

A roadmap to directed enzyme evolution and screening systems for biotechnological applications

Ronny Martínez^{1*} and Ulrich Schwaneberg¹

¹ RWTH Aachen University, Worringerweg 3, D-52074 Aachen, Germany

ABSTRACT

Enzymes have been long used in man-made biochemical processes, from brewing and fermentation to current industrial production of fine chemicals. The ever-growing demand for enzymes in increasingly specific applications requires tailoring naturally occurring enzymes to the non-natural conditions found in industrial processes. Relationships between enzyme sequence, structure and activity are far from understood, thus hindering the capacity to design tailored biocatalysts. In the field of protein engineering, directed enzyme evolution is a powerful algorithm to generate and identify novel and improved enzymes through iterative rounds of mutagenesis and screening applying a specific evolutive pressure. In practice, critical checkpoints in directed evolution are: selection of the starting point, generation of the mutant library, development of the screening assay and analysis of the output of the screening campaign. Each step in directed evolution can be performed using conceptually and technically different approaches, all having inherent advantages and challenges. In this article, we present and discuss in a general overview, challenges of designing and performing a directed enzyme evolution campaign, current advances in methods, as well as highlighting some examples of its applications in industrially relevant enzymes.

Key words: Directed evolution, enzyme engineering, high throughput screening, random mutagenesis.

INTRODUCTION

Proteins are the functional and structural units in biological systems. Among proteins, enzymes have evolved across millions of years to specifically carry out essential functions in living organisms. The use of enzymes as biocatalysts was initially tied to the discovery of enzymes or organisms able to perform a reaction of interest in nature such as in brewing or fermentation processes. However, as immense as natural diversity is, the reaction conditions, reaction chemistry, substrate and product concentrations in which industrial processes are conceived often differ from physiological conditions, resulting in reduced enzymatic efficiency, stability or even complete loss of activity.

Generating and identifying enzymes adapted to non-natural (non-biological) conditions and achieving functional or economically efficient catalytic rates in such conditions is a particularly difficult task; despite being able to successfully modify protein sequence and change enzymatic properties, our understanding of the molecular mechanisms which will allow predicting amino acid sequence changes that would result in a specific behavior remains incomplete.

Successfully engineered enzymes for non-natural conditions enabled biocatalysts to be efficiently used in industrial processes which were traditionally performed chemically. Retroactively, in the past decade, the increasing demand of biocatalyst as replacement for chemical processes has progressively fueled the need for novel and improved enzymes applied to product-driven biocatalysis fields such as the synthesis of pharmaceutical precursors and the production fine chemicals.

Protein engineering strategies comprise rational and semi-rational approaches, directed enzyme evolution and recently the “data-driven” enzyme design (Bommarius et al., 2011).

This review aims to present and discuss the advantages and challenges of designing and performing a directed enzyme evolution campaign, current advances in methods, as well as highlighting some examples of its applications in industrially relevant enzymes.

Enzyme engineering strategies seek to generate suitable enzymes having a higher performance compared to the starting variant. Rational design aims to modify specific amino acids which are known or expected to be directly involved in the activity, stability or substrate/product specificity of the target enzyme, commonly generating single variants of small focused libraries (<100 variants) which can be rapidly screened. Such identification of critical amino acid residues strongly relies on the available information, in the way of structures (X-ray, NMR), catalytic mechanisms, or number of sequences with high identity. In the case of well-known enzymes, rational design allows to focus the mutant library diversity on a few amino acids, thus generating very small libraries. On the other hand, when working with recently isolated enzymes having limited or no available information, or those with few described homologs, assigning amino acids to a specific function becomes a high risk bet. Increasing the confidence of this (educated) guess often means generating tenths of single-saturation mutagenesis libraries, or combinatorial libraries in which theoretical diversity exponentially grows up to millions of variants.

Directed enzyme evolution (Farinas et al., 2001) represents a highly versatile approach for tailoring enzyme properties to the needs in industrial applications through iterative cycles of gene diversity generation and high throughput screening, providing, in addition, valuable insights in structure-function relationships in recently isolated and less studied enzymes (Tee and Schwaneberg, 2007, Wong et al., 2006).

*Corresponding author: Dr. Ronny Martínez, RWTH Aachen University, Lehrstuhl für Biotechnologie, Worringerweg 3, 52074 Aachen, Tel.: +49 241 80 23639, Fax: +49 241 80 22387
E-mail: r.martinez@biotec.rwth-aachen.de

A classic directed evolution experiment comprises three main iterative steps: Design and construction of a mutant library from a parent gene, screening to identify improved variants out of a large mutant pool, and isolating the genes encoding for improved variants (Fig. 1).

Despite the directed enzyme evolution concept seems simple and straightforward, each of its steps is a key process, involving decisions that will ultimately affect the output of the directed evolution campaign. The following sections discuss checkpoints of a directed enzyme evolution campaign, their role in the whole process and the current and novel strategies available for each step.

CHECKPOINTS IN A DIRECTED EVOLUTION EXPERIMENT

Starting point

Enzymes can be evolved towards improving an existing activity, conversion of novel substrates (substrate promiscuity) or performing novel reactions (chemical promiscuity) under non-natural conditions. Enzymes catalyzing a specific chemical reaction can be identified from extensive databases, such as BRENDA (Schomburg et al., 2013). Many naturally occurring enzymes are promiscuous (O'Brien and Herschlag, 1999). When aiming to improve an existing activity, special consideration has to be taken on how much nature has worked on that specific reaction under the natural selection conditions. In the case of the main catalytic activity, a dramatic improvement should not be expected. Higher improvement rates can be expected when expanding side activities, or when laboratory evolution focusses in non-natural aspects of the final application of the target enzyme. Therefore, the optimization of enzymes for novel substrates is more likely to

succeed if the parents already show low catalytic rates towards a target substrate or a closely related molecule (Leemhuis et al., 2009, O'Brien and Herschlag, 1999).

On the other hand, physicochemical properties of the parent, such as thermal and solvent stability, are considered to have an effect on the "evolvability" of the enzyme. A more stable parent is able to withstand a greater number of amino acid substitutions than a less stable homolog. This is not trivial, since catalytically beneficial amino acid substitutions can usually reduce enzyme stability as a side-effect (Bloom et al., 2006, Martinez et al., 2013). The ability of a parent enzyme to accept multiple destabilizing amino acid substitutions increases the probability of identifying beneficial substitutions. Enzymes isolated from thermophilic organism are therefore often a first choice as parents for directed evolution. However, it is reported that thermophilic enzymes show a lower activity at lower reaction temperatures (Demirjian et al., 2001, Kumar and Nussinov, 2001). A suitable combination of activity and stability of the selected parent is a main parameter to judge on an appropriate starting point for a directed evolution campaign.

Additionally, multiple homologous enzymes partially fulfilling only activity or stability requirements can be recombined using gene recombination methods, such as DNA shuffling (Coco et al., 2001), StEP (Zhao et al., 1998), SCOPE (O'Maille et al., 2002) or PTRec (Marienhagen et al., 2012) in order to generate, from that limited diversity, a hybrid molecule with both appropriate activity and high stability. An overview on the developments in diversity generation can be found in a recent review (Ruff et al., 2013).

Library construction

Rather than introducing genetic diversity by accumulation of mutations and recombination over generations, in a directed evolution experiment genetic diversity is generated by introducing random mutations in the parent gene in a single event over multiple rounds, yielding a population of mutant variants that can be screened for a desired trait (Fig. 1).

Methods to generate mutant libraries range from codon focused to random unbiased mutagenesis, and every approach has advantages and drawbacks.

A straightforward brute-force random mutagenesis approach would consist in generating a mutant library covering the whole diversity available within the protein sequence space of the target protein, and subsequent screening for improved variants. Brought to numbers, the theoretical sequence space of unique amino acid peptides with 10 amino acids is $1 \times 10^{10} = 1.024 \times 10^{13}$ (20 amino acid building blocks per position) Even at the impressive screening rate of one million variants per hour –seven days a week– the first year of screening would only cover less than 0.1% percent of the theoretical sequence space. Given that enzymes range from 180 to 600 amino acids, strategies and methods for understanding how sequence randomization can be grasped are required, in order to generate sequence diversity enriched in variants having a higher chance of improvement for a given trait (Shivange et al., 2009).

Understanding how generated sequence diversity translates into functional diversity goes hand to hand with understanding protein properties that ought to be improved. In the case of localized enzyme properties, such

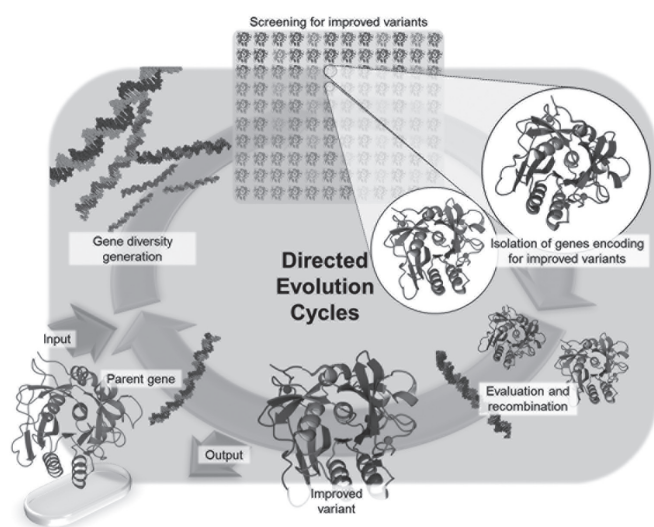


Figure 1. Schematic representation of a directed enzyme evolution campaign. A typical directed evolution campaign comprises iterative rounds in which a parent gene coding for an enzyme of interest is subjected to random mutagenesis, generating a mutant library. Enzyme variants are expressed and screened for a given property. Improved variants are isolated, confirmed and mutations found are evaluated and recombined generating an overall improved variant which can be used as an input for the next round.

as substrate binding (enantio-selectivity, substrate specificity or regio-selectivity), a focused mutant library approach can be performed; when a 3-dimensional model or a crystal structure of the enzyme is available. Successful attempts are reported in which the selection of one or multiple amino acid substitutions were sufficient to alter enzyme enantio-selectivity and substrate specificity (Jakoblinnert et al., 2013, Jakoblinnert et al., 2013, Reetz, 2008, Reetz, 2009, Reetz et al., 2005, Wong et al., 2007). Through single or simultaneous site-saturation mutagenesis (SSM), specific codons are randomized using degenerate codons (NNN; N = A, T, C or G) yielding all 64 possible codons in a single SSM (Table I). Considering an oversampling factor of 3 for a ~95 percent chance of full coverage (Reetz et al., 2008), screening 192 randomly selected variants would be necessary to cover the diversity generated. Simultaneous SSM generates a diversity increment in which saturating two simultaneous NNN codons means screening 12,288 variants (64 x 64 codons), whereas three simultaneous codons saturation demands screening 786,432 variants (Table I). This diversity explosion is due to the genetic redundancy of the NNN codon. By employing NNK codons (K = guanine or thymine) representing 32 possible codons, still coding for all 20 amino acids, the screening demands of single, double and triple SSM libraries are reduced to 96, 3,072 and 98,304, respectively (Table I). The ability to control the codon alphabet can be used to construct SSM libraries enriched in specific amino acids to achieve a double purpose: to reduce the generated diversity, thus the screening effort and to focus the chemical properties of the introduced amino acids according to a predicted effect in the protein. The degenerate codon NDT (N = A, T, C or G; D = A, G or T) has been proposed to further decrease the chemical redundancy since it reduces the diversity to 12 amino acids (Gly, Phe, Ile, Leu, Val, Tyr, His, Cys, Ser, Asn, Asp, Arg), representing a balanced sample of aromatic, aliphatic, non-polar, polar, negative and positively charged residues. The effectiveness of a NDT library was shown to be 13-fold higher compared to NNK after screening 5000 clones (Reetz et al., 2008), supporting the idea that a genetically and chemically non-redundant codon randomization approach will greatly decrease screening efforts

without hindering –given the reduced amino acid diversity– the chance to find improved variants.

Following this approach, the type and amount introduced amino acids in SSM libraries can nowadays be controlled by selecting degenerate codons for virtually any subset of amino acid residues (Firth and Patrick, 2005, Mena and Daugherty, 2005), which combined with the computational prediction of important amino acids positions and catalytic mechanisms form the basis for the construction of “smart” libraries.

There are, however, enzyme properties which are not understood, such as solvent related properties (solubility, co-solvent resistance), salt and surfactant effects, allosteric inhibition and temperature activity and stability, where focused libraries cannot be applied. Suggestions based in collections of mutational and activity data can be obtained using enzyme structure-function relationship driven databases and servers such as hotspot wizard (Pavelka et al., 2009), MAP (Wong et al., 2006) and MAP 3D (Verma et al., 2012). Random mutagenesis libraries offer the possibility to improve enzymes in virtually any property that can be reflected in a screening assay, offering in addition to improved variants, novel structure-function relationship data for the target enzyme.

Random mutagenesis libraries are commonly generated using PCR-based methods differing mainly in how they tackle the mutational bias of DNA polymerases, the organization of the genetic code, and the limitation of single nucleotide exchanges.

Commonly used DNA polymerases (Taq, Pfu, Mutazyme, Φ 29) show a mutational bias towards transitions (purine to purine and pyrimidine to pyrimidine; nucleotide substitutions: A \leftrightarrow G, T \leftrightarrow C) over transversions (purine to pyrimidine or pyrimidine to purine; amino acid substitutions: A \leftrightarrow T, A \leftrightarrow C, T \leftrightarrow G, G \leftrightarrow C) (Wong et al., 2006). In the context of the genetic code, transition mutations result in amino acids chemically similar or identical to the original residue (Wong et al., 2007), causing a limited chemical diversity in the constructed library.

Another challenge presented by organization of the genetic code is the uneven representation of the 20 amino acids, ranging from six codons per amino acids in the case of Leu, Ser and Arg, to only one codon in the case of Trp and Met.

TABLE I

Codon diversity generated by site-saturation mutagenesis (SSM) using different degenerate codons

Generated codon diversity for NNN (64 codons, 20 amino acids), NNK (32 codons, 20 amino acids) and NDT degenerate codons (12 codons, 12 amino acids). An oversampling of 3 is necessary for sampling the whole diversity with a ~95% confidence (Reetz et al., 2008). The diversity explosion is dramatically reduced when using NDT codon for simultaneous SSM, however only 12 amino acids are represented (Gly, Phe, Ile, Leu, Val, Tyr, His, Cys, Ser, Asn, Asp, Arg)

Simultaneously saturated codons	Type of degenerate codons used in site-saturation mutagenesis with number different and unique combinations on the DNA level					
	NNN		NNK		NDT	
	Diversity	Clones for ~95% coverage	Diversity	Clones for ~95% coverage	Diversity	Clones for ~95% coverage
1	64	192	32	96	12	36
2	4096	12288	1024	3072	144	432
3	262,144	786,432	32,768	98,304	1728	5184
4	1.68 x 10 ⁷	5.03 x 10 ⁷	1.05 x 10 ⁶	3.15 x 10 ⁶	20,736	62,208
5	1.07 x 10 ⁹	3.22 x 10 ⁹	3.36 x 10 ⁷	1.01 x 10 ⁸	248,832	746,496

N = adenine [A], thymine [T], cytosine [C] or guanine [G]

K = guanine [G] or thymine [T]

D = adenine [A], guanine [G] or thymine [T]

In practice, this means that random libraries have a naturally biased amino acid distribution, thus further limiting their protein diversity. Furthermore, codon diversity in PCR-based mutagenesis methods is introduced a single nucleotide at a time, given the inability of DNA polymerases to elongate multiple nucleotide mismatches (Wong et al., 2007).

All abovementioned challenges are conveniently integrated in the genetic code most likely as a protection mechanism of the biological system to hamper the accumulation of mutations and drastic protein modifications. Codon redundancy, DNA polymerases bias towards transitions and their inability to introduce consecutive mutations ensure a very small but necessary genetic variability, in the way of silent mutations of substitutions with chemically similar amino acids.

Error-prone PCR (epPCR) is the workhorse of random library generating methods, relying in the introduction of mutations in the template gene by performing a PCR reaction in error-inducing conditions such as in presence of Mn^{2+} , unbalanced nucleotide concentrations or excessive PCR cycles.

Error-prone PCR libraries, despite being of straightforward construction, are affected by diversity challenges presented by the organization of the genetic code discussed above and the introduced mutations are transition biased, resulting in chemically similar amino acid substitutions. As a result, in average, less than 40% of all possible amino acid substitutions are possible using epPCR methods (Wong et al., 2006).

Diverse library generation methods, developed to overcome mutational biases, have been presented and discussed in detail, offering a more complete diversity at amino acid level (Shivange et al., 2009). The addition of nucleotide analogues to bypass mutational bias, Random Insertion and Deletion (RID) mutagenesis, Random Insertional-deletional Strand Exchange (RAISE) mutagenesis and Sequence Saturation Mutagenesis (SeSaM) offer libraries with enhanced diversity (Wong et al., 2007). In particular, SeSaM (Wong et al., 2004) and its transversion enriched version SeSaM-Tv (Mundhada et al., 2011, Wong et al., 2008) offer the additional benefits of generating transversion enriched libraries and the ability of incorporate subsequent mutations in a codon, which increases the obtained codon diversity from a 39.5% (single mutation in a codon) to 83% coverage of the natural diversity (Shivange et al., 2009).

The main difference between biased or unbiased random mutagenesis libraries lays in the probability of generating more complete (or less incomplete) representation of the actual diversity that can be generated from the parent gene. However, given the immense theoretical diversity present in a random mutagenesis library and the possible screening limitations, researchers should not expect to identify “the best” variant from screening a random mutagenesis library. Instead, significant enzyme improvement can be generated using random mutagenesis by performing several subsequent rounds of screening in a directed evolution campaign, where one or more improved enzyme variants identified from screened are used as parents for the next round of random mutagenesis.

Library screening

A reliable enzyme production system is a pre-requisite for a suitable high throughput screening system. Although enzymes are isolated from numerous organisms, directed evolution

is performed only in a few different host organisms with *Escherichia coli*, *Bacillus subtilis* and *Saccharomyces cerevisiae* being the standard hosts, based on their high transformation efficiencies and well established genetic manipulation tools. Further considerations on selecting an appropriate host in directed evolution have been discussed, describing bacterial, yeast, insect and mammalian cells as feasible hosts for directed enzyme evolution (Pourmir and Johannes, 2012). Important is that enzyme production is sufficient for activity and is homogenous across the host population, which can be statistically assessed by measuring enzymatic activity from multiple initial cultures of the selected host.

Discussions regarding theoretical diversity in random mutagenesis methods lose relevance if there isn't a screening or selection platform that can sample a meaningful percentage of the generated library.

High throughput assay development usually deals with a balance between substrate selection, assay complexity, and detection limits. Protein engineers are often presented with the challenge of not being able to use the “real” substrate due to availability, economic or complexity reasons. Using a model substrate for a library screening campaign, though initially more convenient, might lead to optimization of the enzyme towards the model substrate, in detriment of the “real” substrate.

Assay complexity has usually an inverse relationship with throughput, being selection systems, relying on the link between cell growth and enzyme functionality, those providing the highest throughput/complexity relationship and reducing the need for special instruments (Fig. 2A). Selection screening follows library transformation into the host, allowing screening large libraries, by growing transformed cells in a selective medium lacking essential nutrients or containing toxic compounds. Selection assays are therefore mostly limited for detoxifying enzymes or enzymes synthesizing compounds necessary for cell growth and survival. Selection assays can be designed to screen for enantio-selective enzymes (Boersma et al., 2008, Reetz et al., 2008), where the desired enantiopure substrate releases essential nutrients upon reaction, while the undesired enantiomer was supplied as an inhibitor or poison releasing substrate, allowing the growth only of the variants having the desired activity.

Following in the complexity scale and throughput are agar plate assays, in which substrate conversion around single colonies generates a visual signal directly or indirectly in the agar medium (Fig. 2A). Chromogenic or fluorogenic substrate conversion can be observed by change on colonies color, fluorescence, or generation of halos around the active colonies, allowing identifying active variants among those growing on the agar plate. Agar plate screening methods are, however, not suitable for quantifying catalytic activities of individual variants in the library, making them ideal as pre-screening methods.

Microtiter plate (MTP) is the most used screening format for library screening, mainly because it performs as a miniature cuvette system in which the optimal ratio between assay complexity and throughput is achieved by current mainstream technology (Fig 2B). A large diversity of enzymatic assays has been ported to the MTP format and an equal amount has been developed specifically for the format. Single colonies isolated from a mutant library are inoculated

and grown in 96-well plates (master plate), replicated into an expression plate where enzyme is produced and extracted by removing of the biomass (extracellular enzymes) or by cell disruption (intracellular enzymes), and subsequently the assay is performed in 96- or 384- well format. The main advantage of MTP screening systems is the wide availability of analytical tools and standardized equipment, enabling quantitative activity measurements of each screened variant thus providing a more complete dataset for hit evaluation compared to all other high throughput screening methods. On the other hand, MTP assays have a throughput of 10^3 - 10^5 variants per round, falling nowadays, for 96-well plates, to the medium throughput category.

Further miniaturization of the system leads to the use single host cells as femto-liter scale reaction compartments and individual assessment using flow cytometry (Fig 2C). A key requirement in this approach is to consistently keep the link between the genotype and the phenotype of the cell, meaning

that in the context of the reaction, active enzyme variants should label exclusively the cell in which they are produced. This is achieved by the accumulation of a fluorescent product in the cell, enabling the identification and recovery of the active variants. A whole cell cytochrome P450 monooxygenase assay for *o*dealkylation was reported in which the fluorogenic 7-benzyloxy-3-carboxy-coumarin Ethyl ester fluorogenic substrate is converted and retained in the *E. coli* cell due to the spontaneous cleavage of the ethyl ester group to the corresponding carboxylic acid. An epPCR library of cytochrome P450 BM3 monooxygenase was analyzed at a rate of 3.6×10^7 events per day using flow cytometry, active cells were sorted in agar plates, and the active variant enriched re analyzed and screened by flow cytometry. After a three rounds of enrichment, variants with up to 7-fold increased activity were identified, validating the direct cell labeling approach (Ruff et al., 2012). Success in this approach relies heavily in the ability of the

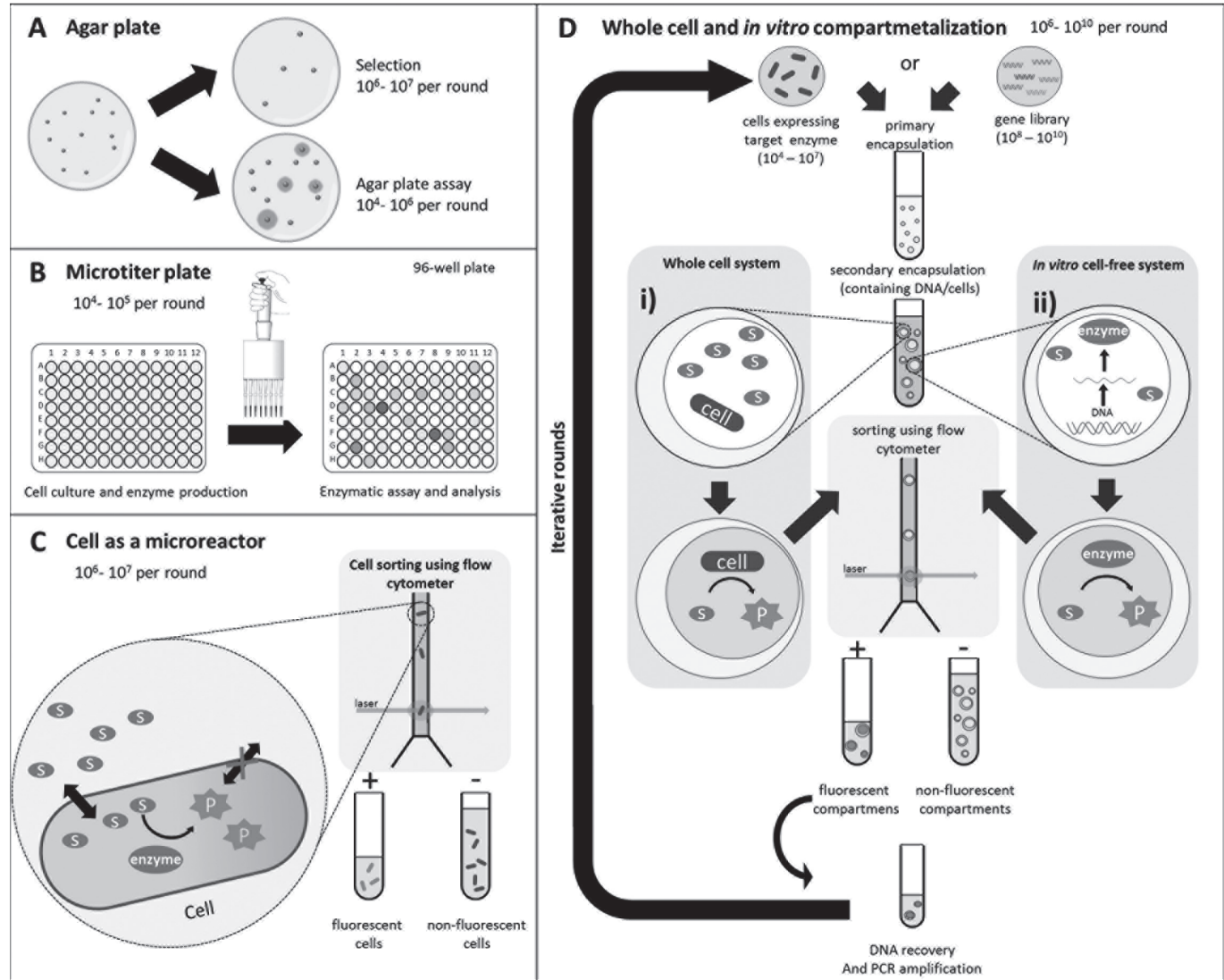


Figure 2. Summary of library screening approaches for directed enzyme evolution. A) Agar plate growth-based selection and enzymatic assay screening. **B)** Microtiter plate screening **C)** Whole-cell based reaction compartment screening. **D)** Whole-cell (i) and *in vitro* (ii) compartmentalization screening approaches.

cell to uptake the substrate and to accumulate the fluorescent product, allowing the identification of active variants.

The product accumulation challenge has been addressed by the development of microencapsulation methods in which cells producing target enzyme are entrapped together with the substrate and product in water/oil/water micro-emulsions (< 70 μm) which are subsequently analyzed at speed of thousands per second by flow cytometry (Griffiths and Tawfik, 2006) (Fig. 2D, i). This approach allows keeping the high throughput analysis of flow cytometry whilst allowing a wider chemical range of substrates and products, which must not necessarily remain inside the cell producing the active variant (Tu et al., 2011).

The microencapsulation approach, combined with flow cytometry analysis, enabled to completely skip the cell host by producing the enzyme directly from the DNA library by *in vitro* protein translation and was termed as *in vitro* compartmentalization (IVC) (Miller et al., 2006, Tawfik and Griffiths, 1998) (Fig 2D, ii). *In vitro* (cell-free) enzyme production allows a simpler biochemical and genetic environment for the selection pressure, enabling for example the production and screening of toxic-proteins, the addition of non-natural cofactors or modified amino acids for labeling the translated protein (Wang et al., 2009). IVC prevents the gene diversity loss inherited from the cloning and transformation steps, by performing protein translation directly from the DNA obtained in the PCR reaction used to generate the mutant library. The use of water-in-oil-in-water emulsions and cell-free protein expression allows reducing experimental time by removing cell growth steps in cloning, transformation and protein production. Despite the conceptual elegance of the IVC screening method and promising potential for directed enzyme evolution, a small number of examples describing routine application of *in vitro* compartmentalization are reported, compared to the technical significance of the method, including directed evolution of a methyltransferase, a beta-galactosidase, and a thiolactanase (Aharoni et al., 2005, Mastrobattista et al., 2005, Tawfik and Griffiths, 1998). This suggests that the IVC approach is not yet established in most protein engineering laboratories likely due to multiple technical challenges (Lu and Ellington, 2012, Nishikawa et al., 2012). An important challenge of water-in-oil-in-water microencapsulation approaches is the generation of mono-dispersed homogeneous droplets containing a single inner water phase, which must be stable in an aqueous-based shear fluid compatible with pressure fluctuations of flow cytometry devices. Current droplet formation approaches include two-step emulsification by dispersion and extrusion, which can generate a population of droplets around a specific diameter. The droplet size variation, however, can reach more than 100% between the smallest and biggest subpopulations (Mastrobattista et al., 2005).

Microfluidics technology has recently been introduced in the field of directed evolution as an alternative approach to overcome the encapsulation challenge in IVC. Homogeneous single and double emulsion generation have been shown to be a mature technology with broad applications in physics, chemistry and biology (Theberge et al., 2010). Successful approaches on expressing green from single DNA templates fluorescent protein GFP paved the way for the use of microfluidic encapsulation in directed evolution experiments (Dittrich et al., 2005). Besides the encapsulation of mutant libraries and protein translation, microfluidic devices have

the capacity to integrate quantitatively detect single cell enzymatic assays such as the phosphatase driven cleavage of 3-o-methylfluorescein phosphate in the periplasm of encapsulated *E. coli* cells (Huebner et al., 2008) and the indroplet cell lysis-based screening system for detecting activity of an epPCR library of a promiscuous sulfatase towards phosphonate using the fluorogenic bis-(methylphosphonyl)-fluorescein (Kintses et al., 2012). The latter example is especially interesting, since the microfluidic device included droplet fluorescence detection and a dielectrophoresis-based sorting system which allowed recovering droplets showing phosphonate cleaving activity for DNA isolation and transformation into *E. coli* cells for the next round of directed evolution. Very recently, a completely *in vitro* high-throughput microfluidic screening system was reported for specific application in directed enzyme evolution using β -galactosidase as a proof of principle (Fallah-Araghi et al., 2012), allowing detection and sorting of up to 2000 droplets per second.

Screening systems available for directed enzyme evolution are varied both in conceptual and technical complexity. Until today, MTP screening platforms are generally used for protein engineers due to their flexibility and automation. Emerging platforms, though conceptually and technically more advanced in terms of throughput, are yet to reach a routine maturity to replace MTP platforms from their privileged position.

Analysis of the output from a directed evolution round

The identification of improved enzyme variants after library screening is still far from the end of a directed evolution campaign. It is infrequent that a directed evolution round yields a single "better" variant; commonly, a population of 3 to 5 variants showing various levels of increased performance compared to the parent enzyme are identified each screening round. Depending on the mutational load applied in the library, each of the improved variants might contain one or more amino acid substitutions. A common approach is to "clean" selected variants by performing single site-directed mutagenesis (SDM) on the parent gene, thus identifying amino acids substitutions responsible of the observed activity changes. Furthermore, mutational cleaning, in addition to isolating meaningful amino acid substitutions, results in identifying synergistic interactions between amino acid substitutions and also removing neutral and non-beneficial amino acid substitutions which are thought to decrease inherent stability and thus the "evolability" of the enzyme for future rounds (Bloom et al., 2006). Experimental characterization of cleaned variants is also convenient when structure-function relationships are investigated, since changes in activity and stability can be directly attributed to the present amino acid substitutions. A more complex scenario is when a large number of selected variants have multiple amino acid substitutions. A statistical-based approach is reported in which sequence-activity data is analyzed for a population of variants, in order to extract and label the beneficial/detrimental factor of single amino acid substitutions and proposing new combinations expecting increased activity, without the need of constructing the single mutants (Brouk et al., 2010). Using this approach and a broad dataset of sequence-activity obtained from directed evolution and rational design, 2-phenylethanol degradation by a multicomponent toluene 4-monooxygenase

from *Pseudomonas mendocina* was improved up to 7.3 times in average from the initially found improved variants only by recombination of known amino acid substitutions from previous rounds (Brouk et al., 2010). The main advantage of the statistical approach is that, similar to directed evolution, no previous structural information is necessary to propose new combinations, though it may be complemented by it. The interaction between amino acid substitutions, however, is regarded as additive –the sum of their individual effect– thus underestimating synergistic effects between substitutions.

Large populations of amino acid substitutions in different improved variants can be analyzed computationally, when structural information is available, by identifying clusters of amino acid substitutions within the targeted protein structure. The latter allows constructing of variants that contain amino acid substitutions that “can interfere with each other” in order to identify and analyze combinatorial effects in smaller subpopulations. Such an analysis allow to identify the most beneficial combinations in each cluster, and subsequently integrate them a single variant (Bocola et al., 2004).

CURRENT CHALLENGES OF ENZYME EVOLUTION

It is clear that for each step in directed evolution, there are multiple methods and strategies that can –and have been shown to– be successful identifying improved enzymes. Choosing the appropriated approach affects not only in how easy or hard will be to actually generate and identify the desired enzyme variant, but also the information that will be generated along with the process. Nowadays, challenges in directed enzyme evolution can be divided into conceptual challenges and technical challenges.

The ultimate challenge in protein engineering is to understand and therefore predict protein behavior. A main conceptual challenge of directed enzyme evolution is to predict and relate enzyme evolutive behavior in the laboratory with that of natural evolution. Laboratory evolution allows to study and modify enzymes as an isolated entity in the context of the selected pressure to which is subjected during the screening campaign. On the other hand, natural evolution the selection pressure is multifactorial, occurs over an extended period of time, and those selection pressures are variable across different evolutionary periods. It is proposed that nature evolves enzymes through accumulation of neutral or even deleterious mutations and, in contrast to directed evolution, allows the continuity of inactive sequences (Romero and Arnold, 2009), suggesting that every natural enzyme has a certain percentage accumulation of neutral mutations and that inactive sequences can over time recover or generate enzymes with new activities, for example after gene duplication events. Considering that starting points for directed evolution are natural enzymes, there is always an uncertainty on how the neutral mutations will translate when a unique evolutionary pressure is applied to the enzyme. An interesting way to convey the enzyme response to the pressure applied in laboratory evolution is the concept of protein fitness landscape, which is the representation of the sequence space (genotype) of a given protein as (theoretical) single variants and adding a third measure of fitness (phenotype) for each of these variants with respect to one or more applied evolutionary pressures (screening assay) (Carneiro and Hartl, 2010, Romero and Arnold, 2009). The shape of this landscape

ranges from the “Mount Fuji” shape, in which the uphill climb is always connected by increasingly active sequences reaching finally a single optimum sequence to the “rugged badlands” landscape in which many optima are present but separated by low activity of inactive sequences. The former landscape allows through navigating through multiple active sequences and through different evolutive paths, to reach the optimum; the latter landscape, however, presents a nearly impossible way to navigate between the different optima, since directed evolution does not allow inactive sequences to be used in further rounds of directed evolution.

The genetic diversity generation has migrated from a technical challenge to a conceptual challenge. Nowadays, routinely used methods for generating mutant libraries range from single site-saturation mutagenesis using the Quickchange method to the transversion enriched full sequence saturation mutagenesis (SeSaM). This range of possibilities has generated fruitful debate about which is the most effective strategy to achieve rapid and dramatic improvements in directed evolution. The “low mutagenic load” approach, where in average 1 or 2 amino acid substitutions are introduced, generates libraries with a high percentage of active variants where incremental improvements (2-3 fold activity/stability) are obtained, relying in several iterative rounds of evolution to generate a dramatic improvement on the desired property.

On the other hand, a high mutational rate approach generates libraries with a low percentage of active clones, relying heavily on a high throughput screening system to identify and recover active variants for further analysis. A higher mutagenic rate could generate epistatic effects that could result in alternative adaptation pathways which are not reachable by low mutational rate approaches (Salverda et al., 2011).

Random mutagenesis and focused libraries have migrated from an excluding to a complementary relationship; a successful directed evolution campaign will often start by epPCR or SeSaM as an “explorative” screening round in order to identify relevant amino acid positions which can be subsequently saturated and recombined using SDM or SSM to obtain optimized variants and generate structure-function knowledge. Conversely, in semi rational active-site focused campaigns, random mutagenesis libraries can be screened in parallel for increase stability, in order to identify stabilizing amino acid substitutions, which can be introduced in improved variants in order to increase enzyme evolvability in further rounds.

Reducing the timescale of a directed evolution campaign is the main challenge for studying the evolutionary behavior of enzymes. Generally, directed evolution campaigns generate isolated examples of improved variants and hypotheses on the effect of the introduced amino acid substitutions, rarely reporting a long-term evolutionary story. The lineage of a P450 monooxygenase evolved towards propane hydroxylation was studied, revealing how the original P450 BM3 monooxygenase which was able to hydroxylate C₁₂-C₂₀ fatty acids transformed, over 20 amino acid substitutions, into a P450 propane monooxygenase. Key changes in the substrate binding pocket of P450 BM3 allowed the appearance of a wide range of substrate activities, including propane. Further changes on substrate recognition mechanisms enabled a narrower substrate profile around propane, ending in a propane-specific P450_{PMO} variant (Fasan et al., 2008). Reducing the time required for a directed evolution round requires overcoming two major

bottlenecks: i) the growth rate of the host organism, which affects transformation, pre-culture and protein production timescales, and ii) the library screening throughput. Currently, the *in vitro* compartmentalization approach (IVC), due to the application of cell-free enzyme production and ultra-high throughput screening capabilities, seems to be the most promising methodology in which multiple rounds of directed enzyme evolution can be achieved in the matter of weeks or even days. An advanced IVC technology will allow a higher throughput per round of directed evolution, thus increasing the probabilities of identifying significantly improved enzyme variants (through more rounds of directed evolution per campaign) and allowing the study of the evolutionary behavior of enzymes under different selection pressures.

GENERATING AND IMPROVING INDUSTRIAL ENZYMES BY DIRECTED EVOLUTION

Industrial (“white”) biotechnology (Frazzetto, 2003) comprises mainly the production of chemicals, fuels and materials using isolated enzymes (biocatalysis) or enzyme-producing living cells (fermentation). The rapid development of industrial biotechnology in the last years can be attributed to two main driving forces; the increased economic efficiency of the biotechnological production compared to traditional chemical processes, and the current social demands for sustainable, environment-friendly industrial processes (Tang and Zhao, 2009). The global demand for replacing fossil resources by renewable agricultural crops or low-value side products as raw starting materials ensures that the market share of chemical produced by industrial biotechnology (fuels, plastics, fine and bulk chemicals) will continue to grow in the next years (Soetaert and Vandamme, 2006).

Enzymes are important part of industrial biotechnology, the use proteases and lipases for the detergent industry, α -amylases for the production of glucose, the use of phytases for the animal feed industry and the biosynthesis of riboflavin (vitamin B2) and cobalamin (vitamin B12), are just the classical examples of enzymes applied to industrial production. Nowadays, enzymes have taken an important role in the production of fine chemicals, pharmaceuticals, food additives and supplements, colorants, vitamins, pesticides, bio-plastics, solvents, bio-plastics, bulk chemicals and biofuels (Kumar and Singh, 2012, Soetaert and Vandamme, 2006).

The use of enzymes, however, is limited when their economic efficiency, stability or substrate/product specificity do not meet the demands of the chemical processes in an industrial scale. Protein engineering in a rational or evolutive or combined manner is currently the preferred approach for tailoring biocatalysts for specific needs in products or production. As an example, we discuss two directed enzyme evolution campaigns in which hydrolases were tailored for specific industrial applications.

Adaptation of biomass degrading enzymes for non-conventional solvents

During the last decade, chemical industries are evaluating biomass –based alternatives for bio– and chemical manufacturing as a way to secure energy production (Grande and Domínguez de María, 2012). In that mindset, cellulose depolymerization has become a key step in biorefineries to produce fermentable sugars. Cellulases, together with

hemicellulases, cellobioases, endo- and exoglucanases, are able to perform such depolymerization under mild conditions. Biomass pretreatment, nevertheless, includes several mechanical and chemical steps to decrease cellulose crystallinity, increase its availability for dissolution. Ionic liquids (IL) and more recently deep eutectic solvents (DES) have been proposed as “green” solvents for cellulose pretreatment (Domínguez de María and Maugeri, 2011); however, high IL or DES content usually leads to rapid enzyme inactivation, hindering the subsequent depolymerization steps. A cellulase, CelA2, isolated from a metagenomic library constructed from a biogas plant (Ilmberger et al., 2012) was subjected to directed evolution in order to increase its resistance to choline chloride: glycerol (ChCl:Gly) as a co-solvent. In the first round, an CelA2 epPCR library was generated and screened using a developed 4-methylumbelliferyl- β -D-cellobioside fluorescent MTP assay, yielding a variant, 4D1, harboring 6 amino acid substitutions (Leu21Pro; Leu184Gln; His288Arg; Lys299Ile; Asp330Gly; Asn442Asp) with an increased specific activity in 30% (v/v) ChCl:Gly. In the second evolution round, 6 single SSM libraries were constructed, in order to identify which amino acid substitutions were responsible for the activity change. Screening result revealed that only variants from the library at position 288 showed significant increase in activity, identifying a substitution that increased the resistance to ChCl:Gly (His288Arg) showing a similar behavior as 4D1, and a substitution that resulted in a more than 8-fold increase in specific activity compared to the original CelA2 cellulase, with a similar inactivation rate in ChCl:Gly. (Lehmann et al., 2012). This cellulase evolution study is an interesting example how mutational cleaning approach can work to identify by single SSM, among the six amino acid substitutions present in variant 4D1, that the His288Arg substitution is responsible for increased co-solvent resistance, and further identifying an additional substitution (His288Phe) that dramatically increased specific activity (Fig. 3 B).

Optimizing and developing subtilisin proteases for detergent industry

Alkaline proteases from *Bacillus* species are one of the first success examples of the use of bacterial enzymes in end-products. Since the 1960’s subtilisin proteases have been applied for the detergent industry as active additives for the degradation of protein stains, and in the beginning of the 1990’s the first engineered enzymes entered the laundry market, establishing themselves as industrial benchmarks (Maurer, 2004). The application of subtilisin proteases in detergents is still challenged by market changes and the type of product in which they are applied such as improving proteolytic performance at low temperatures for energy efficient “cold” washing cycles, and storage stability in liquid detergents. Subtilisins are produced as extracellular enzymes, which greatly simplifies protease separation from cell biomass and facilitates downstream purification procedures (Gupta et al., 2002). In order to generate stable subtilisins able to perform in low temperatures, the activity at 15°C of a *Bacillus gibsonii* alkaline protease (BgAP) was increased through 3 rounds of SeSaM-based directed evolution. Libraries were screened in parallel towards increased activity and towards increased thermal stability by measuring proteolytic activity at 15°C and after incubation at 58°C. Directed BgAP evolution yielded a

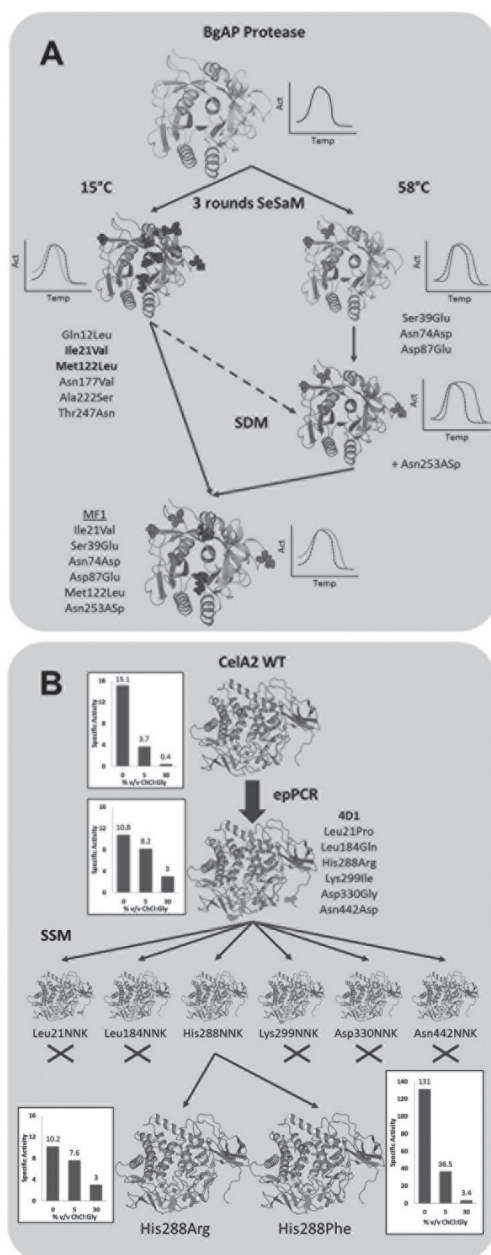


Figure 3. Enzyme evolution strategies applied to two industry relevant enzymes. A) Directed evolution of *Bacillus subtilis* alkaline protease (BgAP) using a parallel screening towards high activity at low temperatures (left) and thermal resistance (right) with 3 rounds of sequence saturation mutagenesis (SeSaM). Two single variants with improved temperature-activity profiles were identified and identified amino acid substitutions were combined to generate a single enzyme variant with a wider temperature profile compared to the wild type (Martinez et al., 2013). B) Directed evolution of a metagenome isolated cellulase (CelA2) towards increased performance in aqueous solutions of the deep eutectic solvent choline chloride with glycerol (ChCl:Gly). After screening an error-prone PCR library, variant 4D1 was identified that harbored 6 amino acid substitutions. In order to identify single amino acid substitutions, 6 site-saturation mutagenesis (SSM) libraries were constructed and screened, revealing that substitution His288Arg was responsible of the observed improvement in 4D1. Additionally substitution His288Phe was identified, which results in an 8-fold increased specific activity, compared to wild type (Lehmann et al., 2012).

set of BgAP variants with increased specific activity at 15°C and increased thermal resistance. Recombination of both sets of amino acid substitutions resulted finally in the variant MF1 with a 1.5-fold increased specific activity and an over 100 times prolonged half-life at 60°C (224 min compared to 2 min of the WT BgAP) (Martinez et al., 2013). Through iterative combination of these amino acid subsets it was shown that improved thermal resistance (requiring strong interactions) and improved activity (often requiring flexibility) can be obtained within a single protease variant harboring six amino acid substitutions.

Using a similar strategy and a novel fluorescent microtiter plate screening assay for peroxy acid (R-COOOH) detection based in the hypobromite induced *o*-dearylation of 7-(4'-aminophenoxy)-3-carboxy coumarin. The screening system was validated by improving the perhydrolytic activity of a subtilisin Carlsberg variant for methyl-butyrate (Despotovic et al., 2012). Peroxy acids are typical bleaching agents (Maurer, 2004) and their enzymatic *in situ* production in the cleaning process would be desirable. Oxidative resistance of the peroxy acid producing subtilisin Carlsberg variant was improved by site saturation mutagenesis in positions Trp216 and Met221 and further recombination. Variant M4 (W216M, M221) had an increase of ~5 fold in the peroxyacetic acid concentration necessary to decrease the enzyme activity by 50% (3.1 mM for M4 versus 0.6 mM for the parent) (Vojcic et al., 2013).

CLOSING REMARKS AND OUTLOOK

Each individual directed enzyme evolution campaign is a challenge, in which the best possible enzyme for a specific trait should be generated and identified using the quickest and most promising strategy. Selecting the appropriate evolution strategy implies considering the expected improvement from the selected parent, the conceptual and technical limitations of library generation and screening systems, and how to evaluate the output of each evolution round, in order to identify and select amino acid substitutions to be carried over to the next round.

Directed evolution is nowadays a general and often validated approach to generate tailored enzymes for specific applications providing additionally valuable information regarding how enzymes adapt and hints in how selection under simple evolutionary pressure can be extrapolated to natural evolution.

The future of directed enzyme evolution as a research field seems quite promising; the increasing demand of engineered enzymes, especially for industrial applications, will continue to fuel research and development of faster, better, more efficient mutagenesis and screening methods for tailoring biocatalysts.

Due to rapid advances in diversity generation and screening technologies, it is not difficult to envision a near future in which protein engineering –and especially directed protein evolution– shares a primary role in developing biocatalysts for industrial and end-user processes, leading the way in understanding of protein adaptation and evolutionary behavior.

ACKNOWLEDGEMENTS

The authors thank Dr. Marco Bocla for critical discussion and editing role for this article.

REFERENCES

- AHARONI, A., G. AMITAI, K. BERNATH, S. MAGDASSI and D. S. TAWFIK (2005). High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. *Chem Biol.* 12, 1281-1289
- BLOOM, J. D., S. T. LABTHAVIKUL, C. R. OTEY and F. H. ARNOLD (2006). Protein stability promotes evolvability. *PNAS.* 103, 5869-5874
- BOCOLA, M., N. OTTE, K.-E. JAEGER, M. T. REETZ and W. THIEL (2004). Learning from Directed Evolution: Theoretical Investigations into Cooperative Mutations in Lipase Enantioselectivity. *ChemBioChem.* 5, 214-223
- BOERSMA, Y. L., M. J. DRÖGE, A. M. VAN DER SLOOT, T. PIJNING, R. H. COOL, B. W. DIJKSTRA and W. J. QUAX (2008). A Novel Genetic Selection System for Improved Enantioselectivity of *Bacillus subtilis* Lipase A. *ChemBioChem.* 9, 1110-1115
- BOMMARIUS, A. S., J. K. BLUM and M. J. ABRAHAMSON (2011). Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* 15, 194-200
- BROUK, M., Y. NOV and A. FISHMAN (2010). Improving Biocatalyst Performance by Integrating Statistical Methods into Protein Engineering. *Appl. Environ. Microbiol.* 76, 6397-6403
- CARNEIRO, M. and D. L. HARTL (2010). Adaptive landscapes and protein evolution. *PNAS.* 107, 1747-1751
- COCO, W. M., W. E. LEVINSON, M. J. CRIST, H. J. HEKTOR, A. DARZINS, P. T. PIENKOS, C. H. SQUIRES and D. J. MONTICELLO (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotech.* 19, 354-359
- DEMIRJIAN, D. C., F. MORIS-VARAS and C. S. CASSIDY (2001). Enzymes from extremophiles. *Curr. Opin. Chem. Biol.* 5, 144-151
- DESPOTOVIC, D., L. VOJCIC, R. PRODANOVIC, R. MARTINEZ, K.-H. MAURER and U. SCHWANEBERG (2012). Fluorescent Assay for Directed Evolution of Perhydrolases. *J. Biomol. Screen.* 17, 796-805
- DITTRICH, P. S., M. JAHNZ and P. SCHWILLE (2005). A New Embedded Process for Compartmentalized Cell-Free Protein Expression and On-line Detection in Microfluidic Devices. *ChemBioChem.* 6, 811-814
- DOMÍNGUEZ DE MARÍA, P. and Z. MAUGERI (2011). Ionic liquids in biotransformations: from proof-of-concept to emerging deep-eutectic-solvents. *Curr. Opin. Chem. Biol.* 15, 220-225
- FALLAH-ARAGHI, A., J.-C. BARET, M. RYCKELYNCK and A. D. GRIFFITHS (2012). A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip.* 12, 882-891
- FARINAS, E. T., T. BULTER and F. H. ARNOLD (2001). Directed enzyme evolution. *Curr. Opin. Biotechnol.* 12, 545-551
- FASAN, R., Y. T. MEHARENNNA, C. D. SNOW, T. L. POULOS and F. H. ARNOLD (2008). Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* 383, 1069-1080
- FIRTH, A. E. and W. M. PATRICK (2005). Statistics of protein library construction. *Bioinformatics.* 21, 3314-3315
- FRAZZETTO, G. (2003). White biotechnology. *EMBO Rep.* 4, 835-837
- GRANDE, P. M. and P. DOMÍNGUEZ DE MARÍA (2012). Enzymatic hydrolysis of microcrystalline cellulose in concentrated seawater. *Bioresour. Technol.* 104, 799-802
- GRIFFITHS, A. D. and D. S. TAWFIK (2006). Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol.* 24, 395-402
- GUPTA, R., Q. BEG, S. KHAN and B. CHAUHAN (2002). An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Appl. Microbiol. Biotechnol.* 60, 381-395
- HUEBNER, A., L. F. OLGUIN, D. BRATTON, G. WHYTE, W. T. S. HUCK, A. J. DE MELLO, J. B. EDEL, C. ABELL and F. HOLLFELDER (2008). Development of Quantitative Cell-Based Enzyme Assays in Microdroplets. *Anal. Chem.* 80, 3890-3896
- ILMBERGER, N., D. MESKE, J. JUERGENSEN, M. SCHULTE, P. BARTHEN, U. RABAUSCH, A. ANGELOV, M. MIENTUS, W. LIEBL, R. SCHMITZ and W. STREIT (2012). Metagenomic cellulases highly tolerant towards the presence of ionic liquids—linking thermostability and halotolerance. *Appl. Microbiol. Biotechnol.* 95, 135-146
- JAKOBLINNERT, A., A. VAN DEN WITTENBOER, A. V. SHIVANGE, M. BOCOLA, L. HEFFELE, M. ANSORGE-SCHUMACHER and U. SCHWANEBERG (2013). Design of an activity and stability improved carbonyl reductase from *Candida parapsilosis*. *J. Biotechnol.*
- SCHWANEBERG (2013). Reengineered carbonyl reductase for reducing methyl-substituted cyclohexanones. *Protein Eng. Des. Sel.* 26, 291-298
- KINTSES, B., C. HEIN, MARK F. MOHAMED, M. FISCHLECHNER, F. COURTOIS, C. LAINÉ and F. HOLLFELDER (2012). Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution. *Chem Biol.* 19, 1001-1009
- KUMAR, A. and S. SINGH (2012). Directed evolution: tailoring biocatalysts for industrial applications. *Crit. Rev. Biotech.* 0, 1-14
- KUMAR, S. and R. NUSSINOV (2001). How do thermophilic proteins deal with heat? *CMLS, Cell. Mol. Life Sci.* 58, 1216-1233
- LEEMHUIS, H., R. M. KELLY and L. DIJKHUIZEN (2009). Directed evolution of enzymes: Library screening strategies. *IUBMB Life.* 61, 222-228
- LEHMANN, C., F. SIBILLA, Z. MAUGERI, W. R. STREIT, P. DOMÍNGUEZ DE MARIA, R. MARTINEZ and U. SCHWANEBERG (2012). Reengineering CelA2 cellulase for hydrolysis in aqueous solutions of deep eutectic solvents and concentrated seawater. *Green Chem.* 14, 2719-2726
- LU, W.-C. and A. D. ELLINGTON (2012). In vitro selection of proteins via emulsion compartments. *Methods.*
- MARIENHAGEN, J., A. DENNIG and U. SCHWANEBERG (2012). Phosphorothioate-based DNA recombination: an enzyme-free method for the combinatorial assembly of multiple DNA fragments. *BioTechniques.* 0, 1-6
- MARTÍNEZ, R., F. JAKOB, R. TU, P. SIEGERT, K.-H. MAURER and U. SCHWANEBERG (2013). Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol. Bioeng.* 110, 711-720
- MASTROBATTISTA, E., V. TALY, E. CHANUDET, P. TREACY, B. T. KELLY and A. D. GRIFFITHS (2005). High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a ²-Galactosidase by Fluorescence-Activated Sorting of Double Emulsions. *Chem Biol.* 12, 1291-1300
- MAURER, K.-H. (2004). Detergent proteases. *Curr. Opin. Biotechnol.* 15, 330-334
- MENA, M. A. and P. S. DAUGHERTY (2005). Automated design of degenerate codon libraries. *Protein Eng. Des. Sel.* 18, 559-561
- MILLER, O. J., K. BERNATH, J. J. AGRESTI, G. AMITAI, B. T. KELLY, E. MASTROBATTISTA, V. TALY, S. MAGDASSI, D. S. TAWFIK and A. D. GRIFFITHS (2006). Directed evolution by in vitro compartmentalization. *Nat Meth.* 3, 561-570
- MUNDHADA, H., J. MARIENHAGEN, A. SCACIOC, A. SCHENK, D. ROCCATANO and U. SCHWANEBERG (2011). SeSaM-Tv-II Generates a Protein Sequence Space that is Unobtainable by epPCR. *ChemBioChem.* 12, 1595-1601
- NISHIKAWA, T., T. SUNAMI, T. MATSUURA and T. YOMO (2012). Directed Evolution of Proteins through In Vitro Protein Synthesis in Liposomes. *J. Nucleic Acids.* 2012, 11
- O'BRIEN, P. J. and D. HERSCHLAG (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol.* 6, R91-R105
- O'MAILLE, P. E., M. BAKHTINA and M.-D. TSAI (2002). Structure-based Combinatorial Protein Engineering (SCOPE). *J. Mol. Biol.* 321, 677-691
- PAVELKA, A., E. CHOVANCOVA and J. DAMBORSKY (2009). HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 37, W376-W383
- POURMIR, A. and T. W. JOHANNES (2012). Directed evolution: selection of the host organism. *CSBJ.* 2,
- REETZ, M. T. (2008). Directed Evolution of Stereoselective Hybrid Catalysts. 1-30
- REETZ, M. T. (2009). Directed Evolution of Enantioselective Enzymes: An Unconventional Approach to Asymmetric Catalysis in Organic Chemistry. *J. Org. Chem.* 74, 5767-5778
- REETZ, M. T., M. BOCOLA, J. D. CARBALLEIRA, D. ZHA and A. VOGEL (2005). Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angew. Chem. Int. Ed. Engl.* 44, 4192-4196
- REETZ, M. T., H. HOBENREICH, P. SONI and L. FERNANDEZ (2008). A genetic selection system for evolving enantioselectivity of enzymes. *Chem. Commun.* 0, 5502-5504
- REETZ, M. T., D. KAHAKEAW and R. LOHMER (2008). Addressing the Numbers Problem in Directed Evolution. *ChemBioChem.* 9, 1797-1804
- ROMERO, P. A. and F. H. ARNOLD (2009). Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol.* 10, 866-876
- RUFF, A. J., A. DENNIG and U. SCHWANEBERG (2013). To get what we aim for – progress in diversity generation methods. *FEBS Journal.* 280, 2961-2978

- RUFF, A. J., A. DENNIG, G. WIRTZ, M. BLANUSA and U. SCHWANEBERG (2012). Flow Cytometer-Based High-Throughput Screening System for Accelerated Directed Evolution of P450 Monooxygenases. *ACS Catal.* 2, 2724-2728
- SALVERDA, M. L. M., E. DELLUS, F. A. GORTER, A. J. M. DEBETS, J. VAN DER OOST, R. F. HOEKSTRA, D. S. TAWFIK and J. A. G. M. DE VISSER (2011). Initial Mutations Direct Alternative Pathways of Protein Evolution. *PLoS Genet.* 7, e1001321
- SCHOMBURG, I., A. CHANG, S. PLACZEK, C. SÖHNEN, M. ROTHER, M. LANG, C. MUNARETTO, S. ULAS, M. STELZER, A. GROTE, M. SCHEER and D. SCHOMBURG (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 41, D764-D772
- SHIVANGE, A. V., J. MARIENHAGEN, H. MUNDHADA, A. SCHENK and U. SCHWANEBERG (2009). Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* 13, 19-25
- SOETAERT, W. and E. VANDAMME (2006). The impact of industrial biotechnology. *Biotech. J.* 1, 756-769
- TANG, W. L. and H. ZHAO (2009). Industrial biotechnology: Tools and applications. *Biotech. J.* 4, 1725-1739
- TAWFIK, D. S. and A. D. GRIFFITHS (1998). Man-made cell-like compartments for molecular evolution. *Nat Biotech.* 16, 652-656
- TEE, K. L. and U. SCHWANEBERG (2007). Directed Evolution of Oxygenases: Screening Systems, Success Stories and Challenges. *Comb. Chem. High Throughput Screening.* 10, 197-217
- THEBERGE, A. B., F. COURTOIS, Y. SCHAEERLI, M. FISCHLECHNER, C. ABELL, F. HOLLFELDER and W. T. S. HUCK (2010). Microdroplets in Microfluidics: An Evolving Platform for Discoveries in Chemistry and Biology. *Angew. Chem. Int. Ed. Engl.* 49, 5846-5868
- TU, R., R. MARTINEZ, R. PRODANOVIC, M. KLEIN and U. SCHWANEBERG (2011). A Flow Cytometry-Based Screening System for Directed Evolution of Proteases. *J. Biomol. Screen.* 16, 285-294
- VERMA, R., U. SCHWANEBERG and D. ROCCATANO (2012). MAP2.03D: A Sequence/Structure Based Server for Protein Engineering. *ACS Synthetic Biology.* 1, 139-150
- VOJCIC, L., D. DESPOTOVIC, K.-H. MAURER, M. ZACHARIAS, M. BOCOLA, R. MARTINEZ and U. SCHWANEBERG (2013). Reengineering of subtilisin Carlsberg for oxidative resistance. *Biological Chemistry.* 394, 79
- WANG, Q., A. R. PARRISH and L. WANG (2009). Expanding the Genetic Code for Biological Studies. *Chem Biol.* 16, 323-336
- WONG, T. S., D. ROCCATANO, D. LOAKES, K. L. TEE, A. SCHENK, B. HAUER and U. SCHWANEBERG (2008). Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): A random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. *Biotech. J.* 3, 74-82
- WONG, T. S., D. ROCCATANO and U. SCHWANEBERG (2007). Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. *Biotech. J.* 2, 133-142
- WONG, T. S., D. ROCCATANO and U. SCHWANEBERG (2007). Challenges of the genetic code for exploring sequence space in directed protein evolution. *Biocatal. Biotransfor.* 25, 229-241
- WONG, T. S., D. ROCCATANO and U. SCHWANEBERG (2007). Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. *Environ. Microbiol.* 9, 2645-2659
- WONG, T. S., D. ROCCATANO, M. ZACHARIAS and U. SCHWANEBERG (2006). A Statistical Analysis of Random Mutagenesis Methods Used for Directed Protein Evolution. *J. Mol. Biol.* 355, 858-871
- WONG, T. S., K. L. TEE, B. HAUER and U. SCHWANEBERG (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res.* 32, e26
- WONG, T. S., D. ZHURINA and U. SCHWANEBERG (2006). The Diversity Challenge in Directed Protein Evolution. *Comb. Chem. High Throughput Screening.* 9, 271-288
- ZHAO, H., L. GIVER, Z. SHAO, J. A. AFFHOLTER and F. H. ARNOLD (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotech.* 16, 258-261

