Enzyme directed evolution, a powerful algorithm to design tailor-made biocatalysts for the industry

Felipe de Salas, Marta Pérez-Boada and Susana Camarero

Centro de Investigaciones Biológicas Margarita Salas, CSIC. Ramiro de Maeztu 9, 28040 Madrid, Spain

Today the demand of enzymes for industrial applications is continuously increasing. The use of enzymatic biocatalysts offers numerous advantages over chemical processes due to the higher catalytic selectivity, reduced processing time, lower energy input and milder reaction conditions of the former, that result into cost-efficient, sustainable and eco-friendly processes. Enzymes are specific to certain reactions which allow enzymatic products to be tailored for specific needs. Hence, applications of microbial enzymes in the pharmaceutical, detergents, food and beverage, textile, leather, pulp and paper and other industries are numerous and increasing rapidly over less sustainable conventional methods. However, naturally occurring enzymes have evolved to act on natural conditions and their application requires tailoring the enzymes to work under the non-natural conditions of the industrial processes. Through the directed evolution of enzymes, we can adjust them to operate in new reaction conditions and optimise their catalytic activity towards recalcitrant or non-natural substrates, or make them able to catalyse new chemical reactions. In addition, commercialization of the enzymes requires scaled up production in industrial strains to increase the insufficient enzyme levels frequently provided by wild microbial producers. World-leader companies such as Novozymes, BASF, DuPont Danisco, Royal DSM and Codexis produce and commercialize tailor made biocatalysts.

Enzyme directed evolution

In the field of protein engineering, directed molecular evolution arose in the 90's as a powerful alternative to rational approaches to design biocatalysts with null or very little structural knowledge of the enzyme required. According to the Darwinian Theory, the diversity of life was created by random mutation and natural selection. Over many generations, beneficial mutations accumulate resulting in a successively improved phenotype. The power and simplicity of the evolution algorithm attracted scientists and engineers to try to implement this same approach to the molecular design.

Frances H. Arnold (CALTECH, US) was the pioneer to put into practice the concept of evolutionary engineering at the molecular level (Figure 1). She reported the first improved enzyme developed through iterative rounds of mutagenesis and screening applying a specific selective pressure. One year later, Willem P. Stemmer described the recombination of homologous genes (DNA-shuffling) as a breakthrough technology to accelerate enzyme directed evolution. Since then, Arnold's and many other groups refined directed evolution to design new enzymes working on unnatural conditions, to produce biofuels or for doing all kinds of complex chemistry. Due to her contribution in enzyme directed evolution to design biocatalysts with broad applications, from pharmaceuticals to renewable fuels, Frances H. Arnold was awarded the Nobel Prize in Chemistry in 2018. The emergence of advanced directed evolution techniques at the protein, pathway and genome level have sped up evolutionary engineering at outstanding levels during the last years.

From its beginnings, the directed evolution of enzymes has significantly expanded the repertoire of biocatalysts useful in the

chemical and biotechnology industry. For instance, Novozymes, the word leader company that introduced the use of proteases and lipases in laundry detergents, was the first to market variants of these enzymes improved by directed evolution in the 90's. Now enzyme directed evolution is an essential part of many industrial processes for manufacturing new enzymes for the pharmaceutical, chemical or food industries.

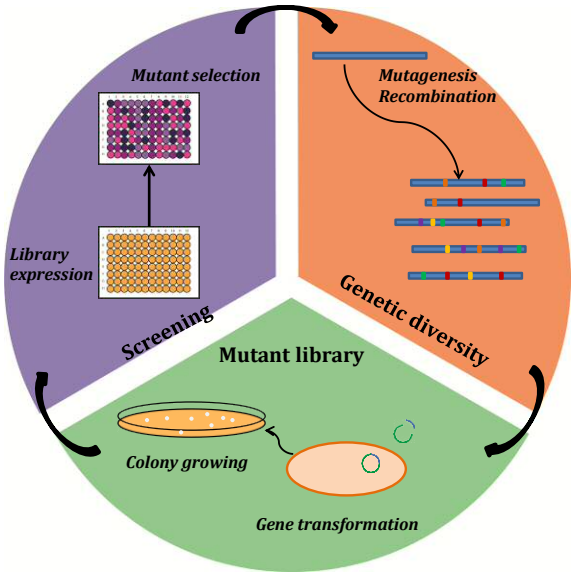

Figure 1: Schematic representation of a directed evolution cycle

Fundamentals of directed evolution

Enzyme directed evolution consists on iterations of mutagenesis of the starting gene and selection of the fittest mutants in each evolution cycle. By applying ad-hoc selective pressure during the screening of the mutant libraries, it is possible to direct the evolution of the enzyme towards the desired property in a step-wise manner.

- *Selection of microbial strain:*

Several bottlenecks, such as low enzyme production yields, little knowledge on physiology and secretion routes or lack of molecular biology tools are encountered when working directly with natural microbial producers. These drawbacks make mandatory the expression of the enzymes in more suited microorganisms. The latter usually display one or more of the following characteristics: high growth and expression rates, easy genetic manipulation, generally recognized as safe (GRAS) or low nutritional requirements. The thoroughly described physiology, fast growth, easy manipulation and great number of standard protocols available make *Escherichia coli* the preferred system for the directed evolution of prokaryotic enzymes (Figure 2). However, the specific modifications required by eukaryotic proteins often result in non-functional enzymes produced by *E. coli*, what heavily hinders its use for expression and engineering of eukaryotic biocatalysts. By contrast, eukaryotic systems such as some yeasts are capable of performing the post-translational modifications (e.g. glycosylation, signal peptide cleavage, etc) required to obtain fully active secreted enzymes.

The yeast *Pichia pastoris* offers high protein secretion yields in up-scaled enzyme production. However, its application as host for enzyme directed evolution is hampered by its low transformation efficiencies and difficult gene recovery, given that most expression vectors are integrated in the yeast genome and display low integrative efficiencies. On the other hand, *Saccharomyces cerevisiae* often provides low protein yields and hyperglycosylated enzymes what discourage its use as industrial production system. However, this yeast shows excellent properties for expressing the libraries of enzyme mutants obtained in directed evolution campaigns. These include high transformation efficiencies, with a range of available multi-copy vectors not integrated in the genome, which enable the easy and efficient genetic manipulation of the laboratory strains. Besides, the homologous DNA recombination machinery of *S. cerevisiae,*

with proof-reading activity, allows the *in vivo* recombination and cloning of PCR products in the linearized vector in one single step.

Therefore, a possible strategy is to use *S. cerevisiae* as host to conduct the directed evolution rounds required to achieve the improved enzyme, and then transfer the final evolved gene to *P. pastoris* for up-scaled production.

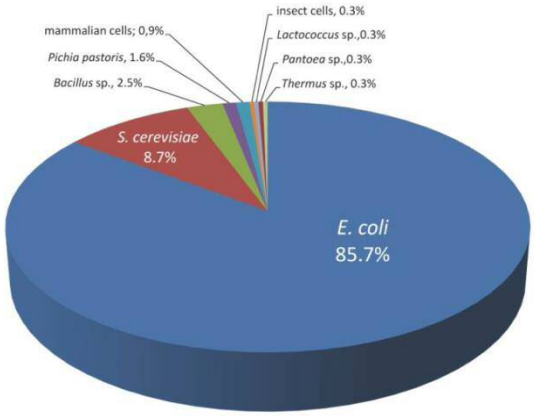

Figure 2: Preferred systems of choice for enzyme directed evolution (from Pourmir and Johannes, 2012).

- *Genetic diversity generation:*

In natural evolution, genetic diversification of the offspring enables the natural selection of the most fitted phenotypes and the inheritance of their genotypes during the evolution of the species. In laboratory evolution, mutagenesis and recombination of the parent genes allow to create the genetic diversity that results in the fitness difference that is subsequently explored and selected.

Usually, when there is no or little structure-function knowledge of the enzyme under study, genetic diversity is created by random mutagenesis of the whole gene using error-prone PCR or by recombination of homologous genes. The latter allows the accumulation of neutral or beneficial mutations from different parent genes without jeopardizing the functionality of the enzyme. DNA shuffling, consisting in random reassembly by PCR of the gene fragments obtained after DNase digestion of the parent genes, is one of the most used methods for DNA recombination. After this, quite a few other approaches have been developed for recombination of homologous and non-homologous genes. On the other hand, thanks to the homologous recombination machinery of some systems such as *S. cerevisiae*, *in vivo* recombination of PCR products and cloning is feasible only by creating overlapping fragments with the linearized vector.

When there is previous structure-function knowledge of the enzyme, mutagenesis can be focused to specific protein hotspots, that is, regions or positions associated with the activity or stability of the enzyme. The simplest focused evolution strategy is the saturation mutagenesis of a single position to explore the effect of all possible amino acid substitutions. Combinatorial saturation mutagenesis of different positions at the same time exponentially increases the number of possible mutation combinations that are impossible to explore in a screening assay. This problem can be partially overcome by using codons with reduced degeneration or following iterative saturation mutagenesis strategies that facilitate the systematic exploration of different regions and discover possible synergisms among mutations. The design of small libraries enriched in functional proteins (smart libraries) significantly reduces the screening effort.

In silico analyses and computational calculations greatly facilitate de design of smart libraries providing the identification of hotspots for enzyme engineering or predicting beneficial amino acid substitutions. Several software and web servers such as GLUE-It, CAVER CASTER, HotSpot Wizard, PoPMuSiC are publicly available for scientists working on enzyme design.

- ***High-throughput screening:***

One of the main obstacles of enzyme directed evolution is the development of a suitable selection or screening method to explore the mutant libraries that preserves the linkage between genotype and phenotype, allowing to detect slight changes in fitness among clones. The selective pressure has to be correctly applied during the screening according to the enzyme properties sought ("*you get what you screen for*", Frances Arnold).

Screening methods rely on the inspection of individual phenotypes by spatial separation of clones and the rapid assessment of optical features such as colour, fluorescence, luminescence or turbidity. However, most biomolecules are not associated with directly observable phenotypes and require a readily detectable reporter (GFP, luciferase, etc.) to enable the detection of gene expression. Another option is the use of surrogate substrates to generate a fluorescent, luminescent or colorimetric signal that is proportional to the enzymatic activity of interest. The optical signal can be directly screened in the colonies grown in agar plates or after expression of the enzyme library in liquid culture (microtiter plates). Since the increment in activity towards a particular substrate often entails the loss of other beneficial properties as enzyme stability or substrate promiscuity, multiple-screening assays are commonly performed and combined with stability assays (Figure 3).

In the above described screening methods, the capacity of the high-throughput measurement limits the exploration of the variability of the mutant libraries to a few thousands of clones spatially separated. Instead, a mass population can be interrogated at the level of individual cells using the cell membrane or cell wall to maintain the genotype-phenotype association. This is the case of high-throughput screening by flow cytometry, which enables the classification of fluorescence activated cells by measuring the fluorescence of individual cells and the automated separation of subpopulations by electrostatic deflection. This methodology has been further developed to perform miniaturized directed evolution experiments in microfluidic droplets, offering an ultrahigh-throughput screening platform that has revolutionized the scale and speed of screening as compared with microtiter-plates based systems.

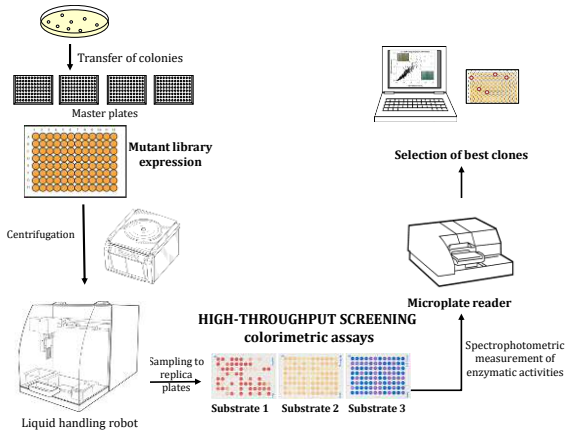

Figure 3: Schematic representation of a multiple high-throughput screening colorimetric assay carried out in our lab for the directed evolution of fungal ligninolytic enzymes in S. cerevisiae.

Engineering of enzymes for the pulp and paper industry

Pulp and paper industry is one of the largest business sector in the world. The production of paper and paperboard stands around 407 million metric tons per year.

Enzymes secreted by saprophytic fungi during biodegradation of plant biomass serve as valuable biocatalysts to transform cellulose, hemicelluloses and lignin, the main polymeric constituents of lignocellulose. Several of these enzymes are currently applied in the paper pulp industry for deinking, stickies removal, pitch control or pulp bleaching or refining. Enzymatic reactions have the potential to reduce manufacturing costs and the emission of toxic pollutants while creating novel high-value products.

For instance, lignocellulose degrading enzymes can also be applied to valorise leftover industrial lignocellulosic fractions to obtain new molecules usable as building blocks for the production of new bio-based materials or chemicals. This is precisely the goal of the European Project **WoodZymes** (www.woodzymes.eu), where underutilised wood fractions that are released in the side streams of kraft pulp mills will be valorised by the application of novel lignocellulolytic enzymes with extremophilic properties.

Fungal laccases are ligninolytic oxidases with significant potential as biocatalysts for diverse reactions due to their wide substrate promiscuity and low catalytic requirements (they only require O_2 from the air for their activation). These enzymes are naturally acidic and not always thermostable. In **WoodZymes** we are engineering fungal laccases by directed evolution, in order to develop extremophilic properties to the enzymes that make them tolerant to the extreme industrial operation conditions of elevated temperature and pH typical of kraft pulping, or to improve their activity towards lignin.

The goal is to apply the engineered extremophilic laccases in kraft pulp mills to valorise the lignin fraction released in the black-liquors, as well as to integrate the developed enzymes in the pulp bleaching sequence to enhance pulp properties, and in fibreboard manufacture to reduce the need of fossil-based resins as adhesives.

Conclusions

Enzyme directed evolution is a powerful algorithm that mimics the natural evolution process, but at a molecular level and laboratory time scale. Through iterations of mutagenesis and selection, it is possible to modify the enzyme's intrinsic properties so as to obtain tailor-made industrial biocatalysts active and stable at the required operation conditions. The application of enzymes as biotechnological tools in the conversion of plant biomass enables to obtain new bio-based products while at the same time enhancing the sustainability, efficiency and circularity of the industrial processes and reducing their environmental impact and energy requirements.

WoodZymes project has received funding from the Bio Based Industries Joint Undertaking (JU) under grant agreement No 792070. The JU receives support from the EU's Horizon 2020 research and innovation programme and the Bio Based Industries Consortium. This article reflects only the author's view and the JU is not responsible for any use of the information it contains.