow enzyme that enables efficient synthesis of (4R,6R)-actinol in a one-pot reduction system. *ChemBioChem*, **16** (3), 440–445.
21. Lonsdale, R. and Reetz, M.T. (2015) Reduction of α,β -unsaturated ketones by old yellow enzymes: mechanistic insights from quantum mechanics/molecular mechanics calculations. *J. Am. Chem. Soc.*, **137**, 14733–14742.

22. (a) Fitzpatrick, T.B., Amrhein, N., and Macheroux, P. (2003) Characterization of YqjM, an Old Yellow Enzyme homolog from *Bacillus subtilis* involved in the oxidative stress response. *J. Biol. Chem.*, **278** (22), 19891–19897; (b) Stuermer, R., Hauer, B., Hall, M., and Faber, K. (2007) Asymmetric bioreduction of activated C=C bonds using enoate reductases from the Old Yellow Enzyme family. *Curr. Opin. Chem. Biol.*, **11** (2), 203–213.
23. Schmidt, M., Hasenpusch, D., Kahler, M., Kirchner, U., Wiggernhorn, K., Langel, W., and Bornscheuer, U.T. (2006) Directed evolution of an esterase from *Pseudomonas fluorescens* yields a mutant with excellent enantioselectivity and activity for the kinetic resolution of a chiral building block. *ChemBioChem*, **7** (5), 805–809.
24. Cheeseman, J.D., Tocilj, A., Park, S., Schrag, J.D., and Kazlauskas, R.J. (2004) Structure of an aryl esterase from *Pseudomonas fluorescens*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, **60** (Pt. 7), 1237–1243.
25. (a) Ortiz de Montellano, P.R. (2005) *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd edn, Springer, Berlin; (b) Isin, E.M. and Guengerich, F.P. (2007) Complex reactions catalyzed by cytochrome P450 enzymes. *Biochim. Biophys. Acta, Gen. Subj.*, **1770** (3), 314–329; (c) Munro, A.W., Girvan, H.M., and McLean, K.J. (2007) Variations on a (t)heme--novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat. Prod. Rep.*, **24** (3), 585–609; (d) Ortiz de Montellano, P.R. (2010) Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.*, **110** (2), 932–948; (e) Shaik, S., Cohen, S., Wang, Y., Chen, H., Kumar, D., and Thiel, W. (2010) P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem. Rev.*, **110** (2), 949–1017; (f) Urlacher, V.B. and Girhard, M. (2012) in *Enzyme Catalysis in Organic Synthesis*, 3rd edn, vol. 3 (eds K. Drauz, H. Gröger, and O. May), Wiley-VCH Verlag GmbH, Weinheim, pp. 1227–1267.
26. (a) Lonsdale, R., Harvey, J.N., and Mulholland, A.J. (2010) Inclusion of dispersion effects significantly improves accuracy of calculated reaction barriers for cytochrome P450 catalyzed reactions. *J. Phys. Chem. Lett.*, **1** (21), 3232–3237; (b) Lonsdale, R., Harvey, J.N., and Mulholland, A.J. (2010) Compound I reactivity defines alkene oxidation selectivity in cytochrome P450cam. *J. Phys. Chem. B*, **114** (2), 1156–1162.
27. Hogg, J.A. (1992) Steroids, the steroid community, and Upjohn in perspective: a profile of innovation. *Steroids*, **57** (12), 593–616.
28. (a) Fasan, R., Chen, M.M., Crook, N.C., and Arnold, F.H. (2007) Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew. Chem. Int. Ed.*, **46** (44), 8414–8418; (b) Xu, F., Bell, S.G., Lednik, J., Insley, A., Rao, Z., and Wong, L.L. (2005) The heme monooxygenase cytochrome P450cam can be engineered to oxidize ethane to ethanol. *Angew. Chem. Int. Ed.*, **44** (26), 4029–4032.
29. (a) Roiban, G.D. and Reetz, M.T. (2015) Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.*, **51** (12), 2208–2224; (b) Whitehouse, C.J., Bell, S.G., and Wong, L.L. (2012) P450(BM3) (CYP102A1): connecting the dots. *Chem. Soc. Rev.*, **41** (3), 1218–1260; (c) Lewis, J.C., Coelho, P.S., and Arnold, F.H. (2011) Enzymatic functionalization of carbon-hydrogen bonds. *Chem. Soc. Rev.*, **40** (4), 2003–2021; (d) Fasan, R. (2012) Tuning P450 enzymes as oxidation catalysts. *ACS Catal.*, **2** (4), 647–666; (e) Khatri, Y., Hannemann, F., Girhard, M., Kappl, R., Meme, A., Ringle, M., Janocha, S., Leize-Wagner, E., Urlacher, V.B., and Bernhardt, R. (2013) Novel family members of CYP109 from *Sorangium cellulosum* So ce56 exhibit characteristic biochemical and biophysical properties. *Biotechnol. Appl. Biochem.*, **60** (1), 18–29; (f) Holtmann, D., Fraaije, M.W., Arends, I.W., Opperman, D.J.,

- and Hollmann, F. (2014) The taming of oxygen: biocatalytic oxyfunctionalisations. *Chem. Commun.*, **50** (87), 13180–13200.
30. (a) Kille, S., Zilly, F.E., Acevedo, J.P., and Reetz, M.T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.*, **3** (9), 738–743; (b) Agudo, R., Roiban, G.D., and Reetz, M.T. (2012) Achieving regio- and enantioselectivity of P450-catalyzed oxidative CH activation of small functionalized molecules by structure-guided directed evolution. *ChemBioChem*, **13** (10), 1465–1473; (c) Agudo, R., Roiban, G.-D., Lonsdale, R., Ilie, A., and Reetz, M.T. (2015) Biocatalytic route to chiral acyloins: P450-catalyzed regio- and enantioselective α -hydroxylation of ketones. *J. Org. Chem.*, **80** (2), 950–956; (d) Roiban, G.D., Agudo, R., Ilie, A., Lonsdale, R., and Reetz, M.T. (2014) CH-activating oxidative hydroxylation of 1-tetralones and related compounds with high regio- and stereoselectivity. *Chem. Commun.*, **50** (92), 14310–14313; (e) Roiban, G.D., Agudo, R., and Reetz, M.T. (2014) Cytochrome P450 catalyzed oxidative hydroxylation of achiral organic compounds with simultaneous creation of two chirality centers in a single C-H activation step. *Angew. Chem. Int. Ed.*, **53** (33), 8659–8663; (f) Ritter, C., Nett, N., Acevedo-Rocha, C.G., Lonsdale, R., Kräling, K., Dempwolff, F., Hoebenreich, S., Graumann, P.L., Reetz, M.T., and Meggers, E. (2015) Bioorthogonal enzymatic activation of caged compounds. *Angew. Chem. Int. Ed.*, **54**, 13440–13443.
 31. (a) Narhi, L.O. and Fulco, A.J. (1986) Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J. Biol. Chem.*, **261** (16), 7160–7169; (b) Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., Page, C.C., and Dutton, P.L. (2002) P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem. Sci.*, **27** (5), 250–257; (c) Dubey, K.D., Wang, B., and Shaik, S. (2016) Molecular dynamics and QM/MM calculations predict the substrate-induced gating of cytochrome P450 BM3 and the regio- and stereoselectivity of fatty acid hydroxylation. *J. Am. Chem. Soc.*, **138**, 837–845.
 32. (a) Haines, D.C., Tomchick, D.R., Machius, M., and Peterson, J.A. (2001) Pivotal role of water in the mechanism of P450BM-3. *Biochemistry*, **40** (45), 13456–13465; (b) Haines, D.C., Chen, B., Tomchick, D.R., Bondlela, M., Hegde, A., Machius, M., and Peterson, J.A. (2008) Crystal structure of inhibitor-bound P450 BM-3 reveals open conformation of substrate access channel. *Biochemistry*, **47** (12), 3662–3670.
 33. Weber, E., Seifert, A., Antonovici, M., Geinitz, C., Pleiss, J., and Urlacher, V.B. (2011) Screening of a minimal enriched P450 BM3 mutant library for hydroxylation of cyclic and acyclic alkanes. *Chem. Commun.*, **47** (3), 944–946.
 34. (a) Hutchison, C.A. III., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.*, **253** (18), 6551–6560; (b) Smith, M. (1994) Synthetic DNA and biology (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.*, **33** (12), 1214–1221.
 35. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W.H. Freeman and Company, New York.
 36. (a) Carter, P.J., Winter, G., Wilkinson, A.J., and Fersht, A.R. (1984) The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell*, **38** (3), 835–840; (b) Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P., and Estell, D.A. (1987) Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **84** (5), 1219–1223; (c) Wells, J.A. (1990) Additivity of mutational effects in proteins. *Biochemistry*, **29** (37), 8509–8517; (d) Horovitz, A. (1996) Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des.*, **1** (6), R121–R126; (e) Shortle, D. (1992) Mutational studies of protein structures

- and their stabilities. *Q. Rev. Biophys.*, **25** (2), 205–250; (f) Laskowski, M. Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M.W., Kohr, W.J., Park, S.J., Parks, K., Schatzley, B.L. *et al.* (1987) Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry*, **26** (1), 202–221; (g) Nelson, H.C. and Sauer, R.T. (1985) Lambda repressor mutations that increase the affinity and specificity of operator binding. *Cell*, **42** (2), 549–558; (h) Mildvan, A.S., Weber, D.J., and Kuliopulos, A. (1992) Quantitative interpretations of double mutations of enzymes. *Arch. Biochem. Biophys.*, **294** (2), 327–340; (i) Mildvan, A.S. (2004) Inverse thinking about double mutants of enzymes. *Biochemistry*, **43** (46), 14517–14520; (j) Huang, Z., Wagner, C.R., and Benkovic, S.J. (1994) Nonadditivity of mutational effects at the folate binding site of *Escherichia coli* dihydrofolate reductase. *Biochemistry*, **33** (38), 11576–11585; (k) Wagner, C.R., Huang, Z., Singleton, S.F., and Benkovic, S.J. (1995) Molecular basis for nonadditive mutational effects in *Escherichia coli* dihydrofolate reductase. *Biochemistry*, **34** (48), 15671–15680.
37. (a) Skinner, M.M. and Terwilliger, T.C. (1996) Potential use of additivity of mutational effects in simplifying protein engineering. *Proc. Natl. Acad. Sci. U.S.A.*, **93** (20), 10753–10757; (b) Aita, T., Uchiyama, H., Inaoka, T., Nakajima, M., Kokubo, T., and Husimi, Y. (2000) Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. *Biopolymers*, **54** (1), 64–79; (c) Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.*, **15** (5), 403–411; (d) Sullivan, B.J., Nguyen, T., Durani, V., Mathur, D., Rojas, S., Thomas, M., Syu, T., and Magliery, T.J. (2012) Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability. *J. Mol. Biol.*, **420** (4–5), 384–399; (e) Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R., and Gaillardin, C. (2002) Engineering the thermostability of *Bacillus licheniformis* α -amylase. *Biologia*, **57**, 203–211.
 38. Reetz, M.T., Prasad, S., Carballeira, J.D., Gumulya, Y., and Bocola, M. (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.*, **132**, 9144–9152.
 39. (a) Reetz, M.T., Wilensek, S., Zha, D., and Jaeger, K.-E. (2001) Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angew. Chem. Int. Ed.*, **40** (19), 3589–3591; (b) Reetz, M.T., Puls, M., Carballeira, J.D., Vogel, A., Jaeger, K.-E., Eggert, T., and Thiel, W. (2007) Learning from directed evolution: further lessons from theoretical investigations into cooperative mutations in lipase enantioselectivity. *ChemBioChem*, **8**, 106–112.
 40. Recent examples of cooperative non-additive mutational effects in directed evolution of stereoselective enzymes [41]: (a) Zhang, Z.-G., Lonsdale, R., Sanchis, J., and Reetz, M.T. (2014) Extreme synergistic mutational effects in the directed evolution of a Baeyer-Villiger monooxygenase as catalyst for asymmetric sulfoxidation. *J. Am. Chem. Soc.*, **136**, 17262–17272; (b) Bartsch, S., Kourist, R., and Bornscheuer, U.T. (2008) Complete inversion of enantioselectivity towards acetylated tertiary alcohols by a double mutant of a *Bacillus subtilis* esterase. *Angew. Chem. Int. Ed.*, **47**, 1508–1511; (c) Reetz, M.T., Soni, P., Acevedo, J.P., and Sanchis, J. (2009) Creation of an amino acid network of structurally coupled residues in the directed evolution of a thermostable enzyme. *Angew. Chem. Int. Ed.*, **48**, 8268–8272; (d) Sandström, A.G., Wikmark, Y., Engström, K., Nyhlén, J., and Bäckvall, J.-E. (2012) Combinatorial reshaping of the *Candida antarctica* lipase a substrate pocket for enantioselectivity using an extremely condensed

- library. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (1), 78–83.
41. Reetz, M.T. (2013) The importance of additive and non-additive mutational effects in protein engineering. *Angew. Chem. Int. Ed.*, **52** (10), 2658–2666.
 42. Bornscheuer, U.T. and Kazlauskas, R.J. (2005) *Hydrolases in Organic Synthesis: Regio- and Stereoselective Biotransformations*, 2nd edn, Wiley-VCH Verlag GmbH, Weinheim.
 43. Nardini, M., Lang, D.A., Liebeton, K., Jaeger, K.-E., and Dijkstra, B.W. (2000) Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation: the prototype for family I.1 of bacterial lipases. *J. Biol. Chem.*, **275** (40), 31219–31225.
 44. Reetz, M.T., Carballeira, J.D., Peyralans, J., Höbenreich, H., Maichele, A., and Vogel, A. (2006) Expanding the substrate scope of enzymes: combining mutations obtained by CASTing. *Chem. Eur. J.*, **12**, 6031–6038.
 45. Reetz, M.T. and Sanchis, J. (2008) Constructing and analyzing the fitness landscape of an experimental evolutionary process. *ChemBioChem*, **9**, 2260–2267.
 46. Gumulya, Y., Sanchis, J., and Reetz, M.T. (2012) Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *ChemBioChem*, **13** (7), 1060–1066.
 47. (a) Weinreich, D.M., Delaney, N.F., DePristo, M.A., and Hartl, D.L. (2006) Darwinian evolution can follow only very few mutational pathways to fitter proteins. *Science*, **312**, 111–114; (b) Hall, B.G. (2002) Predicting evolution by in vitro evolution requires determining evolutionary pathways. *Antimicrob. Agents Chemother.*, **46**, 3035–3038.
 48. Stemmer, W.P.C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370** (6488), 389–391.
 49. 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**, 1193–1196.

Index

a

Adaptive Substituent Reordering Algorithm (ASRA) 99
 additive mutational effects 287
 Agilent technology 93
 aggregation 213
Agrobacterium radiobacter epoxide hydrolase, amino acid residues 250
 alcohol dehydrogenases (ADHs) 178–179
 aldehyde dehydrogenases 179–180
 aldolases 167, 184
 AlleyCat 255
 amino acid
 – residues 128
 – residues in *A. radiobacter* epoxide hydrolase 250
 – saturation mutagenesis 142
 amino acid bias 59, 62–65, 67
 amino acid dehydrogenases 180–181
 aminoacyl-tRNA synthetase 186
 aminoglycoside aminotransferase 185
 ANT computational tool 96
 antibody engineering 16, 187
 apo-myoglobin 243
 aptamers 4
Arthrobacter sp. transaminase 131
 artificial metalloenzymes 84, 237–260
 artificial shuffling 88
Aspergillus niger epoxide hydrolase (ANEH)
 134, 136, 138, 269, 271
 – epoxide hydrolase 269
 – feruloyl esterase from 209
 – mechanism 270
 – rac-1 catalyzed by 269
 assembly of designed oligonucleotides (ADO)
 89
 – benefits and limitations 90
 – concept 90

avidin-Whitesides system 238
 axial chirality 178

b

Bacillus megaterium, CYP 284
Bacillus stearothermophilus host organism 6
Bacillus subtilis spore display; laccase; lipase
 205, 210
 – CNA 225
 – epPCR and 63, 64
 Baeyer-Villiger monooxygenases (BVMOs)
 167, 178
 β -galactosidase (BGAL) 119
 – X-ray data of 119
 – structure 121
 – variants 121
 B-FIT approach 211
 B-FIT based thermostabilization 135
 B-FITTER computational tool 97
 biased mutation-assembly 87–88
 BVMOs 167, 178

c

Candida antarctica A lipase (CALA)
 145–149, 169
Candida antarctica B lipase (CALB) 75, 91,
 169
Candida macedoniensis ene-reductase
 AKU4588 (CmOYE) 277
 carbon dioxide fixation 16
 carbonic anhydrase 171
 carotenoid synthase 187
 cassette mutagenesis 7
 CASTER 96
 CASTp computational tool 96
 catalyst properties, simultaneous optimization
 117
 catalytic antibodies 187

- CAVER computational tool 97
 - cellulases 167
 - chorismate mutase 77
 - circular permutation 91
 - codon degeneracy 131, 142, 147, 150
 - KCA 144
 - NDT 136, 143, 148
 - NNK 131, 134
 - Combinatorial Active-Site Saturation Test (CAST) 75, 125, 127, 131, 135, 168, 245
 - residues 129
 - Combinatorial Multiple-Cassette Mutagenesis (CMCM) 86, 87, 125
 - combinatorial solid-phase gene synthesis 159
 - continuous evolution 4
 - computational tools 96–100
 - consensus approach 70, 216
 - Constrained Network Analysis (CNA) approach 224
 - ConSurf computational tool 96
 - cooperative effects 116, 130
 - cooperative mutational effects 287–295
 - CorNet computational tool 96
 - CRISPR-Cas9 nuclease 16
 - 22c-trick mutagenesis method 154, 159
 - Cu(II)-phthalocyanine 240
 - CUPSAT computational tool 97
 - cutinase 182
 - cyclodextrin glycosyltransferase 167
 - cyclododecanone monooxygenase (CDMO) 144
 - cyclopropanation 258, 259
 - Fe-carbenoid mediated 259
 - cytochrome P450 monooxygenases (CYPs) 282
- d**
- deacetoxycephalosporin C synthase 186
 - decarboxylases 183
 - deep mutational scanning 44
 - Degrado-algorithm 219, 226
 - depurination 5
 - Diels-Alder cycloadditions 240, 244
 - dioxygenases 181
 - directed evolution
 - enzymes, *see* enzymes directed evolution
 - history of 4
 - directed evolution stereoselectivity concept 14
 - display systems 35
 - divide and combine approach 226, 227
 - DNA shuffling 12, 118–121, 125, 136
 - application 207
 - epPCR and 125, 130–132
 - experiment 125
 - single-gene shuffling *versus* 86
 - DNA polymerases 16, 186
 - domain swapping 92
 - double codon saturation mutagenesis (DCSM) 152
 - dual selection system 36
 - Dynamic Disulfide Discovery (DDD) 221
- e**
- enantioselectivity 115, 238
 - PAL 123
 - ene-reductase, OYE 273
 - ene-reductases 182, 273–279
 - enzyme activity 116
 - enzyme ancestral resurrection 237
 - enzyme-coupled assay 47
 - enzyme promiscuity 237–260
 - enzymes, directed evolution
 - applications of 16
 - definition 1
 - history of 4
 - selected examples 167
 - epoxide hydrolase (EchA) 171–172, 250, 269
 - error-prone polymerase chain reaction (epPCR) 5, 9, 14, 60–70, 122, 139
 - application 9, 207
 - conventional 10
 - cycle 122
 - DNA shuffling 125, 130, 132
 - four cycles of 14
 - illustration 60
 - whole-gene mutagenesis techniques 60
 - error-prone rolling circle amplification (RCA) 66
 - esterases 167, 169, 279–281
 - expanded genetic code 67, 84
- f**
- FamClash computational tool 96
 - family shuffling 86
 - fatty acids 284
 - Fe-carbenoid mediated cyclopropanation 259
 - Fe-Fe hydrogenase 182
 - FireProt approach 223
 - fitness landscapes 137–139, 292–295
 - flavin reductase 182
 - fluorescence-activated cell sorter (FACS) technology 13
 - FoldXR computational tool 97
 - Framework for Rapid Enzyme Stabilization by Computational libraries (FRESCO) 221

fructosyl amine oxidase 182
 fucosidase, from galactosidase 119

g

galactase oxidase 177
 galactosidase, fucosidase from 119
 gas chromatography high-throughput assay
 49–50, 52
 gene mutagenesis methods 59, 115
 – circular permutation 91
 – combinatorial solid-phase gene synthesis
 159
 – comparative studies 118
 – computational tools 96
 – epPCR 60
 – ISM 135
 – recombinant 85
 – saturation mutagenesis 70, 130, 142, 154
 – SCSM 149
 – solid-phase combinatorial 92
 – TCSM 151
 genetically modified plants 16
 genetic complementation 4, 28
 genetic complementation *Geobacillus*
kaustophilus error prone strain 205
 glucose oxidase 178
 GLUE/GLUE-IT computational tools 96
 glycosidases 172
 glycoside hydrolase 139–141
 glycosyltransferases 167, 174
 glyphosate degradation 181
 growth selection pressure 29–35
 Grubbs-Hoveyda Ru-catalyst 259, 260

h

haloalkane dehalogenase 183
 halohydrin dehalogenase 183
 history of directed evolution 4–16
 homology model 123
 homophenylalanine dehydrogenase 29
 HotSpot Wizzard computational tool 96
 HPLC high-throughput assay 49, 52
 human estrogen sulfotransferase 185
 human glutathione transferase 185
 human serum albumin (HSA) 240, 241
 hybrid catalysts 238
 hydroxynitrilase lyases 67, 185
 hypermutagenic PCR 62

i

immunofluorescence, antibody-based 141
 incorporating synthetic oligonucleotides via
 gene reassembly (ISOR) 87, 89
 in vitro coevolution strategy 118

in vitro compartmentalization (IVC) 42
 IPRO computational tool 96
 ISOR mutagenesis method 87, 89
 isotopically labeled pseudo-enantiomers 48
 ITCHY mutagenesis method 87
 iterative saturation mutagenesis (ISM) 76,
 123, 125–127, 133–135
 – B-FIT 212, 213
 – free energy 138
 – optimal pathways 135
 – pathways 135
 – promiscuity 246, 247
 – protein-protein interfacial sites 215
 – reminiscent of 168
 – Sp3GH98 141

k

kanamycin nucleotidyltransferase 6
 Kemp elimination 251, 252, 255, 256
 Kunkel mutagenesis method 5, 66

l

laccases 177
 L-aspartic acid 9
 limonene epoxide hydrolase (LEH) 273
 lipase 14–15, 135, 167–169
 local minima 136–139, 213
 local structural entropy (LSE) 226
 lock-and-key Fischer hypothesis 75, 124
Lucilia cuprina 207, 208

m

malonyl-CoA synthase 187
 manganese Schiff base complex 243
 MAX strategy 75
 MBLOSUM computational tool 96
 megaprimer approach 9, 72
 megaprimer PCR 72, 75, 94
 megaprimer PCR of whole plasmid
 (MEGAWHOP) 63
 meso-epoxides 148
 metabolic pathway engineering 16
 Michaelis-Menten equation 269
 microfluidic devices 42–43
 microtiter plates 1
 Mn-corroles 240
 molecular biological techniques 118
 molecular biology, saturation mutagenesis and
 148
 molecular dynamics (MD) 291
 – calculations 271
 molecular mechanics (MM) methods 268,
 278
 monoamine oxidases 181–182

- multiple sequence alignment (MSA) 70, 142, 143, 217
- multiplexing GC and HPLC 52
- mutagenesis method 6
- mutational effects, in fitness landscapes 287
- mutator strain 6, 182
- myoglobin-based Cr-catalyst 242

- n**
- N*-acyl acrylic acid 238
- N*-acyl amino acid racemase (NAAAR) 29
- Natto-kinase 186
- nitrilases 172
- NMR, B-FIT approach 212
- non-additive mutational effects 287–295
- noncontiguous recombination (NCR) 219, 220
- non-overlapping oligonucleotides 73
- Nov statistics 80, 82–83
- nucleases 167

- o**
- OE-PCR, *see* overlap extension polymerase chain reaction (OE-PCR)
- oil-in-droplet emulsions 42–43, 87
- Old Yellow Enzymes (OYE) 273
- olefin metathesis 260
- oligonucleotide gene shuffling (DOGS) 87
- oligonucleotide shuffling 12
- OmniChange saturation mutagenesis 75
- OSCARR mutagenesis method 75
- overlap extension polymerase chain reaction (OE-PCR) 8, 11, 71, 73
- oversampling 77
- oversampling factor 80
- oxidative stability 7, 206

- p**
- P450-BM3 sequencing 157
- Patrick/Firth algorithm 13, 77, 79
- penicillin G acylase 170
- phage display 13, 187
- phenyl acetone monooxygenase (PAMO) 143–145, 147, 148
 - crystal structure 143
- phenylalanine aminomutase 185
- phosphonate inhibitor 34
- phosphorylases 174
- phosphotriesterases 167, 171
- phytases 170
- P450 monooxygenases 167, 175–177, 258, 281–286
- pollution control 16
- polyketide synthase 187
- pooling strategies in screening 50
- Polymerase chain reaction (PCR)
 - amplification 62, 63
 - hypermutagenic 62
 - megaprimer 72, 75, 94
 - with Sloning libraries 95
- PoPMuSiC algorithm 208
- PoPMuSiC computational tool 97
- primer design 150, 154
- primer extension 7
- product inhibition 1, 179
- promiscuity 237
 - catalytic behavior 238
 - catalytic profile 245
 - ISM 246, 247
 - P450-BM3 258
 - Whitesides system 238, 245, 247
 - X-ray crystallography 242
- ProSAR computational guide 97–99
- PROSIDE method 206
- protein engineering study 205
- protein-protein interfaces 215
- protein sequence-activity relationships (ProSAR) 98, 99
- protein thermostabilization 205, 226
 - divide and combine approach 227
 - FRESCO 97, 221
 - PcDTE 216
- ProxiMAX method 75
- pseudo-enantiomers 32
- Pseudomonas aeruginosa* lipase (PAL) 14, 122, 288
 - binding pocket 124
 - comparative studies of 126
 - crystal structure 124, 127
 - directed evolution 122
 - enantioselectivity 123
- Pseudomonas aeruginosa* esterase 37
- Pseudomonas cichorii* D-tagatose-3-epimerase (PcDTE) 215, 216
- Pseudomonas fluorescens* esterase (PFE) 248
 - catalytic profiles 282
 - enantioselectivity 279
 - rac-3 catalyzed 280
 - 3D homology model 281
- pyrrolysyl-tRNA synthetase 185, 186

- q**
- QM/MM methods 268
- quantitative structure-activity relationships (QSARs) 98
- quantum mechanics (QM) methods 268, 278
- Quick Quality Control (QQC) 81, 84, 135, 154, 156, 158

- QuikChange™ protocol 67, 71, 72
 Q-values for library quality control 81, 154, 156
- r**
- random chimeragenesis on transient templates (RACHITT) 87
 random drift mutagenesis (RNDM) 87
 random insertion/deletion (RID) mutagenesis 67, 70
 RDA-PCR mutagenesis method 87
 recombined extension truncated templates (RETT) 87
 recombinant gene mutagenesis methods 84–92
 reduced amino acid alphabets 77–82
 reductive amination 180, 131–133, 183
 ribosomal display 13
 RNA 4
 RNA polymerase 186
 robotic colony picker 1
 robustness of enzymes 205
 – ancestral/consensus approaches 216
 – B-FIT approach 211
 – CNA approach 224
 – DNA shuffling 207
 – epPCR 207
 – FireProt approach 223
 – FRESCO approach 97, 221
 – ISM 215
 – SCHEMA approach 219
 rolling circle amplification (RCA) 66, 67
 Rosetta algorithms 101, 219, 244, 250
- s**
- Saccharomyces cerevisiae* phosphatase 4
 saturation mutagenesis (SM) 70, 120, 121, 123, 131
 – application of 205
 – CALA binding pocket 146
 – efficient method for 74
 – four libraries 132
 – guidelines 130
 – at multi-residue site 131
 – in protein engineering 154
 – purpose 70
 – SCSM 149
 – systematization 142
 – TCSM 151
 – techno-economical analyses 154
 SCADS computational tool 226
 SCHEMA approach 219
 SCRATCHY mutagenesis method 87
 screening systems 44–52
 selection systems 27–44
 sequence-independent site-directed chimeragenesis (SISDC) 89
 sequence saturation mutagenesis (SeSaM) 67
 – advantage 67
 – stages 68
 serum paraoxonase 87
 single code saturation mutagenesis (SCSM) 149
 single codon saturation mutagenesis (SCSM) 150
 single-gene shuffling 86
 single-stranded DNA (ssDNA) 86
 SIRCH computational tool 96
 site saturation mutagenesis (SSM) 70
 site-directed mutagenesis
 – application of 205
 – QuikChange™ protocol 67
 site-specific mutagenesis 7, 168
 Sloning approach 93, 95
 SM-based gene mutagenesis methods 168
 solid-phase combinatorial gene synthesis 92–95
 Sp3GH98
 – crystal structure 141
 – evolutionary pathways 141
 splicing by overlap extension (SOE) 11
 staggered extension process (StEP) 86, 87
 statistical computationally assisted design strategy (SCADS) 226
Stenotrophomonas maltophilia flavoprotein monooxygenase (SMFMO) 250
 step economy in directed evolution 148
 stereoselective enzyme, selected examples 167
 stereoselectivity 116
 strategies for gene mutagenesis applications 115
 streptavidin 246
 structure-based combinatorial protein engineering (SCOPE) 87
 substrate walking 117, 118, 131, 132
 sulfoxidation 176, 180, 242
 SwiftLib computational tool 96
 synergistic mutational effects 287–295
- t**
- tabacco epi-aristolochene synthase (TEAS) 226
 tagatose epimerase 175
 Tang mutagenesis approach 154, 158–159
Taq polymerase 60, 61, 67, 186

techno-economical analysis of saturation

mutagenesis 154

TEM-1 β -lactamase 27

testosterone 285, 286

Thermoethanolicus brockii alcohol
dehydrogenase (TbSADH) 152

thermostability 115

– optimization 116, 205–227

Thermotoga maritima thermophilic organism,
tHisF 242

Thermus aquaticus, Taq polymerase epPCR
60

thiamine diphosphate enzymes 185

tHisF 243, 244

toluene monooxygenases 180

TopLib computational tool 80, 96

transaminases 131–133, 183

transketolases 184

Trichoderma reesei cellulase, B-FIT approach
213, 215

triple code saturation mutagenesis (TCSM)
151

two-parameter optimization 116–117

u

universal blood 140

UV-Vis plate reader 44–45

v

vaccine production 16

VisualCNA 225

w

Whitesides system, promiscuity 238, 245,
247

wild-type (WT) 1

– *Aspergillus niger* epoxide hydrolase 134,
136, 138, 269, 271

– ene-reductase 273

– LEH 149

– and mutant BPO-A1 haloperoxidases 209

– PAL 14, 122, 123

– TEAS 226

x

X-ray, BGAL 119

X-ray crystallography, promiscuity 242

xylanase 174, 210

y

yeast display 13

z

Zinc-finger proteins 167