

Protein Engineering as a Tool for the Development of Novel Bioproduction Systems

Uwe T. Bornscheuer

Abstract Protein engineering is the most important tool to overcome the limitations of natural enzymes to facilitate their efficient use in biocatalysis, but also in biotransformations using engineered whole cell systems. In the past decade numerous novel concepts have been developed to create mutant libraries more efficiently, taking advantage of improved molecular biology methods and better expression systems for the target proteins. Most important are sophisticated designs of mutant libraries guided by bioinformatic tools. These achievements were complemented with the development of advanced high-throughput screening or selection methods using, for instance, fluorescence-activated cell sorting (FACS) or microfluidic systems. The latter are prime examples useful to establish novel bioproduction systems as cell-free protein biosynthesis allows protein expression and subsequent sorting of desired enzyme variants with these methods.

Keywords Bioinformatics · Directed evolution · Enantioselectivity · High-throughput screening · Protein engineering · Rational protein design · Substrate spectrum

Abbreviations

BLAST	Basic local alignment search tool
CAST	Combinatorial active site-test
E	Enantioselectivity
FACS	Fluorescence-activated cell sorting
FADS	Fluorescence-activated droplet sorting
HRP	Horseradish peroxidase
ISM	Iterative saturation mutagenesis
IVC	In vitro compartmentalization
IVTT	In vitro transcription and translation

U. T. Bornscheuer (✉)

Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry,
Greifswald University, Felix-Hausdorff-Str. 4,
17487 Greifswald, Germany
e-mail: uwe.bornscheuer@uni-greifswald.de

PCR	Polymerase chain reaction
PFE	<i>Pseudomonas fluorescens</i> esterase
QM/MM	Quantum mechanical/molecular mechanics
WT	Wildtype

Contents

1	Introduction.....	26
2	Basic Strategies in Protein Engineering	27
2.1	Methods for the Creation of Mutant Libraries	29
2.2	Assay Systems	31
3	Cell-Free Protein Expression	33
3.1	<i>In Vitro</i> Compartmentalization	34
3.2	Microfluidic Systems.....	35
4	Conclusions.....	37
	References.....	37

1 Introduction

For more than a century enzymes have found numerous applications as biocatalysts for synthetic organic chemistry due to their high stereoselectivity required for the creation of chiral compounds. They are also widely used for processes where mild reaction conditions are important, such as in the conversion of labile compounds, in protection/deprotection steps or where chemo- and regioselectivity are needed, in laundry applications, or food processing. Also the pretreatment of renewable resources for the degradation of lignocellulose to make biofuels recently became important.

Traditionally, only wildtype enzymes for single-step biocatalysis or microorganisms for whole cell biotransformations were used, but the advent of recombinant DNA technology enabled the cloning and functional expression of the enzyme of interest. Nowadays, complete pathways in microorganisms can be altered and this area is called metabolic engineering and/or synthetic biology.

Natural (wildtype) enzymes have evolved in nature to solve a distinct task in the host organism, which usually means that they have a specific substrate range and selectivity and are active under reaction conditions determined by the host environment. Substrate concentrations are usually very low (in the μM range) and pH- and temperature regions are rather narrow. Intracellular enzymes are often highly specific and intensively regulated by, for example, feedback inhibition whereas extracellular enzymes typically are more stable and show a broader substrate scope as their main task is the degradation of nutrients for the host (i.e., lipases split triglycerides to provide fatty acids, amylases degrade carbohydrate polymers to provide sugars).

Efficient application of enzymes in biocatalysis and biotransformations on an industrial scale requires substantial improvements of wildtype enzymes to make them robust. Typical goals of protein engineering are therefore improved substrate specificity (narrow or broader substrate scope depending on the application, altered regio- and stereoselectivity, increased reaction rate, elimination of substrate/product inhibition), and better tolerance to reaction conditions (pH, heat, solvents, reactive components, high substrate concentrations in the molar range). In addition, protein engineering can help to improve efficient protein manufacturing by improving expression and folding of active protein, to simplify purification and to achieve better storage stability. Furthermore, protein engineering is an important tool to investigate evolutionary relationships of proteins within a superfamily, to study its reaction mechanism, but also catalytic promiscuity.

Successful protein engineering for industrial applications should target these various properties simultaneously keeping in mind that changing one property (i.e., substrate scope) can also affect other properties (e.g., expression, stability). For instance, in one of our projects, we identified a highly enantioselective esterase variant obtained by directed evolution bearing three amino acid substitutions [1]. Unfortunately, although *E. coli* produced high amounts of the soluble wildtype esterase, only small amounts of the triple mutant were produced, most of it in an unfolded insoluble form. Back mutations identified a double mutant that was both highly enantioselective and efficiently produced in soluble form.

Modern tools developed for protein engineering helped to solve various issues related to the application of enzymes, which has been summarized in our recent review [2] stating that “in the past, an enzyme-based process was designed around the limitations of the enzyme; today, the enzyme is engineered to fit the process specifications.”

In this contribution first some basic concepts for protein engineering are briefly given together with some selected examples. Finally, its potential for novel bio-production systems is elucidated. Readers are referred to a broad number of books, book chapters, and review articles for a broader coverage of protein engineering [2–12].

2 Basic Strategies in Protein Engineering

Protein engineering usually follows two major concepts: rational design or directed (molecular) evolution [5, 6, 12] (Fig. 1).

Rational protein design requires the 3D-structure of a protein or at least a good homology model. A high-resolution structure of an enzyme is usually determined by x-ray crystallography and is deposited at the Brookhaven protein database (www.pdb.org). Preferentially, the structure also contains a substrate (or substrate analogue) bound to the active site as this provides detailed information about the residues involved in substrate binding and pockets binding substrate substituents. A model of the protein structure can be generated automatically based on a known

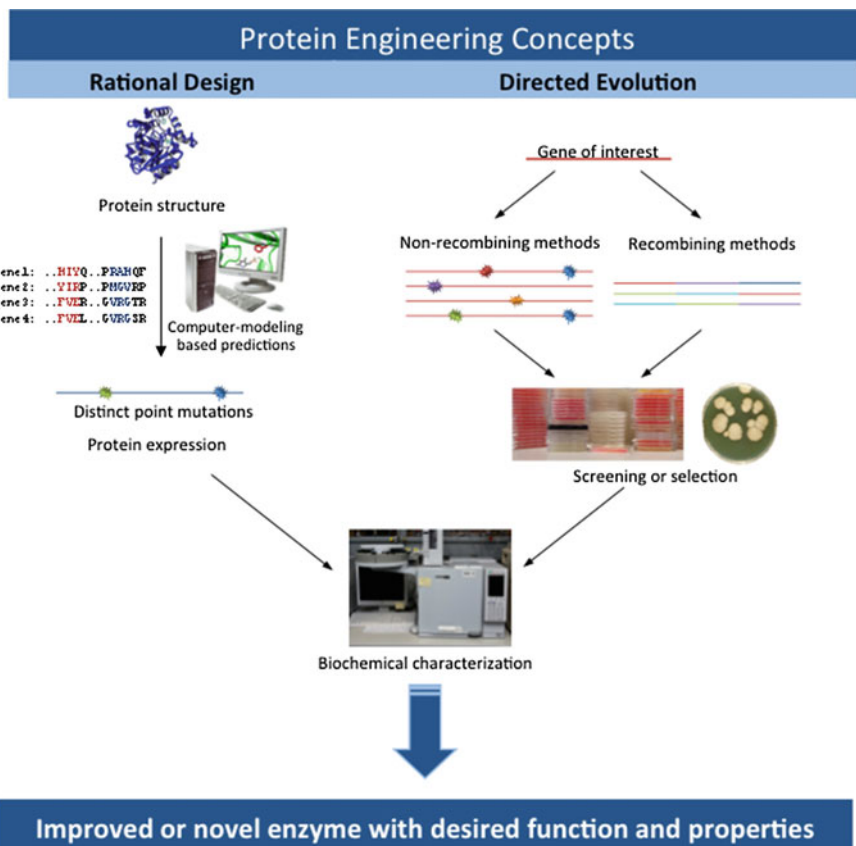


Fig. 1 Overview of protein engineering strategies

homologous x-ray structure (e.g., by Phyre [13] or Robetta [14]). More recently, a method was proposed to predict a protein structure from evolutionary and sequence variations [15]. These structures then serve as a basis to perform molecular modeling for the identification of key residues to be changed in order to improve the protein in the desired direction. Once amino acid residues are identified from a structural analysis these are introduced into the protein on the gene level by site-directed mutagenesis and after production of the mutant it can be analyzed for altered function. The success of this approach largely depends on the information available and the modeling approach used. For instance, if the substrate scope or stereoselectivity of the enzyme should be altered, a simple homology model without detailed knowledge about the binding pockets and active site region will barely enable correct predictions. If, however, a 3D-structure with a bound transition state analogue is available, then modern modeling software should allow a rather reliable prediction, which amino acids should be changed. Software for molecular modeling now runs on standard personal computers and is

rather easy to use (e.g., Yasara; [16]), tools for sequence comparison (e.g., BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi> [17]) and alignment (e.g., ClustalW: <http://www.ebi.ac.uk/Tools/msa/clustalw2/> [18, 19]) enable quick identification of novel sequences and conserved motifs. The most sophisticated rational protein design uses QM/MM (quantum mechanical/molecular mechanics) simulations, but this requires extended skills and large computational resources.

In contrast, directed evolution only requires a functional expression of the gene to produce the protein of interest. The protein-encoding gene is then subjected to random mutagenesis (e.g., by error-prone PCR) or gene shuffling resulting in very large mutant libraries. These are then either screened for the desired property—usually 10^3 – 10^4 variants in the microtiter plate format—or subjected to selection methods (e.g., by complementation of pathways in the host organism, but also FACS or microfluidic systems can be used; see Sects. 3.1 and 3.2). Best hits are usually subjected to further rounds of directed evolution until the protein shows the desired properties. Finally, the position and type of amino acid substitutions are determined and the enzyme is biochemically characterized in detail. Often back mutations or combinations are performed to further improve the enzyme or elucidate the specific influence of a given amino acid substitution. Many detailed protocols for directed evolution are available, including several books [20–23] and reviews [3, 9, 24–26].

Both strategies have their advantages and limitations: rational protein design depends on the availability of the protein 3D structure and sufficient knowledge of the contribution of residues to enzyme properties whereas directed evolution strongly depends on suitable mutagenesis methods and high-throughput screening or selection methods to identify the desired variants in a fast and reliable manner. More recently, researchers combined both methods and dubbed this focused directed evolution or semi-rational design. Thus, information available from related protein structures, families and mutants already identified, is combined and then used for targeted randomization of certain areas of the protein [8]; see below.

One challenge in directed evolution experiments is the coverage of a sufficiently large sequence space: the creation and analysis of as many variants as possible. When considering a protein (enzyme) consisting of 200 amino acids, the number of possible variants of a protein by introduction of M substitutions in N amino acids can be calculated from the formula $19^M[N!/(N-M)!M!]$ (from the 20 proteinogenic amino acids, one is already present in the wildtype enzyme; therefore the value 19). Thus, for two random mutations already more than seven million variants are possible; with three or more substitutions, the creation and screening of a library becomes very challenging.

2.1 Methods for the Creation of Mutant Libraries

The still most often used method for the creation of libraries is the error-prone PCR in which conditions are used that lead to the introduction of approximately

1–10 nucleotide mutations per 1,000 base pairs [27]. This is achieved by changing the reaction conditions, that is, to use Mn^{2+} salts instead of Mg^{2+} salts (the polymerase is magnesium-dependent), the use of *Taq* polymerase from *Thermomyces aquaticus*, and variations in the concentrations of the four desoxynucleotides (dNTPs). It should be noted, that due to the biased exchange of nucleotides and the degenerate genetic code, only an average of 6 of all possible 19 amino acid substitutions are accessible by this method. This can be overcome by the use of an optimized polymerase such as Mutazyme from Stratagene. Another approach utilizes mutator strains, for example, the *E. coli* derivative *Epicurian coli* XL1-Red, which lacks DNA repair mechanisms [28]. The introduction of a plasmid bearing the gene encoding the protein of interest leads to mutations during replication. Both methods introduce point mutations, and several iterative mutation rounds followed by identification of best variants are usually required to obtain a biocatalyst with the desired properties.

Alternatively, methods of recombination (also referred to as sexual mutagenesis) can be used. The first example was the DNA (or gene-) shuffling developed by Stemmer, in which DNase degrades the gene followed by recombination of the fragments using PCR with and without primers [29, 30]. This process mimics natural recombination and has been proven in various examples as a very effective tool to create desired enzymes. Later this method was further refined and termed “DNA family shuffling” or “molecular breeding,” enabling the creation of chimeric libraries from a family of genes.

The Arnold laboratory developed two variations of DNA shuffling: The staggered extension process (StEP) is based on a modified PCR protocol using a set of primers and short reaction times for annealing and polymerization. Truncated oligomers dissociate from the template and anneal randomly to different templates leading to recombination. Several repetitions allow the formation of full-length genes [31]. Many other methods are covered in the reviews and books cited above.

In the past few years researchers started to combine rational protein design and random mutagenesis. The major advantage is that if one knows that certain enzyme regions are linked to a desired property without a hint as to which specific amino acid substitution is the best, randomizing only a few amino acids can drastically reduce the screening effort. For instance, randomization of three positions corresponds to $20^3 = 8,000$ combinations. Keeping in mind the degenerate genetic code, about 100,000 variants need to be screened to cover all of the combinations. Although this number is still very high, it is substantially smaller than the sequence space to be covered in standard error-prone PCR libraries. Furthermore, the number of mutants can be further reduced by using the NDT codon (encoding only 12 of the 20 proteinogenic amino acids, but all types of amino acids). Another concept is the combinatorial active-site saturation test (CAST). For this, a small set of amino acids in the vicinity of the active site is chosen and mutated randomly followed by screening for best hits. One single CAST approach can result in an impressive enhancement of enzyme features as demonstrated for an epoxide hydrolase [32, 33]. ISM, iterative saturation mutagenesis, introduces an evolutionary factor into the CASTing strategy. In one

example, the best variants from a first CASTing round served as templates for the next round. Bearing in mind that mutations can have additive or cooperative effects, the order of positions to be mutated and screened might be crucial, leading to different pathways. Following all pathways is work intensive, as each step in a pathway includes a whole library that needs to be screened [34]. Reetz and coworkers have shown that only 8 paths increased the enantioselectivity of an epoxide hydrolase. ISM was also successfully used to improve the thermoactivity (from 45 to 93 °C) [35] and organic solvent stability of a lipase drastically from *Bacillus subtilis* [36]. These and further examples are nicely summarized in recent reviews by Reetz [9, 37, 38].

A bioinformatics-based approach is the ProSAR-strategy (protein sequence activity relationships) developed by Fox et al. [39]. This was used to guide the design of a halohydrin dehalogenase by data obtained from different protein engineering approaches (rational design, random mutagenesis, gene shuffling, saturation mutagenesis), sorting the single mutations into beneficial, neutral, and deleterious ones and keeping only the beneficial mutations for further rounds of mutagenesis. The best halohydrin dehalogenase contained 35 amino acid substitutions leading to a 4,000-fold increased volumetric productivity in the synthesis of a Lipitor (Atorvastatin) side chain [39]. In my group we have used the software 3DM to guide protein engineering of enzymes from the α/β -hydrolase enzyme superfamily [40, 41] and this served as basis to create “small, but smart” libraries. The 3DM allows the design of mutant libraries in a manner so that only amino acids frequently occurring at a given position are covered. This concept enabled us to improve the enantioselectivity ($E_{\text{Mutant}} = 80$, $E_{\text{WT}} = 3.2$) of an esterase from *Pseudomonas fluorescens* (PFE) by mutating only four residues near the active site. Also activity was improved 240-fold [42]. The same approach was used to improve the thermostability of PFE by 8 °C [43].

2.2 Assay Systems

The major challenge in directed evolution is the identification of desired variants within the mutant libraries. Suitable assay methods should enable a fast, very accurate and targeted identification of desired biocatalysts out of libraries comprising usually 10^3 – 10^6 mutants. In principle, two different approaches can be applied, namely screening or selection.

Selection-based systems have been used traditionally to enrich certain microorganisms. For in vitro evolution, selection methods are less frequently used as they usually can only be applied to enzymatic reactions that occur in the metabolism of the host strain. On the other hand, selection-based systems allow a considerably higher throughput compared to screening systems (see below). Often, selection is performed as a complementation: an essential metabolite is produced only by a mutated enzyme variant. For instance, a growth assay was used to identify monomeric chorismate mutases. Libraries were screened using media

lacking L-tyrosine and L-phenylalanine [44]. In a similar manner, complementation of biochemical pathways has also been used to identify mutants of an enzyme involved in tryptophan biosynthesis. HisA and TrpF (isomerases involved in the biosynthesis of histidine and tryptophan, respectively) have a similar tertiary structure, and the aminoaldose substrates used are very similar except for a different residue at the amino functionality. Using random mutagenesis and selection by complementation on media lacking tryptophan, several HisA variants that catalyze the TrpF reaction both in vivo and in vitro were identified [45]. One of these variants also retained significant HisA activity.

Stemmer's group subjected four genes of cephalosporinases (enzyme causing antibiotic resistance by hydrolysis of cephalosporins) from *Enterobacter*, *Yersinia*, *Citrobacter*, and *Klebsiella* species to error-prone PCR or DNA-shuffling. Libraries from four generations (a total of 50,000 colonies) were assayed by selection on agar plates with increasing concentrations of Moxalactam (a β -lactam antibiotic). Only those clones could survive that were able to hydrolyze the β -lactam antibiotic. The best variants from error-prone PCR gave only an eightfold increased activity, but the best chimeras from multiple gene-shuffling showed 270–540-fold resistance to Moxalactame [46]. Sequencing of a mutant revealed low homology compared to the parental genes and a total of 33 amino acid substitutions and seven crossovers were found. These changes would have been rather impossible to achieve using error-prone PCR and single-gene shuffling only; thus, these investigations demonstrated the power of DNA-shuffling.

Mutants of an esterase from *Pseudomonas fluorescens* (PFE) produced by directed evolution using the mutator strain *Epicurian coli* XL1-Red were assayed for altered substrate specificity using a selection procedure [28]. The key to the identification of improved variants acting on a sterically hindered 3-hydroxy ester—which was not hydrolyzed by the wild-type esterase—was an agar plate assay system based on pH-indicators, thus leading to a change in color upon hydrolysis of the ethyl ester. Parallel assaying of replica-plated colonies on agar plates supplemented with the glycerol derivative of the 3-hydroxy ester was used to refine the identification, because only *E. coli* colonies producing active esterases had access to the carbon source glycerol, thus leading to enhanced growth and, in turn, larger colonies. By this strategy, a double mutant was identified that efficiently catalyzed hydrolysis.

Screening-based systems (not to be confused with the use of the term “screening” for the identification of microorganisms) are much more frequently used. Due to the very high number of variants generated by directed evolution, common analytical tools such as gas chromatography and HPLC are less useful, as they are usually too time consuming. High-throughput GC–MS or NMR techniques have also been described, but these require the availability of rather expensive equipment and, in the case of screening for enantioselective biocatalysts, the use of deuterated substrates. Phage display, ribosome display, and fluorescence-activated cell sorting (FACS) are the methods of choice to screen very large mutant libraries containing on the order of $>10^6$ variants [47–49], but they are not generally applicable. Examples for the use of FACS are given in Sect. 3.2.

The most frequently used methods are based on photometric and fluorimetric assays performed in microtiter plate (MTP)-based formats in combination with high-throughput robot assistance. These allow a rather accurate screening of several tens of thousands of variants within a reasonable time, and provide information about the enzymes investigated, notably their activity, by determining initial rates or endpoints and their stereoselectivity by using both enantiomers of the compound of interest. One versatile example is the use of umbelliferone derivatives. Esters or amides of umbelliferone are rather unstable, especially at extreme pH and at elevated temperatures, but these ether derivatives are very stable as the fluorophore is linked to the substrate via an ether bond. Only after the enzymatic reaction and treatment with sodium periodate and bovine serum albumin (BSA), is the fluorophore released [50]. A few other techniques are outlined in [Sects. 3.1](#) and [3.2](#). Three recent extensive reviews cover the use of protein engineering methods to obtain better enzymes for organic synthesis [7–9].

3 Cell-Free Protein Expression

Protein engineering is commonly based on the use of a microorganism (*Escherichia coli* in most cases) to create the mutant library. This, however, can be encountered with disadvantages. Even if the enzyme of interest is functionally expressed without formation of inclusion bodies, the presence of many other enzymes in the host cell can lead to substantial background activities affecting the assay to identify the desired mutants, undesired side reactions can occur, or the amount of the recombinant enzyme is too low to measure the activity over the background. Also it can be laborious to prepare samples for the measurement as steps such as cell disruption, centrifugation, and transfer of cell lysate to an assay plate are required. Furthermore, if the protein of interest is toxic to the host, then this approach will not be applicable.

Consequently, researchers developed alternatives such as display of the mutant library on the cell surface [51–53], which could be successfully established for *E. coli* and *Saccharomyces cerevisiae*. With this surface display, for instance, enantioselective variants of an esterase could be identified [54, 55]. A more widely used method is cell free protein expression. This is based on *in vitro* transcription and translation (IVTT) of a DNA (or RNA) template using extracts from *E. coli* or wheat germ containing all components necessary for protein biosynthesis [56]. IVTT is nowadays a well-established method: several systems are commercially available, it permits the direct use of PCR templates of mutant libraries, and the protein of interest is produced in reasonable amounts and at high purity. Thus IVTT can help to overcome the issues mentioned above for whole cell expression. Furthermore, a direct linkage between genotype and phenotype can be established facilitating the identification of the encoding gene once the desired enzyme variant has been found in an assay. More details about IVTT can be found elsewhere in this issue (see [Cell-Free Systems: Functional Modules for Synthetic and](#)

[Chemical Biology](#)). IVTT also turned out to be a very useful concept for protein engineering where in vitro compartmentalization (IVC) or microfluidic systems are used for library production and screening of desired variants.

3.1 In Vitro Compartmentalization

In vitro compartmentalization (IVC) allows the physical separation of each enzyme variant in a protein-engineering experiment. In contrast to standard screening methods (see [Sect. 2.2](#)) where mutant libraries are transferred to microtiter plates to enable separate measurement of each variant, in IVC either whole cells or single proteins from IVTT productions are separated. Initially IVC was developed using *E. coli* by creating water-in-oil emulsions of tiny droplets containing statistically a single *E. coli* cell, each expressing a different enzyme variant. These droplets can have volumes in the femtoliter range and diameters around 1 μm . A 1 ml *E. coli* culture can contain $> 10^8$ cells and it was shown that IVC allows $> 10^{10}$ droplets in 1 ml of emulsion, the equivalent of $> 100,000$ microtiter plates. Later, the same principle was transferred to IVTT systems in which the components needed for protein expression, the template gene, and a substrate are encapsulated in the emulsified droplets (Fig. 2) [47, 57]. Combined with the relative ease of preparation, high stability over a broad range of temperature, pH, and salt concentrations, IVC represents a useful tool especially for miniaturization and parallelization of enzymatic reactions. However, IVC also has disadvantages. First, the amount of active protein produced per droplet is very small and hence the detection limit for the enzymatic activity can be too low to be measured properly. Second, the substrate must be entrapped in the droplets (independent whether an *E. coli* cell or IVTT are used for protein expression) when the emulsion is prepared. If the substrate is unstable over a longer period of incubation time, then the activity measurement may give misleading results. Third, a fluorescent product is usually formed to overcome the issues with the detection limit and to enable identification of desired proteins with a fluorescence-activated cell sorter (FACS). Substrate and product differ substantially, however, from the structure of the real target molecule and thus the risk is high that mutants are identified, which show optimal activity or selectivity for the surrogate substrate, but do not show identical properties towards the real substrate of interest in accordance to the first law of directed evolution “You get what you screen for.” Finally, the IVC approach only allows endpoint measurements, but barely the determination of enzyme kinetics. This becomes very important, if the wildtype enzyme already shows activity for the substrate used and variants are searched for with improved activity increasing the risk for false-positive results.

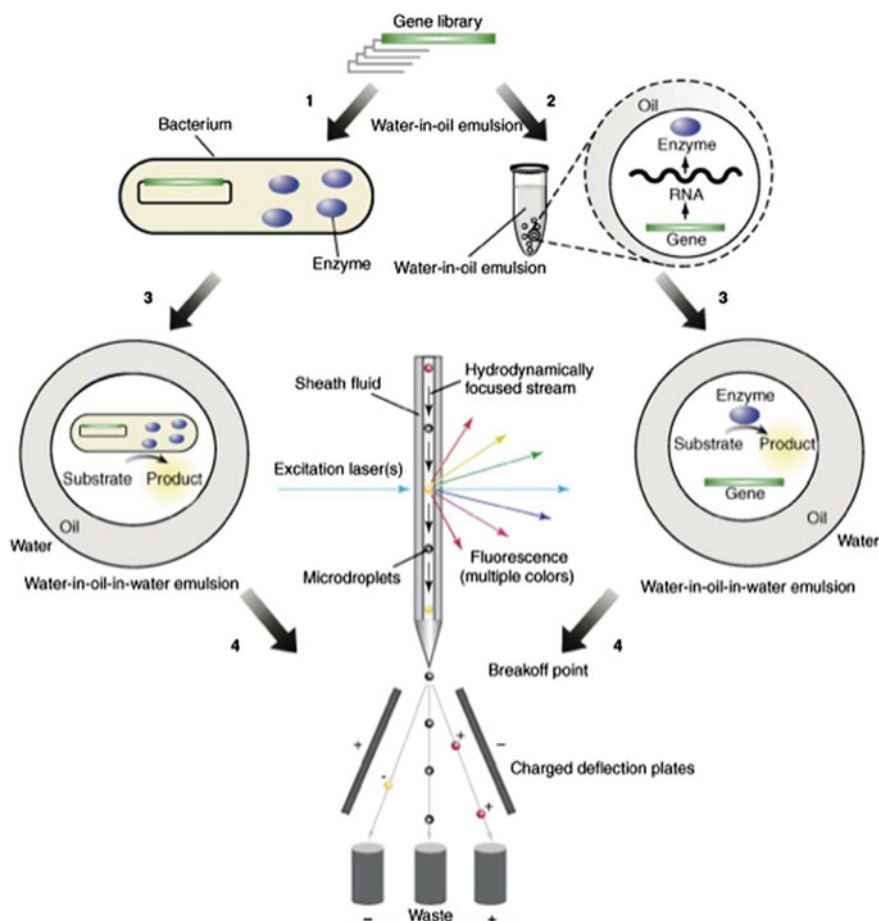


Fig. 2 Principle of compartmentalization for the selection of mutant enzymes. A gene library is either transformed into a bacterium (1) or subjected to IVTT in water-in-oil emulsions in the presence of the substrate (2). In the next step, water-in-oil-in-water emulsions are prepared (3). Enzyme variants produced in each of these compartmentalized systems convert the substrate into a fluorescent product (4). Finally, these droplets are then analyzed by FACS. Droplets containing the desired product are sorted out maintaining the physical linkage between phenotype and genotype. The nucleotide and protein sequence of the mutant can then be determined and the protein produced on a larger scale to perform proper biochemical characterization (reproduced from Ref. [57] with permission from Elsevier)

3.2 Microfluidic Systems

Microfluidic systems have recently also become important for protein engineering. This tool allows the creation of microdroplets in a controlled and sophisticated manner. For instance, aqueous microdroplets can be formed at a rate of up to 10,000 per second to form a monodisperse single emulsion or double emulsion.

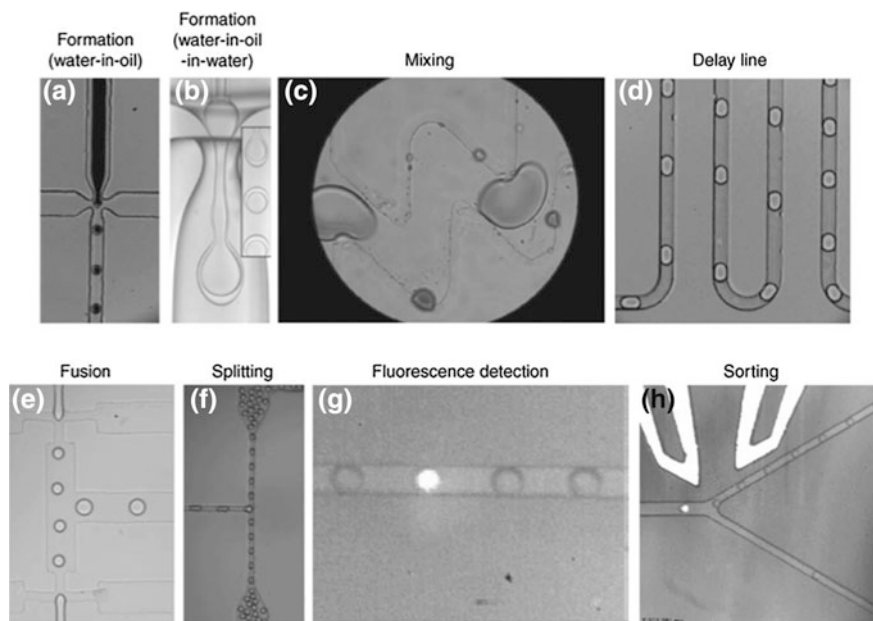


Fig. 3 Formation and manipulation of microdroplets in microfluidic systems (reproduced from Ref. [57] with permission from Elsevier)

Major advantages are that systems can be designed to fuse droplets, allowing the addition of substrate from one droplet to another droplet containing the protein of interest, then subdividing them, incubating them in delay lines to prolong enzymatic reaction times, and they can be sorted, for example, by FADS, dielectrophoresis, or electrocoalescence (Fig. 3). Thus, microfluidic systems can overcome some limitations of IVC while keeping its advantages.

The application of microfluidic systems for protein engineering has been successfully demonstrated for the directed evolution of horseradish peroxidase (HRP) [58]. Mutant libraries created by error-prone PCR were expressed on the surface of the yeast *S. cerevisiae* EBY100, encapsulated, and subjected to the microfluidic device. Analysis of an initial population of $\sim 10^7$ cells by fluorescence measurement allowed sorting the brightest droplets. Biochemical characterization revealed that after a second round of sorting, variants of HRP could be identified as being 10 times faster than the wildtype with activities close to the diffusion limit. The entire screening took less than 10 h using <150 μ l total reagent volume. This without doubt represents a dramatic reduction of costs and time compared to standard robotic systems. In another example, a microfluidic workflow was developed in which cytoplasmatic protein expression in *E. coli*, cell lysis, and reaction progress by fluorescent measurement could be established at the picoliter droplet scale using water-in-oil-droplets. This system could be used to identify arylsulfatase variants with significantly improved activity and expression levels [58]. More recently, the

Griffiths group could demonstrate that the microfluidic system can also be combined with IVTT [59]. Here single genes were compartmentalized in aqueous droplets, dispersed in an inert carrier oil, and amplified by PCR. The droplets containing 30,000 copies of each gene were then fused with droplets containing the IVTT ingredients and the substrate for the fluorogenic assay. With this concept, the authors were able to separate mixtures of *lacZ* genes encoding an active β -galactosidase from *lacZmut* genes encoding an inactive variant at an impressive speed of 2,000 droplets per second.

4 Conclusions

The methods and examples covered in this short contribution should demonstrate that protein engineering has reached a mature level to improve properties of the enzyme of interest in a rather straightforward manner. Researchers now can choose from a variety of concepts for the creation and design of mutant libraries. Furthermore, the recently developed tools for in vitro transcription and translation in combination with in vitro compartmentalization now enable searching within very large libraries ($>10^6$ members) using either FACS or microfluidic systems to find the improved variants. This can lead to a substantial reduction in costs and labor to solve a given protein engineering problem and might become competitive to rather classical systems where high-throughput screening is performed in microtiter plates.

References

1. Schmidt M, Hasenpusch D, Kähler M, Kirchner U, Wiggernhorn K, Langel W, Bornscheuer UT (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. *Chem Bio Chem* 7:805
2. Bornscheuer UT, Huisman G, Kazlauskas RJ, Lutz S, Moore J, Robins K (2012) Engineering the third wave in biocatalysis. *Nature* 485:185
3. Davids T, Schmidt M, Böttcher D, Bornscheuer UT (2013) Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr Opin Chem Biol* 17:125. <http://dx.doi.org/10.1016/j.cbpa.2013.02.022>
4. Buchholz K, Kasche V, Bornscheuer UT (2012) Biocatalysts and enzyme technology, 2nd edn. Wiley-VCH, Weinheim
5. Lutz S, Bornscheuer UT (eds) (2009) Protein engineering handbook, vol 1 and 2. Wiley-VCH, Weinheim
6. Lutz S, Bornscheuer UT (eds) (2012) Protein engineering handbook, vol 3. Wiley-VCH, Weinheim
7. Strohmeier GA, Pichler H, May O, Gruber-Khadjawi M (2011) Application of designed enzymes in organic synthesis. *Chem Rev* 111:4141
8. Behrens GA, Hummel A, Padhi SK, Schätzle S, Bornscheuer UT (2011) Discovery and protein engineering of biocatalysts for organic synthesis. *Adv Synth Catal* 353:2191

9. Reetz MT (2011) Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew Chem Int Ed* 50:138
10. Steiner K, Schwab H (2012) Recent advances in rational approaches for enzyme engineering. *Comput Struct Biotechnol J* 2:e201209010
11. Wenda S, Illner S, Mell A, Kragl U (2011) Industrial biotechnology—the future of green chemistry? *Green Chem* 13:3007
12. Kazlauskas RJ, Bornscheuer UT (2009) Finding better protein engineering strategies. *Nat Chem Biol* 5:526
13. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 4:363
14. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32:W526
15. Marks DS, Hopf TA, Sander C (2012) Protein structure prediction from sequence variation. *Nat Biotechnol* 30:1072
16. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* 77(Suppl 9):114
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403
18. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38:W695
19. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947
20. Arnold FH, Georgiou G (eds) (2003) Directed enzyme evolution: screening and selection methods. *Methods Mol Biol Humana Press*, Totawa
21. Arnold FH, Georgiou G (eds) (2003) Directed evolution library creation: methods and protocols., *Methods in Molecular Biology Humana Press*, Totawa
22. Brakmann S, Johnsson K (eds) (2002) Directed molecular evolution of proteins. Wiley-VCH, Weinheim
23. Brakmann S, Schwienhorst A (eds) (2004) Evolutionary methods in biotechnology: clever tricks for directed evolution. Wiley-VCH, Weinheim
24. Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. *Nat Chem Biol* 5:567
25. Neylon C (2004) Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acid Res* 32:1448
26. Böttcher D, Bornscheuer UT (2010) Protein engineering of microbial enzymes. *Curr Opin Microbiol* 13:274
27. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 2:28
28. Bornscheuer UT, Altenbuchner J, Meyer HH (1998) Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of Epothilones. *Biotechnol Bioeng* 58:554
29. Stemmer WPC (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Nat Acad Sci U S A* 91:10747
30. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370:389
31. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16:258
32. Reetz MT, Bocola M, Carballeira JD, Zha DX, Vogel A (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew Chem Int Ed* 44:4192

33. Reetz MT, Wang LW, Bocola M (2006) Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed* 45:1236
34. Reetz MT, Sanchis J (2008) Constructing and analyzing the fitness landscape of an experimental evolutionary process. *Chem Bio Chem* 9:2260
35. Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew Chem Int Ed* 45:7745
36. Reetz MT, Soni P, Fernandez L, Gumulya Y, Carballeira JD (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
37. Reetz MT (2013) The Importance of additive and non-additive mutational effects in protein engineering. *Angew Chem Int Ed*. 52:2658. doi:[10.1002/anie.201207842](https://doi.org/10.1002/anie.201207842)
38. Acevedo-Rocha CG, Kille S, Reetz MT (2013) Iterative saturation mutagenesis: a powerful approach to engineer proteins by simulating Darwinian evolution. *Methods Mol Biol* (submitted)
39. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S, Grate J, Gruber J, Whitman JC, Sheldon RA, Huisman GW (2007) Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotechnol* 25:338
40. Kuipers RKP, Joosten HJ, van Berkel WJ, Leferink NG, Rooijen E, Ittmann E, van Zimmeren F, Jochens H, Bornscheuer U, Vriend G, dos Santos VA, Schaap PJ (2010) 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* 78:2101
41. Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Bottcher D, Joosten HJ, Bornscheuer UT (2010) The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *Chem Bio Chem* 11:1635
42. Jochens H, Bornscheuer UT (2010) Natural diversity to guide focused directed evolution. *Chem Bio Chem* 11:1861
43. Jochens H, Aerts D, Bornscheuer UT (2010) Thermostabilization of an esterase by alignment-guided focussed directed evolution. *Prot Eng Des Sel* 23:903
44. MacBeath G, Kast P, Hilvert D (1998) Redesigning enzyme topology by directed evolution. *Science* 279:1958
45. Juergens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R (2000) Directed evolution of a (beta-alpha) 8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Nat Acad Sci U S A* 97:9925
46. Crameri A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 391:288
47. Aharoni A, Griffiths AD, Tawfik DS (2005) High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol* 9:210
48. Becker S, Hobenreich H, Vogel A, Knorr J, Wilhelm S, Rosenau F, Jaeger KE, Reetz MT, Kolmar H (2008) Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew Chem Int Ed* 47:5085
49. Fernández-Álvarez E, Snajdrova R, Jochens H, Davids T, Böttcher D, Bornscheuer UT (2011) A combination of in vivo selection and cell sorting for the identification of enantioselective biocatalysts. *Angew Chem Int Ed* 50:8584
50. Reymond JL, Wahler D (2002) Substrate arrays as enzyme fingerprinting tools. *Chem Bio Chem* 3:701
51. Jose J (2006) Autodisplay: efficient bacterial surface display of recombinant proteins. *Appl Microbiol Biotechnol* 69:607
52. Kondo A, Ueda M (2004) Yeast cell-surface display—applications of molecular display. *Appl Microbiol Biotechnol* 64:28
53. Becker S, Schmoldt HU, Adams TM, Wilhelm S, Kolmar H (2004) Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr Opin Biotechnol* 15:323

54. Becker S, Hobenreich H, Vogel A, Knorr J, Wilhelm S, Rosenau F, Jaeger KE, Reetz MT, Kolmar H (2008) Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes. *Angew Chem Int Ed* 47:5085
55. Becker S, Michalczyk A, Wilhelm S, Jaeger KE, Kolmar H (2007) Ultrahigh-throughput screening to identify *E. coli* cells expressing functionally active enzymes on their surface. *Chem Bio Chem* 8:943
56. Shimizu Y, Kanamori T, Ueda T (2005) Protein synthesis by pure translation systems. *Methods* 36:299
57. Griffiths AD, Tawfik DS (2006) Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* 24:395
58. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Nat Acad Sci U S A* 107:4004
59. Fallah-Araghi A, Baret JC, Ryckelynck M, Griffiths AD (2012) A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip* 12:882



<http://www.springer.com/978-3-642-41520-3>

Fundamentals and Application of New Bioproduction
Systems

Zeng, A.-P. (Ed.)

2013, X, 238 p., Hardcover

ISBN: 978-3-642-41520-3