Saccharomyces cerevisiae in directed evolution
An efficient tool to improve enzymes

David Gonzalez-Perez, Eva Garcia-Ruiz and Miguel Alcalde*

Department of Biocatalysis; Institute of Catalysis; CSIC; Madrid, Spain

Keywords: Directed evolution, Saccharomyces cerevisiae, DNA recombination, random mutagenesis, IvAM, IVOE

Abbreviations: bp, base pair; GIP, Coprinopsis cinerea peroxidase; CLERY, combinatorial libraries enhanced by recombination in yeast; CSM, combinatorial saturation mutagenesis; epPCR, error prone-PCR; ER, endoplasmatic reticulum; FACS, fluorescence activated cell sorter; HTPS, high-throughput screening; IvAM, in vivo assembly of mutant libraries with different mutational spectra; IVOE, in vivo overlap extension; MtL, Myceliophthora thermophila laccase; PeL, Pycnoporus cinnabarinus laccase; SOE, splicing by overlap extension; StEP, staggered extension process; VP, versatile peroxidase

Over the past 20 years, directed evolution has been seen to be the most reliable approach to protein engineering. Emulating the natural selection algorithm, ad hoc enzymes with novel features can be tailor-made for practical purposes through iterative rounds of random mutagenesis, DNA recombination and screening. Of the heterologous hosts used in laboratory evolution experiments, the budding yeast Saccharomyces cerevisiae has become the best choice to express eukaryotic proteins with improved properties. S. cerevisiae not only allows mutant enzymes to be secreted but also, it permits a wide range of genetic manipulations to be employed, ranging from in vivo cloning to the creation of greater molecular diversity, thanks to its efficient DNA recombination apparatus. Here, we summarize some successful examples of the use of the S. cerevisiae machinery to accelerate artificial evolution, complementing the traditional in vitro methods to generate tailor-made enzymes.

Throughout evolution, natural selection promotes the survival of specific organisms at the expense of thousands with trait/s that are not optimal to live in a given environment. Alterations to genes and enzymes are generated by processes such as random mutagenesis, DNA recombination, deletion and/or insertion, augmenting the diversity in this pool. These molecular modifications are then subjected to rigorous and constant testing by environmental factors, selection processes that drive the survival or disappearance of genes and enzymes. Typically, beneficial mutations (or neutral mutations that may become beneficial) accumulate and are recombined in the offspring. After successive generations of strict selective pressure, such mutations can give rise to new phenotypes. In February 2011, the Draper prize (considered the Nobel of Engineering) was awarded to Frances Arnold and Willem Stemmer for the development of Directed Molecular Evolution. This is a tool that has revolutionized the manner in which proteins are manipulated in the laboratory in order to improve their application in distinct industrial settings. By mimicking the mutation, recombination and selection processes that occur naturally in evolution, in vitro evolution provides a means of directing the evolution of genes toward specific goals in a manner that may not occur in a natural environment

Heterologous Functional Expression in Saccharomyces cerevisiae

The in silico analysis of genes/enzymes by computational methods is a valuable approach to engineer “smart” libraries reducing the exploration of the vast protein sequence space. This strategy can be combined with powerful tools for HTP-screening (e.g., fluorescence activated cell sorter (FACS)), providing another twist in enzyme engineering by laboratory evolution. Still, there are 3 basic premises to carry out a laboratory evolution experiment: (1) a suitable functional expression system; (2) reliable screening assays with which to detect improvements introduced after each round of evolution; and (3) the support of in vitro or in vivo methods to create enzyme diversity. The bacteria Escherichia coli is by far the most widely used host in directed evolution as it has a well-described physiology and it reproduces rapidly, making experiments less time-consuming. Moreover, standardized protocols are available to manipulate this bacteria and to rapidly recover the screened variants. While these characteristics generally hold true for prokaryotic proteins, bacterial hosts are less appropriate when working with eukaryotic genes, often resulting in misfolded, deglycosylated, non-functional or altered proteins, and the accumulation of the desired enzyme in inclusion bodies.
These shortcomings can be circumvented by using eukaryotic hosts such as *Pichia pastoris* or *Saccharomyces cerevisiae*. *P. pastoris* can secrete large amounts of proteins and mediate post-translational modifications. However, most vectors available for heterologous expression in *P. pastoris* are integrative—although there are a few exceptions—which together with a low efficiency of integration limits their use for HTP-screening (HTPS) and laboratory evolution. Recent efforts have sought to integrate linear expression cassettes in order to express mutant libraries of hydroxynitrile lyases. Nevertheless, a cumbersome mutant recovery process and poor transformation rates are still big obstacles which discourage scientists to take this approach. Fortunately, *S. cerevisiae* provides a solution to these bottlenecks as it exhibits high transformation efficiencies (from $1 \times 10^7$ to $1 \times 10^8$ transformants/µg DNA depending on the yeast strain), it performs post-translational modifications (e.g., processing of N- and C-terminal ends, glycosylation), and it possesses a fully developed secretory machinery that directs the secretion of proteins into the culture medium (bypassing the tedious lysis steps generally required when working with *E. coli* and avoiding any interference of complex lysate mixtures in the screening assays). *S. cerevisiae* may hyperglycosylate heterologous proteins (in some cases over 50% of the enzyme molecular weight) by the addition of mannose moieties at the Golgi compartment, a side-consequence of difficulties found during the exocytosis. This effect, although generally beneficial for protein stability—at the time that protect the enzyme from proteolytic degradation—generates a pool of isoforms which makes difficult the enzyme purification and biochemical characterization. Interestingly, in recent examples tackled in our laboratory (with high redox potential peroxidases and laccases, see below), mutations discovered by directed evolution helped us to surpass this hurdle by reducing the residence time at the Golgi, which generated new variants whose glycosylation degrees were below 10% showing a noticeable improvement in secretion yields. It is also worth noting that multicity episomal and bi-functional vectors are available to help identify and isolate the variants of interest screened from mutant libraries in *S. cerevisiae*. Finally, *S. cerevisiae* exhibits a high frequency of homologous DNA recombination with proof-reading activity, enabling in vivo recombination of the best mutant hits to occur at stages that prevent the incorporation of new mutations, as usually occurs in classical in vitro recombination protocols. Given these many advantages, *S. cerevisiae* has begun to be heavily exploited for the functional expression of evolved eukaryotic enzymes in the laboratory. Despite the advantages offered by *S. cerevisiae*, there are cases where the initial secretion levels of the target protein are not sufficiently high to perform artificial evolution. However, it has proved possible to adopt different strategies to considerably augment the secretion of such proteins in *S. cerevisiae*. One approach involves the introduction of random mutations in processing regions of the native gene to adjust the nascent
polypeptide to the specific attributes of the proteases found in the secretion route. This approach has been successfully applied in the evolution of the laccase from the ascomycete *Myceliophthora thermophila* (MiL) in *S. cerevisiae* for functional expression. 27 The single most beneficial mutation (producing a 10-fold enhancement in total activity) was found at the C-terminal tail of MiL, and it involved the introduction of a cleavage site for the KEX2 Golgi protease. An alternative strategy involves replacement of the signal peptide of the native protein with another signal leader that is recognized better by *S. cerevisiae*. In particular, the construction of fusion genes with the τ-factor prepro-leader from *S. cerevisiae* can drive protein secretion. 20, 21 Indeed, *S. cerevisiae* can process the native signal peptide of foreign proteins in some cases, as seen with the aspartic proteinase from *Mucor pauillacii* and the glucoamylase from *Aspergillus awamori*, among other examples. 22, 23 However, by replacing the native signal leader with the τ-factor prepro-leader, expression can be significantly enhanced. 24 As ligninolytic enzymes are remarkably difficult to express in non-fungal systems, 25–27 our group has used this approach to enhance the expression of these interesting oxidoreductases in *S. cerevisiae*. In recent studies performed in our laboratory, the native secretion leaders of genes encoding two different high redox potential laccases (PM1L, from basidiomycete *PM1*; and PcL from *Pyronema confluens*; and one peroxidase [VP, the versatile peroxidase from *Phleomus erygiti*) were replaced by the τ-factor prepro-leader. 28–30 The secretion of these fusion constructs was greater than that of these enzymes with their native leader (by at least one order of magnitude). Moreover, secretion could be further augmented by subjecting the entire gene (i.e., the τ-factor prepro-leader plus the mature protein) to directed evolution. This strategy allowed us to adjust both the τ-factor prepro-leader and the gene encoding the mature protein to the subtleties of the yeast secretory pathway. The canonical pre-leader is involved in the orientation and insertion of the nascent polypeptide during translocation to the endoplasmic reticulum (ER). Interestingly, mutations in the hydrophobic core of the pre-leader were discovered during the evolution of PM1L and PM11L that enhanced secretion several fold (A99D and V191D, respectively). Positions 9 and 10 of the pre-leader were further analyzed by constructing individual and double mutants containing the corresponding substitutions: A99D and V191D mutations exerted a 2.2-fold improvement in secretion individually but not when they were introduced together in the same variant. 31 Our results address that slightly increasing the hydrophilicity of the signal pre-leader may have beneficial effects on the interaction between the pre-leader and the signal recognition particle by improving the translocation of the polypeptide chain into the ER. 32 We also detected several interesting mutations in the pre-leader during the directed evolution of PM1L and PcL, that altered the affinity for sugar anchoring (N122K and S58G, respectively). As these positions correspond to 2 of the 3 N-glycosylation sites in the pro-leader, they may affect ER to Golgi protein transport. 33 Recent work has demonstrated that mutations in the τ-factor prepro-leader can enhance heterologous protein secretion in *S. cerevisiae* of a variety of proteins. 34 In fact, some of these mutations that increase secretion were the same as those identified for laccase fusion genes in yeast in our laboratory (position and nucleotide change). Finally, our evolved τ-factor prepro-leaders were fused with native (non-mutated) laccases, enhancing secretion by up to 40-fold and thereby corroborating the significance of the mutations induced by directed evolution. 35 Taken together, these findings suggest that the directed evolution of the τ-factor prepro-leader may give rise to a universal signal peptide for the heterologous expression of foreign proteins in yeast.

**Exploiting the Machinery of *S. cerevisiae* for Directed Enzyme Evolution**

Developing successful directed evolution experiments requires an appropriate array of molecular methods to allow the user to generate diversity. In this context, the power of *S. cerevisiae* cannot be underestimated. *S. cerevisiae* constitutes a simple and efficient vehicle to create libraries for directed evolution, exhibiting a high frequency of homologous DNA recombination with multiple recombination pathways generated by double-strand breaks. 36 A recent study reported that *S. cerevisiae* can recombine up to 38 overlapping single-stranded oligonucleotides and a linear double-stranded vector in just one transformation event. 37 Crossover areas can contain as few as 20 base pairs and as many as 200 homologous nucleotides. The importance of the length of the overlapping ends in the crossover region between the DNA fragment and the linearized plasmid to achieve high recombination efficiency has been demonstrated. Thus, a homologous region of at least 40 base pairs appears to be necessary to obtain recombination efficiencies of over 60%. 38 Recently, the full capacities of *S. cerevisiae* were challenged by a methodology known as DNA assembler, which was used to successfully assemble an entire biochemical pathway in a single step via in vivo homologous recombination. 39 In directed evolution, we use the DNA recombination machinery of *S. cerevisiae* to in vivo clone and recombine mutant libraries with the linearized vector, avoiding tedious ligation steps (Fig. 2A). To perform this type of experiment, it is necessary to engineer overlapping areas of approximately 40 bp of homology with the ends of the linearized vector, coupling the mutants generated to the corresponding screening assay. The number of crossover events among the inserts can also be enhanced (increasing the likelihood of recombining beneficial mutations between templates) by testing different overlapping regions with less homology to the linear vector, although the transformation efficiency may be compromised. In this context, in vivo DNA shuffling based on the *S. cerevisiae* recombination machinery is a powerful tool, speeding up the evolution process by shuffling parental genes with sequence homologies of ~70% at one point in the process where the whole autonomously replicating vector is repaired by the yeast’s in vivo gap repair mechanism (Fig. 2B). One of the first pioneering works of in vivo DNA shuffling was reported by Cherry and coworkers to engineer oxidative stability into the low-medium redox potential peroxidases from *Carpophilus cinnereus* (CiP). 40 Although in vivo DNA shuffling relies on the proof-reading device of *S. cerevisiae*, we observed better improvements
Figure 2. Different methods used to generate diversity using the *S. cerevisiae* toolbox. (A) epPCR from a single template followed by in vivo recombination in *S. cerevisiae*. (B) In vivo DNA shuffling. Several parental genes are recombined and cloned with a linearized vector into *S. cerevisiae* in a single step. (C) epPCR in conjunction with in vivo DNA shuffling. (D) IvAM (In vivo Assembly of Mutant libraries with different mutational spectra). Two or more distinct mutant libraries are generated by epPCR using polymerases with different biases. *S. cerevisiae* is transformed with the mutant libraries together with the linearized plasmid. (E) IVOE for combinatorial saturation mutagenesis or site-directed mutagenesis. The gene is amplified in two independent PCR reactions using mutagenized/degenerate primers. By engineering specific overhangs, the PCR products are then cloned into *S. cerevisiae* together with the linearized plasmid in a single transformation. (F) Mutagenic StEP (Staggered Extension Process). Several parental genes are used as templates during mutagenic StEP, promoting the random introduction of mutations during the short cycles of annealing and extension. The resulting mutant/recombined library is further shuffled by *S. cerevisiae*, together with the linearized plasmid. Stars represent single mutations.
in each cycle of evolution when error-prone PCR (epPCR) products of different templates were recombined in vivo in order to introduce new mutations in conjunction with recombination (Fig. 2C). In addition to our own works, 16,26,27,28,29 in vivo DNA shuffling has also been applied to other studies such as the engineering of chimeric enzymes from four different templates of Trametes C30 laccase with low and high redox potentials. 30

Given the inherent degeneracy of the genetic code and the fact that some errors in the genetic code cause silent mutations, diversity is more fault-tolerant to point mutations. Moreover, epPCR that some errors in the genetic code cause silent mutations, diverging this technique helps to enhance the mutational spectrum 31 (Fig. 2D). IVAM has been applied to the directed evolution of the Ml, in order to confer organic co-solvent tolerance. 32 We identified two beneficial mutations in two consecutive codons during the same cycle of evolution (G614D and E615K), probably induced as a consequence of the IVAM technique. Similarly, we employed IVAM to evolve VP toward thermal stability, raising the T50 by 8°C in the final VP mutant. 17 IVOE (In Vivo Overlap Extension) is a simple protocol applied to semi-rational or rational approaches such as combinatorial saturation mutagenesis (CSM), site-directed mutagenesis, site-directed recombination, insertions and deletions. IVOE is based on conventional SOE (Splicing by Overlap Extension), 33 although several of the in vitro steps in SOE are missing. Our method involves the engineering of mutagenic primers that generate PCR products with homologous regions, both with one another and with the linearized plasmid. These PCR fragments are transformed into the yeast together with the linearized plasmid with the desired mutation/s (Fig. 2E). 33 We previously improved the properties of hot-spot residues in Ml, by combining CSM with IVOE. 34 Moreover, in an attempt to enhance the activity and stability of PM1L, using IVOE, we performed site-directed mutagenesis studies to recover beneficial mutations discarded during the evolutionary pathway. The final PM1L mutant was readily secreted by S. cerevisiae (-8 mg/L) in an active and stable form with regards temperature, pH range and organic co-solvents. 35 We also used IVOE to demonstrate how VP secretion was affected by linking an extra four amino acid N-terminal tail to the mature protein (EAEA). The truncated VP variant was engineered by deletion mutagenesis through IVOE, confirming that the STE13 protease failed to process the extra N-terminal extension in the Golgi compartment of S. cerevisiae. 36 Another interesting application of IVOE developed in our laboratory involves the fusion of different enzymes with the n-factor prepro-leader of S. cerevisiae. 16,26,29

We engineered primers with homologous overhangs in order to generate fragments that were spliced in vivo to produce proteins fused to the n-factor prepro-leader, replacing the native signal peptide.

It is feasible to combine in vitro and in vivo methods for DNA-recombination to perform directed enzyme evolution. Indeed, in vitro DNA recombination and in vivo DNA shuffling were combined to increase the mutagenic spectrum of a given library (a method known as CLERY, Combinatorial Libraries Enhanced by Recombination in Yeast). 37 Similarly, we modified conventional StEP (Staggered Extension Process) 38 to enhance the likelihood of introduction of random mutations in the process (Mutagenic StEP, Figure 2F), and we combined this strategy with in vivo DNA shuffling in the same round of evolution to create a temperature, peroxide and alkalline-pH tolerant VP that was secreted readily by yeast (~22 mg/L). 17

In the past decade, S. cerevisiae has been used widely in the directed evolution of proteins. As a host, S. cerevisiae possesses all the necessary cellular machinery required to secrete active and functional eukaryotic proteins. As a biomolecular toolbox, S. cerevisiae permits new strategies to be designed that boost and direct the evolutionary process, complementing the traditional methods used to tailor enzymes à la carte. We hope that in the near future, S. cerevisiae will serve as a platform to support the directed evolution of artificial operons and metabolic pathways, thereby providing us with a powerful microbial cell factory for synthetic biology.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

This paper is based on research funded by EU Projects (FP7–3D-Nanobiodevices: NMP4-SL-2009–229255; FP7-KBBE-2010–4-26537: Peroxids; COST Action CM0701) and a National project (Evolufacel, BIO2010–19697).
