

1

Continuous Evolution of Proteins *In Vivo*

Alon Wellner¹, Arjun Ravikumar¹, and Chang C. Liu^{1,2,3}

¹University of California, Department of Biomedical Engineering, 3201 Natural Sciences II, Irvine, CA, 92697, USA

²University of California, Department of Chemistry, 1102 Natural Sciences 2, Irvine, CA, 92697, USA

³University of California, Department of Molecular Biology and Biochemistry, 3205 McGaugh Hall, Irvine, CA, 92697, USA

1.1 Introduction

Directed evolution is a powerful approach for engineering new biomolecular and cellular functions [1–3]. In contrast to rational design approaches, directed evolution exploits diversity and evolution to shape the behavior of biological matter by applying the Darwinian cycle of mutation, selection, and amplification of genes and genomes. By doing so, the field of directed evolution has generated important insights into the evolutionary process [4–6] as well as useful RNAs, proteins, and systems with wide-ranging applications across biotechnology and medicine [7–11].

To mimic the evolutionary process, classical directed evolution approaches carry out cycles of *ex vivo* diversification on genes of interest (GOIs), transformation of the resulting gene libraries into cells, and selection of the desired function (Figure 1.1). Each iteration of this cycle is defined as a round of evolution, and as selection stringency increases over rounds, either automatically through competition or manually through changing conditions (or both), this process can lead GOIs closer and closer to the desired function. This overall process makes practical sense for a number of reasons, especially for the goal of protein engineering (i.e. GOI encodes a protein). First, *ex vivo* diversification is appropriate, because test tube molecular biology techniques such as DNA shuffling, site-directed saturation mutagenesis, and error-prone (ep) polymerase chain reaction (PCR) [2] are capable of generating exceptionally high and precise levels of sequence diversity for any GOI. Second, transforming diversified libraries of the GOI into cells is appropriate, because each GOI variant needs to be translated into a protein in order to express its function, and cells, especially model microbes, are naturally robust hosts for protein expression. Third, carrying out selection inside cells is appropriate, because (i) cells automatically maintain the genotype–phenotype connection between the GOI and expressed protein that is necessary for amplification of desired variants,

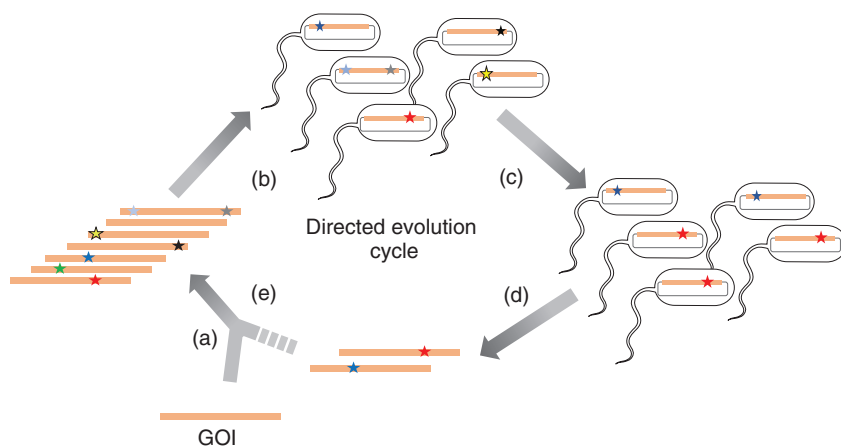


Figure 1.1 A schematic illustration of a typical directed evolution setup. (a) A GOI is diversified *ex vivo*, typically by applying an error-prone PCR to generate a GOI library. (b) The library is then cloned into an expression vector and transformed/transfected into cells that are subjected to (c) outgrowth and selection for enhanced protein activity. (d) Plasmid DNA that is enriched for library members with increased properties is extracted and (e) subjected again to diversification and selection. The directed evolution cycle is iterated until the desired outcome is achieved or until diminishing returns (a plateau is reached).

(ii) we often care about a GOI's function within the context of a cell, especially as metabolic engineering and cell-based therapy applications mature, and (iii) the use of cell survival as the output for a desired protein function allows millions or billions of GOI variants to be simultaneously tested – it is easy to culture billions of cells under selection conditions – in contrast to *ex vivo* screens that are much lower throughput. Survival-based selections are not always immediately available, but one can often find a way to reliably link the desired function of a protein to cellular fitness.

While sensible, the practical requirement that diversification should occur *in vitro* but expression and selection should occur *in vivo* in this classical directed evolution pipeline creates significant suboptimalities. First, the number of steps that can be taken along an adaptive path becomes few, since each round of *in vitro* mutation, transformation, and *in vivo* selection takes several days or weeks to carry out. Second, limited DNA transformation efficiencies result in strong bottlenecks of diversity that can mitigate the probability of finding the most optimal solutions in sequence space. Third, the number of evolution experiments that can be run simultaneously is minimal, because *in vitro* mutagenesis, cloning, and transformation are experimentally onerous, demanding extensive researcher intervention [12]. These shortcomings keep two highly promising categories of experiments largely outside the grasp of classical methods: first is the directed evolution of genes towards highly novel functions that likely require long mutational paths to reach (e.g. the optimization of multi-gene metabolic pathways or the *de novo* evolution of enzyme activity); and second is the large-scale replication of directed evolution experiments, needed in cases when many different functional variants of a gene are desired (e.g. the evolution of

multiple synthetic receptors for a collection of ligands) or when statistical power is required in order to understand outcomes in experimental evolution (e.g. probing the scope of adaptive trajectories leading to resistance in a drug target).

An emerging field of *in vivo* continuous directed evolution seeks to overcome these shortcomings by performing both continuous diversification of the GOI and selection entirely within living cells [13]. In this way, GOIs can be rapidly and continuously evolved through basic serial passaging of cells under selective conditions. This removes the labor-intensive cycling between *in vitro* and *in vivo* steps and the DNA transformation bottlenecks associated with the classical pipeline, creating a new paradigm for directed evolution that is limited only by the generation time of the host cell and the number of cells that can be cultured. These limitations are usually negligible – in most host organisms for directed evolution such as *Escherichia coli* and *Saccharomyces cerevisiae*, generation time is fast (20–100 minutes) and the number of cells that can be cultured is massive (10^8 – 10^9 ml⁻¹) – so the potential power of continuous systems is enormous. Moreover, *in vivo* continuous directed evolution is amenable to high-throughput experiments, because serial passaging is straightforward and can be automated at scale or converted to continuous culture using bioreactors [14–16]. In this chapter, we discuss various systems that partially or fully achieve *in vivo* continuous evolution (ICE).

1.2 Challenges in Achieving *In Vivo* Continuous Evolution

Before discussing how ICE can be realized, we shall first clarify why this is a challenging problem. The difficulty of achieving ICE of GOIs lies in the fundamental relationship between how fast one can mutate an information polymer and its length. Several theories predict that organisms face an “error threshold” at mutation rates on order $1/L$ (where L is the length of the genome), near which selection cannot maintain fitness, leading to gradual decline towards low fitness, or above which one is nearly guaranteed a lethal mutation every cycle of replication, leading to rapid extinction [17–20]. Because cellular genomes are large (e.g. $\sim 5 \times 10^6$ in *E. coli*, $\sim 1.2 \times 10^7$ in *S. cerevisiae*, and $\sim 3 \times 10^9$ in humans), this implies that evolution strongly favors low genomic mutation rates (e.g. $\sim 5 \times 10^{-8}$ substitutions per base [s.p.b.] in *E. coli*, $\sim 10^{-10}$ s.p.b. in *S. cerevisiae* and $\sim 3 \times 10^{-9}$ in human somatic cells) [21–23]. Experiment confirms this prediction. Drake observed empirically that mutation rates scale as $1/L$ across many organisms [17]; evolution experiments have shown that when mutator phenotypes do arise, they are accompanied by fitness costs and only transiently persist [19, 24–26]; and more direct tests in yeast find that there is indeed a mutation-induced extinction threshold at $\sim 1/L$, above which yeast cannot propagate [18]. Yet individual GOIs are small in comparison with genomes, so they are capable of tolerating much higher error rates. In fact, they *require* much higher per base error rates than genomes to generate the same amount of total mutational diversity, because they have fewer bases. Following the $1/L$ scaling, a typical 1 kb GOI should be able to tolerate mutation rates on order $\sim 10^{-3}$ s.p.b.

Therefore, the primary challenge in achieving rapid ICE is how to develop molecular machinery or other strategies that target rapid mutagenesis to only GOIs, allowing the host genome to replicate at mutation rates below its low error thresholds but driving the GOI at the high mutation necessary for fast generation of sequence diversity. When considering the level of targeting in the ideal case, the formidability of this challenge becomes quite apparent. Ideally, one should continuously mutate GOIs at rates close to their error threshold ($\sim 10^{-3}$) to maximize diversification but leave the genomic error rate completely unchanged, as the genome's error rate is evolutionarily optimized for host fitness. In *E. coli*, *S. cerevisiae*, and human cells, this means that on-target versus off-target mutagenesis must differ by 10^6 -fold, 10^7 -fold, and 10^7 -fold, respectively, which is much more than the 10- to 1000-fold targeting required in most synthetic biology problems involving molecular recognition. *How can we achieve such extreme precision in mutational targeting in the cell?*

There is yet another hard challenge in realizing ICE, which has to do with the durability of mutagenesis. Ideally, one wants a high rate of mutagenesis on the GOI to persist indefinitely (or at least for as long as the experimenter cares), so that a protein can traverse long mutational pathways towards desired functions. Because one needs to achieve mutational targeting to the GOI, there is almost always a risk to durability: any mechanism for targeting the GOI over the rest of the genome will necessarily rely on some cis-elements in or surrounding the GOI to mediate the targeting. If these cis-elements become mutated, which is quite likely since they are usually in or near the GOI undergoing rapid mutation, then mutagenesis will slow or stop. Ideally, a continuous evolution system will limit the chance that a cis-element for mutational targeting gets degraded. In the case that it does, an ideal system will remove the GOI containing the mutated cis-element from the population so that it can't fix in the population (through gradual mutational accumulation or a selective sweep if mutagenesis comes with a fitness cost) and end the continuous evolution process prematurely. *How do we achieve architectures for durability?*

Other challenges for ICE include generality across host organisms, the ability to mutate many genes simultaneously, and fine control over mutation rate and spectra; but the most defining ones are targeting and durability. In the remainder of this chapter, we review several *in vivo* continuous directed evolution platforms within the framework of these challenges. We highlight in Section 1.4.4 and note here that our recently developed orthogonal DNA replication (OrthoRep), among systems for ICE, seems uniquely capable of complete precision in mutational targeting (as far as we can tell), and is a highly durable architecture for enforcing prolonged mutagenesis in GOIs. We also highlight, in Section 1.3, phage-assisted continuous evolution (PACE), which has been remarkably successful for continuous biomolecular evolution. Although PACE is not an entirely *in vivo* system, it achieves complete precision in mutational targeting and durability – in fact by not being entirely *in vivo*, as we will explain. We do not discuss several powerful technologies for non-continuous *in vivo* diversification or streamlined diversification methods, such as MAGE [27], CREATE [28], DIVERGE [29], and CPR [30], but note that these are also promising approaches to protein evolution as they address some of the constraints of classical directed evolution methods. A summary of various characteristics of the systems we discuss is provided in Table 1.1.

Table 1.1 Comparison among approaches for *in vivo* continuous evolution.

| Approach | Systems | Mutation rate | Targeting of mutagenesis | Durability of mutagenesis | Number (simple estimates) and location of genes that can be evolved simultaneously | Generality across host organisms | Mutational spectrum | References |
|---|--|---|---|---|---|---|---|-------------|
| Continuous rounds of evolution with a conditionally replicating bacteriophage | PACE | Mutates GOIs at $\sim 10^{-3}$ s.p.b. | Complete targeting to the bacteriophage genome, since <i>E. coli</i> are constantly replaced | Indefinitely continuous since mutagenesis is enforced. In practice, this method is typically implemented for 1–3 weeks | 1–10 genes encoded on bacteriophage genome. | Currently in <i>E. coli</i> . Could be implemented with mammalian cells using non-integrating viruses (e.g. adenovirus). | Fairly unbiased mutational spectrum. | [33] |
| Targeted mutagenesis in <i>E. coli</i> with error-prone DNA polymerase I | ep Pol I/ColE1-based systems, CRISPR-guided DNA polymerases (EvolVR) | GOIs encoded near the ColE1 origin are mutated by ep Pol I at $\sim 10^{-3}$ s.p.b. CRISPR-guided Pol I can induce rates as high as 10^{-2} , but this quickly drops off after the guide region | Targeting with unfused ep Pol I is maximally only ~ 400 -fold. Fusion to nCas9 generally improves targeting to ~ 1000 -fold | Durability remains to be tested. Ep Pol I/ColE1 incurs significant off-target mutagenesis, which could quickly abrogate mutagenesis. EvolVR risks breaking down because it rapidly mutates the gRNA target region | 1–5 genes encoded on a plasmid with ep Pol I/ColE1. 1–20 genes on plasmids or at their endogenous genomic loci with EvolVR, depending on how many targeting sgRNAs one can stably encode. | Both systems are currently in <i>E. coli</i> . EvolVR should be fairly general across hosts, especially with the use of Phi29 DNAP. | ep Pol I mutates ColE1 plasmids with a bias towards transition mutations. EvolVR generates substitutions of all four nucleotide types, in a relatively unbiased manner. If needed, this can be improved through DNAP engineering. | [47–50, 54] |

(Continued)

Table 1.1 (Continued)

| Approach | Systems | Mutation rate | Targeting of mutagenesis | Durability of mutagenesis | Number (simple estimates) and location of genes that can be evolved simultaneously | Generality across host organisms | Mutational spectrum | References |
|--|----------------|--|---|--|---|---|---|-------------------|
| Yeast systems that do not use engineered DNA polymerases for mutagenesis | TaGTEAM, ICE | For TaGTEAM, $\sim 10^{-4}$ s.b.p. at 10 kb regions on both sides of the tetO array. For ICE, 1.5×10^{-4} , if excluding the rate of retro-transposition needed to induce mutagenesis | TaGTEAM offers targeting of genomic GOIs, however with low accuracy. ICE's targeting is theoretically good since off-target regions are not reverse transcribed | Durability remains to be tested. Off-target mutation and the requirement that retro-transposition occurs back into the original locus for continued evolution with ICE will likely affect durability | 1–10 genes on plasmids or at engineered genomic loci | Both systems are currently in yeast. ICE has been demonstrated in several diverged yeast species. TaGTEAM should function in <i>E. coli</i> and mammalian cells. ICE could be implemented in new hosts using retrotransposable elements similar to Ty1. | TaGTEAM generates a broad spectrum of both transitions and transversions. In addition, 25% of mutations are single base deletions. In ICE there is a 1:1 ratio between transitions and transversions. | [55, 58] |

| | | | | | | | | |
|---|--|--|--|---|---|--|--|--------------|
| Somatic hypermutation as a means for targeted mutagenesis of GOIs | Hypermutator B cell line, Ramos cell line, dCas9-AID fusions (such as CRISPRx), T7 RNAP-AID fusion | CRISPRx mutates GOIs at $\sim 5 \times 10^{-4}$ s.p.b. | Efficient targeting. No increase in mutagenesis rate was detected in an off-target locus. The hyperactive AID variant can create dense, highly variable point mutations within a region of 100 bp surrounding an sgRNA target site | Durability remains to be tested. | 1–10 genes on plasmids or at engineered genomic loci with the hypermutator B cell line, Ramos cell line, or T7 RNAP-AID fusion. Dozens of genes at endogenous genomic loci with dCas9-AID fusions | Systems depending on natural SHM are limited to mammalian cells. AID-fusions are currently available in mammalian systems or <i>E. coli</i> , depending on the system. AID fusions should be extensible to all host-types. | AID generates point mutations rather than insertions and deletions, and it favors transitions over transversions. However, repair pathways operate at AID-mutated loci to extend the scope of mutagenesis. | [67–73] |
| Orthogonal DNA replication | OrthoRep | Mutates GOIs at $\sim 10^{-5}$ s.p.b. | Complete orthogonality (at least 100 000-fold targeting) | Indefinitely continuous since mutagenesis is enforced. This method has been implemented for up to 300 generations without any sign of erosion | 1–10 genes encoded on a special orthogonal plasmid | Currently in yeast. Should be extensible to bacteria and mammalian systems using related protein-priming DNAPs. | TP-DNAP1-4-2 strongly favors transition mutations. This can be readily improved through DNAP engineering. | [74, 75, 78] |

Source: Esvelt et al. [33]; Fabret et al. [47]; Alexander et al. [49].

1.3 Phage-Assisted Continuous Evolution (PACE)

The most successful method for continuous protein evolution thus far is the PACE system developed in the lab of David Liu (Figure 1.2) [2, 12, 14, 15, 31–37]. PACE reimagines traditional “rounds” of directed evolution as generations of the M13 bacteriophage life-cycle, thereby transforming a step-wise and labor-intensive procedure into a continuous biological process. In PACE, GOIs are encoded in the M13 genome, and the resulting phage continuously replicate in a vessel (termed “lagoon”) that experiences a constant influx of *E. coli* cells. To create a selection pressure for GOIs to evolve, the activity of interest is coupled to phage survival. This is achieved by deleting the essential gene III (gIII), encoding coat protein III (pIII), from the M13 genome. The host *E. coli* strain is engineered to encode gIII in a genetic circuit that makes pIII expression dose-dependent on the desired activity of the GOI (see the following text for examples); so only phage that successfully evolves the GOI can trigger pIII expression and continue propagating. Due to the rapid generation time of M13 (~10 minutes without selection), evolution in this manner can iterate hundreds of times in just a few days.

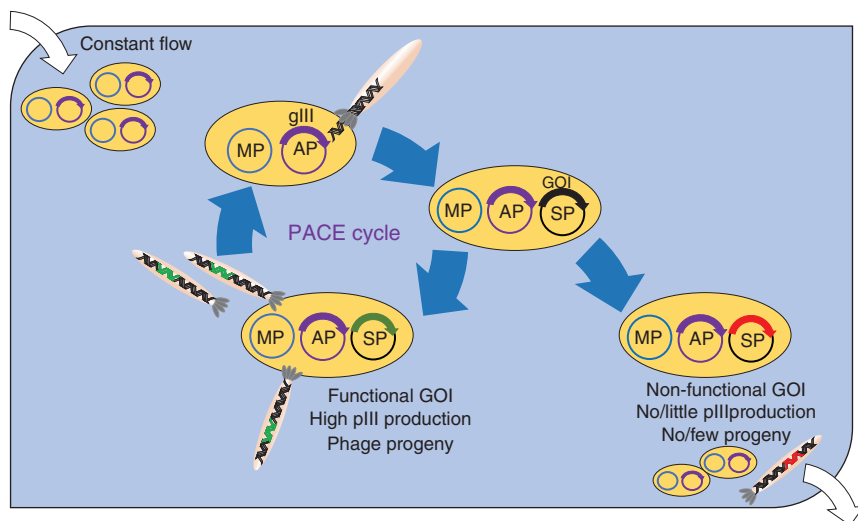


Figure 1.2 Pace. Phage carrying the selection plasmid (SP) encoding the GOI propagates on *E. coli* cells which are constantly flowing into the “lagoon” at a rate that does not permit their propagation but is longer than the phage life cycle, thus permitting phage replication. Upon infection, the SP (as well as the bacterial genome) experiences a high degree of mutagenesis due to the presence of a mutator plasmid (MP). In a PACE experiment, high GOI activity (green) is linked to drive strong gIII expression, resulting in progeny that can then infect incoming *E. coli*. No GOI activity (or a weak one, red) results in poor progeny production, becoming washed away from the lagoon at a larger rate (alongside bacterial cells). The system is designed to run for hundreds of generations without human intervention and result in the evolution of the GOI towards the desired activity. Source: Packer and Liu [2]; Badran and Liu [12]; Carlson et al. [14]; Dickinson et al. [15].

A key parameter in PACE is the *E. coli* flowrate, which should exceed their doubling time but be slower than the phage life cycle, allowing only phage to replicate in the lagoon (on average). Consequently, only phage accumulates mutations, whereas *E. coli* are physically prevented from doing so. High rates of mutation on the phage (and *E. coli*) genome are driven by a mutator plasmid (MP) that is carried by the *E. coli* cells and induced in the lagoon for error-prone M13 replication. The latest version of the MP is able to drive potent mutagenesis at $>10^{-3}$ s.p.b. by combining the effects of six different mutagenesis drivers [38].

Esvelt et al. first demonstrated proof of concept by evolving T7 RNA polymerase (RNAP) to initiate transcription from new promoter sequences [33]. pIII expression was bottlenecked at the level of transcription by encoding promoter sequences unrecognized by wild-type (wt) T7 RNAP (or any *E. coli* RNAPs), thus driving the selection to favor T7 RNAP variants that are able to efficiently recognize the new promoters. After eight days and 200 “rounds” of PACE, new T7 RNAPs emerged that could transcribe from the distant T3 RNAP promoter as efficiently as wt T7 RNAP does from its cognate promoter [33]. Similarly, T7 RNAP variants that efficiently initiate transcription with ATP or CTP, instead of GTP, were evolved. Since that landmark study, the ability to couple T7 RNAP activity to PACE has been exploited in a number of ways, ranging from basic adaptation studies to selections for split T7 RNAP [14, 15, 35–38].

In principle, PACE is applicable for the evolution of any biomolecular function that can be linked to pIII expression; and in just a few years since its inception, this has been realized in a wide range of applications beyond RNAP evolution. A notable example is the evolution of new DNA binding domains. Hubbard et al. employed the classic one-hybrid selection with PACE to evolve transcription activator-like effector nucleases (TALENs) with broadly improved DNA cleavage specificity [34]. Although TALENs are highly promising for gene editing, their major limitation is that they require the 5′ nucleotide of the target sequence to be T [39]. New TALEs (TALENs without the fused nuclease) were evolved with PACE by fusing the DNA binding domain of the canonical CBX8-targeting TALE to the ω subunit of *E. coli* RNAP. The PACE system was designed to include the TALE target sequence upstream of gIII. TALEs that successfully bind the target DNA recruit holoenzyme RNAP around the ω subunit, resulting in subsequent pIII expression. With this TALE selection, the identity of the target sequence can be custom-tailored, in this case, to encode noncanonical 5′ nucleotides. After using an additional negative selection (see below) that inhibited variants with promiscuous substrate specificity, Hubbard et al. were able to evolve TALE variants that displayed two- to fourfold increases in specificity for 5′ A, 5′ C, or 5′ G versus 5′ T, relative to wt TALE.

The one-hybrid PACE format was also used for overcoming one of Cas9’s main limitations, restricted protospacer adjacent motif (PAM) compatibility. This time, Hu et al. fused a catalytically dead variant of *Streptococcus pyogenes* Cas9 (dCas9) to the ω subunit of *E. coli* RNAP [40]. Then, the authors cleverly fed the lagoon with a mixture of host *E. coli* cells bearing a library of target sequences that covers all 64 possible PAM sequences, to select for broadened PAM compatibility. After PACE, several variants were isolated that could efficiently recognize NG, GAA, and

GAT as PAMs. Upon restoration of nuclease catalytic activity to these evolved dCas9 variants, the authors remarkably found that one of them, xCas9, exhibited greater DNA specificity than wt Cas9, even with its newly-gained broad PAM compatibility. This result challenges the widely-held assumption that there must be a trade-off between editing specificity and PAM compatibility and suggests that Cas9 can be improved through laboratory evolution to meet the most demanding challenges of CRISPR-Cas9 applications.

Another important form of PACE is its use with two-hybrid selection for the evolution of high-affinity protein-binders [31]. In the bacterial two-hybrid system, the ω subunit of *E. coli* RNAP is fused to a protein of interest, which is recruited to DNA through its interaction with a target protein. This target protein is fused to a DNA binding domain that localizes the complex at its cognate sequence encoded upstream of a reporter gene. If the protein of interest binds the target protein, then the RNAP holoenzyme can reconstitute around the ω subunit and drive expression of the downstream reporter. Badran et al. adapted this system for PACE using gIII as the reporter. After extensive optimization, Badran et al. were able to use this PACE format to evolve the insecticidal protein, *Bacillus thuringiensis* δ -endotoxin (Bt toxin) Cry1Ac, to bind and inhibit a new receptor in the gut of the insect pest *Trichoplusia ni* (TnCAD) [31]. Although wt Cry1Ac did not detectably bind TnCAD, the evolved variants were able to bind with nM affinity. Significantly, this strategy could overcome widespread Bt toxin resistance, which primarily occurs through mutational changes that inhibit binding to the native receptor of wt Cry1Ac. Badran et al. demonstrated this by showing that evolved Cry1Ac is highly potent at killing *T. ni* that are resistant to wt Cry1Ac. An exciting possibility for the future would be to evolve TnCAD to resist the new Cry1Ac variant, and then iterate this cycle in a study of molecular co-evolution.

Additional positive selections developed for PACE have enabled evolution of proteases that are drug resistant [32] or have altered substrate specificities [41], aminoacyl-tRNA synthetases (aaRSs) that can accept noncanonical amino acids [42], and protein variants with improved soluble expression [43]. Negative selections are also compatible with PACE, and are useful in cases where it is desirable to evolve high specificity towards the target substrate and restrict promiscuity towards others (especially the native substrate). This can be achieved by introducing a dominant negative allele of pIII, pIII-neg, that inhibits phage propagation [14]. The expression of pIII-neg can then be linked to the unwanted activity (e.g. recognition of the T7 promoter by T7 RNAP) for negative selection. (This strategy was successfully employed during TALEN and aaRS evolution.) Selection stringency and mutation rate are also important determinants of PACE outcomes and can be titrated [14, 35]. Lastly, we note that the Isalan lab developed a system related to PACE that accommodates the evolution of multiple genes, starting from combinatorial libraries. With this system, they were able to evolve a panel of orthogonal dual promoter-transcription factor pairs that were used to make multi-input logic gates [44, 45].

Clearly, PACE is a powerful method for continuous protein evolution, but as noted early in this chapter, it is not an entirely *in vivo* system. Rather, M13 serves as a

biological carrier of the GOI from one *E. coli* host cell to the next, with a given cell serving as a host of error-prone replication just once (on average). This ingenious design circumvents the challenges of *in vivo* mutational targeting. Since mutagenesis is induced in the lagoon, where *E. coli* briefly reside without doubling, mutation rates can be elevated entirely through untargeted mechanisms (and temporarily induced to be as high as desired), without consideration for replication of the *E. coli* genome. Even if *E. coli* cells stochastically replicate in the lagoon and become a source of cheater mutations (e.g. constitutive gIII expression), the flow rate ensures that any progeny are quickly diluted out. What's left in the lagoon is a population of M13 that selectively undergoes error-prone replication. In effect, targeting of mutations to the phage genome containing the GOI is complete, as the host *E. coli* is constantly replaced.

PACE also achieves durable mutagenesis by enforcing continuity. Replication of GOIs is intrinsically coupled to mutagenesis, through error-prone replication of the M13 genome. Any phage that escapes mutagenesis through a mutation in the phage genome's origin of replication, for example, must do so at the expense of being replicated. Only variants that continue to accumulate mutations can survive and propagate. And since *E. coli* cells do not persist long enough in the lagoon to evolve, the mutation rate experienced by phage remains unchanged. The durability of PACE is best evidenced by the long mutational trajectories traversed during evolution experiments, which have yielded protein variants with up to 16 mutations [46].

However, because PACE is not entirely *in vivo*, it suffers two major limitations. First, it requires continuous propagation of phage in a population of freshly diluted *E. coli* cells, which has been achieved thus far with a chemostat or turbidostat setup. This greatly limits the throughput and accessibility of PACE experiments, typically to fewer than ten replicates or experiments especially when different selection environments are desired across replicates. Second, PACE is restricted to selections that are linked to phage propagation. This precludes selections for *in vivo* phenotypes like tolerance or metabolism, as well as cell-based selections like fluorescence-activated cell sorting (FACS) or droplet sorting. These limitations motivate the need for continuous directed evolution systems that operate entirely *in vivo*.

1.4 Systems That Allow *In Vivo* Continuous Directed Evolution

1.4.1 Targeted Mutagenesis in *E. coli* with Error-Prone DNA Polymerase I

The first system that was able to perform continuous targeted mutagenesis *in vivo* was published in 2000 by Fabret et al. [47]. It was designed based on the developments in understanding the mechanism of ColE1 plasmid replication in *E. coli*. For plasmids that contain a ColE1 origin of replication, DNA polymerase (DNAP) I (Pol I) is responsible for elongating from the RNA primer that initiates replication at the origin. Pol I will extend for about 400–2000 bp, after which DNAP III

(Pol III), responsible for bulk DNA replication in *E. coli*, replaces Pol I [48]. When using a genome-encoded proofreading-deficient Pol I, genes that were cloned near the ColE1 origin experienced a 6- to 20-fold higher degree of mutagenesis over genes at more remote areas in the plasmid, showing targeting. The system's components were further combined with mismatch repair mutants to raise the mutation rate on GOIs yet another 20- to 40-fold, although significant increases in genomic mutation rates of at least several hundred-fold were observed. As a proof of concept, the authors evolved dominant negative variants of LacI that would outcompete a genomically-encoded wt LacI in binding its cognate operator, LacO. After 30 generations, LacI mutants that caused complete abolishment of wt LacI's binding to LacO were isolated. These variants were altered in their DNA binding domain but still formed tetramers with wt LacI, thereby abolishing LacI's repression at LacO.

Further improvement of the Pol I/ColE1 system was demonstrated in 2003 (Figure 1.3a) [46, 49]. Camps et al. modified the system to express the ep Pol I from a plasmid with a Pol I-independent origin of replication. Then, they used a host *E. coli* strain (J2000) whose genomically-encoded wt Pol I was temperature sensitive (ts) [49]. At restrictive temperatures, the ts Pol I becomes inactive such that only the ep Pol I acts, preventing the high-fidelity ts Pol I from competing for replication at the ColE1 origin. Based on prior studies of Pol I from the same lab [50], Camps et al. engineered a Pol I variant that was exceptionally error-prone, leading to mutation rates as high as 8.1×10^{-4} s.p.b at the GOI when the ts Pol I was inactivated. Mutagenesis expanded to about 3 kb from the ColE1 origin and

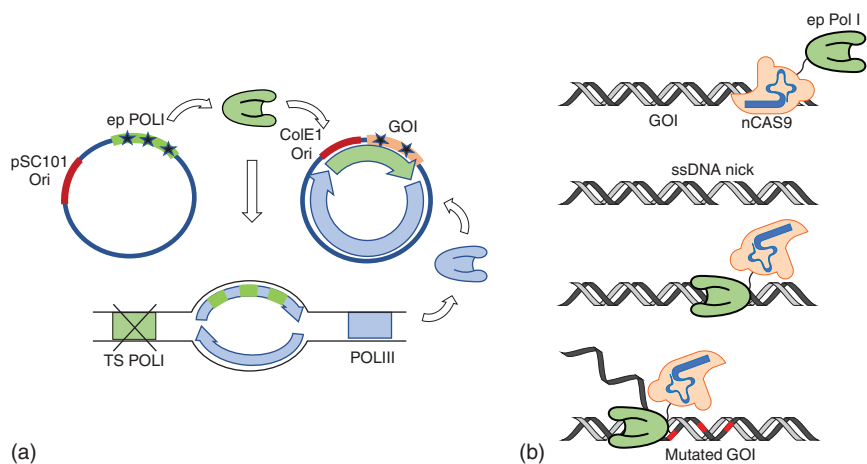


Figure 1.3 Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I. (a) An ep version of Pol I is expressed from a plasmid whose replication is driven by a non-ColE1 origin of replication (ori). The GOI is placed on the target plasmid near the ColE1 ori and thus targeted for mutagenesis. After 1–3 kb of ep replication, Pol III replaces Pol I to replicate the remainder of the plasmid with high fidelity. The genomic allele of POL I is temperature sensitive, such that enhanced mutagenesis can be induced by growth at the restrictive temperature. Source: Alexander et al. [49]; Camps et al. [46]. (b) The EvolvR system is composed of a CRISPR-guided nickase that nicks the target GOI, fused to ep Pol I that performs nick translation.

was evenly distributed within this region, albeit with certain biases in mutational preference. As a proof of concept experiment, Camps et al. demonstrated that their system could be used to evolve enzymes with diverged function by generating TEM-1 β -lactamase mutants that were able to hydrolyze a third-generation lactam antibiotic, aztreonam.

The ep Pol I/ColE1 system has subsequently been applied in a handful of additional directed evolution experiments. For example, Koch et al. used the system to prepare a library of terminal alkane hydroxylases with the aim of evolving variants that can oxidize butane [51]. Although they only used the system for the preparation of mutant libraries (i.e. as a mutator strain) and not for continuous evolution involving serial passaging under prolonged selection conditions, they demonstrated that one can create large libraries of GOI variants directly *in vivo*. In another application, an M13 phagemid with a ColE1 origin was made to encode LuxR and infect *E. coli* harboring the ep Pol I [52]. LuxR is a transcriptional activator and drove the transcription of an antibiotic resistance gene (β -lactamase) controlled by the *lux* promoter in the *E. coli*. Through several cycles of infecting fresh *E. coli*, antibiotic selection, lysis of *E. coli*, and phage isolation, LuxR evolved a 17-fold higher binding affinity to the *lux* promoter sequence.

While the ep Pol I/ColE1 system approaches ICE, it is limited by off-target mutagenesis and low durability. Because Pol I is responsible for Okazaki fragment mending throughout the genome and also participates in DNA repair [53], expressing an ep Pol I causes substantial mutagenesis genome-wide. Targeting of mutations to the GOI does occur – owing to the ColE1 origin, the limited role of Pol I in lagging strand replication, and special growth conditions optimized to time ep Pol I action with growth phases where genome replication activity is low – but is maximally only \sim 400-fold. Therefore, when highly ep Pol Is are used, it is possible that off-target mutagenesis will lower the fitness of the cell, causing fixation of suppressor mutations that abrogate the activity of ep Pol I. Still, the Pol I/ColE1 system represents a landmark development that encouraged the field to pursue new strategies for realizing ICE.

Perhaps the closest conceptual descendant of the ep Pol I/ColE1 system is a new *E. coli* continuous evolution system called EvolvR, which uses CRISPR-guided ep DNAPs to continuously target mutations to GOIs (Figure 1.3b). Rather than rely on the natural targeting of Pol I to ColE1, Halperin et al. [54] fused ep Pol I variants (and other DNAPs) to a nickase Cas9 (nCas9) that would serve two purposes. First, nCas9 would bring the ep Pol I to any GOI encoded on a plasmid or the genome using a guide RNA (gRNA). Second, the nCas9 would nick the target strand, creating a free 3'-OH substrate from which the ep Pol I could extend. Once nCas9 releases the nicked product, it is believed that ep Pol I then latches on and carries out error-prone extension from the nick. This highly clever idea was demonstrated in *E. coli* with a number of ep Pol I variants spanning different mutation rates and activities, as well as with a moderately ep Phi29 DNAP with high processivity. Using the most mutagenic ep Pol I, Halperin et al. measured a mutation rate approaching 10^{-2} s.p.b. (a 7.7 million-fold elevation compared to wt cells) at the first nucleotide 3' of the nCas9-induced nick. While this extreme mutation rate quickly dropped when

moving away from the nick, other Pol I and Phi29 DNAP variants with moderate error rates could achieve mutagenesis windows up to 350 bp. With these characteristics and with the potential to use multiple gRNAs to simultaneously target multiple parts of a gene, EvolvR could readily and efficiently generate sequence diversity on a GOI *in vivo* to support continuous evolution. Indeed, in a proof of principle experiment, Halperin et al. used EvolvR to rapidly evolve spectinomycin resistance by targeting mutagenesis to the *rpsE* gene and found new resistance mutations that were previously unknown.

Future studies and improvements on EvolvR will clarify how well it drives ICE for prolonged periods of time, needed to traverse long mutational pathways. Durability may be difficult in the current architecture, because the mutation rate is maximal at nucleotides within the target region of the gRNA, which if mutated, will reduce the ability of the system to continue inducing mutagenesis. Since the GOI can still be replicated (by high-fidelity host systems) in the absence of EvolvR function, this may result in the fixation of partially adapted GOI mutants that stop mutating, leading to premature cessation of evolution. In addition, EvolvR still has off-target elevations in mutation rate, presumably because ep Pol I or Phi29 can participate in genomic replication and/or because Cas9 has off-target binding. Strategies that use more processive ep DNAPs with no activity in normal genome replication and alternative CRISPR systems that nick outside the critical regions for gRNA targeting may overcome potential issues of targeting and durability. We also anticipate that this system should readily transfer to cell-types other than *E. coli*. Therefore, EvolvR is a highly promising new system for ICE with enormous potential, especially for the multiplexed evolution of genes at their endogenous genomic loci rather than on a plasmid.

1.4.2 Yeast Systems That Do Not Use Engineered DNA Polymerases for Mutagenesis

The first demonstration of continuous targeted mutagenesis *in vivo* in yeast was published in 2013 under the name TaGTEAM (Figure 1.4a), which stands for targeting glycosylases to embedded arrays for mutagenesis [55]. In TaGTEAM, mutagenesis at the GOI is initiated by recruiting a DNA glycosylase, which normally functions as the first step in the base excision repair (BER) pathway responsible for removing chemically altered DNA bases [56]. The authors adopted the yeast 3-methyladenine glycosylase, Mag1p, and fused it to the tet repressor (tetR) that binds a 19-bp operator sequence, tetO. By introducing a non-recombinogenic tetO array (with each tetO site separated by 10–30 bp of random sequence), the tetR-Mag1p fusion could be targeted to GOIs in the chromosome or plasmid. It is presumed that tetR-Mag1p targeting generates a build-up of unprocessed abasic sites at target loci, leading to replication fork stalling and recruitment of ep translesion polymerases [57]. This faulty repair can lead to both point mutations and frameshifts. To test their system for its ability to generate mutagenesis at a GOI, Finney-Manchester et al. introduced a 240X tetO array upstream of a *URA3* auxotrophic marker in a region of chromosome 1 that does not contain nearby essential genes. The distance between the tetO array and the marker was titrated to assess the size of the area subjected to mutagenesis. The

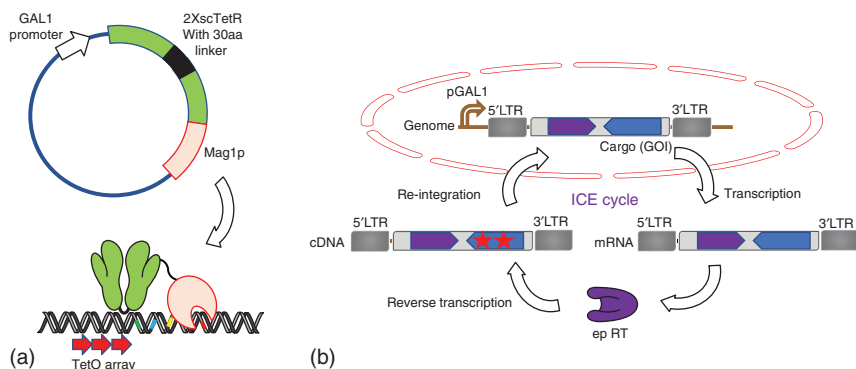


Figure 1.4 Yeast systems for targeted mutagenesis of GOIs. (a) TaGTEAM is achieved by fusing the yeast 3-methyladenine DNA glycosylase, Mag1p, to a tetR DNA-binding domain. Upon expression of the fusion from an inducible galactose promoter, the 20 kb region that is proximal to the tetO array experiences a high degree of mutagenesis. (b) In ICE, the GOI is cloned into an inducible Ty1 retrotransposon in the genome. The ICE cycle begins with inducible transcription of the retroelement followed by ep reverse transcription driven by Ty1's encoded rt. The cycle ends upon re-integration of the mutated cDNA into the genome. Source: Based on Crook et al. [58].

presence of tetR-Mag1p resulted in a >800-fold increase in mutation rate spanning a 10 kb region. However, the off-target mutation rate was also increased 40-fold in the absence of the array, indicating genome-wide mutagenesis by tetR-Mag1p. No direct applications of the system have been published to date, but this mutagenic strategy was important for opening new avenues of thought in the field.

ICE is another notable example of continuous evolution in yeast (Figure 1.4b), introduced in 2016 [58]. ICE adopts a strategy for DNA diversification that is based on the mutagenic properties of the Ty1 retrotransposon element. A GOI is cloned into the Ty1 cassette, which then gets transcribed into an RNA. Next, the RNA is reverse transcribed to form cDNA and reintegrated into the chromosome [59]. The mutagenic properties of the system stem from Ty1's self-encoded reverse transcriptase (rt), which introduces mutations at a rate of $\sim 2.5 \times 10^{-5}$ to $\sim 1.5 \times 10^{-4}$ per base per retrotransposition event [58, 60], thus allowing rapid mutagenesis of Ty1 and its embedded GOI. However, since mutagenesis depends on retrotransposition and the retrotransposition rate of Ty1 with a GOI inserted is low, the high mutation rate of Ty1's rt is only occasionally experienced on the GOI. Therefore, the authors carried out a series of experiments to increase the retrotransposition rate. By fine-tuning various parameters including the cargo's promoter strength, host genotype (i.e. deletions of certain host genes), cell density, temperature, initiator methionine tRNA expression (which acts to prime Ty1 replication), and inclusion of terminators, the authors were able to significantly increase retrotransposition rate. Altogether, the optimization process reached a mutation rate capable of generating up to 1.6×10^7 distinct mutants of a GOI per round per liter cultured [58]. Crook et al. then used ICE in three independent experiments to test the system's ability to evolve genetic material. In the first demonstration, URA3 was evolved for increased resistance to 5-fluoroorotic acid (5-FOA); in the second example, the Spt15p global transcription

regulator was evolved to confer a complex cellular phenotype of butanol resistance; and in the third example, a multigene pathway spanning 4.6 kb and containing two enzymes and a regulatory region was evolved for increased xylose catabolism. Additional experiments will clarify the extent to which ICE continuously mutates GOIs, as the ability for Ty1 elements to semi-randomly spread throughout the yeast genome [61, 62] could potentially complicate analysis, reduce mutational accumulation for the GOI, and diffuse the target of evolution. These issues could potentially be solved by somehow limiting Ty1 integration to a single location in the genome, turning the retrotransposon into a “retrocisposon,” and then increasing the “retrocisposition” rate to access high levels of diversification. In fact, the ability to achieve “retrocisposition” would also be important for reaching continuous evolution in other systems based on retroelement-mediated mutagenesis, such as a recently reported bacterial approach for *in vivo* genome editing and evolution [63]. Nevertheless, ICE is an important example of continuous evolution in yeast.

1.4.3 Somatic Hypermutation as a Means for Targeted Mutagenesis of GOIs

Some groups have harnessed one of nature’s built-in mechanisms for generating targeted DNA diversity, somatic hypermutation (SHM). In SHM, B cells create point mutations in their immunoglobulins (Igs) to drive antibody affinity maturation [64]. The enzyme responsible for SHM is Activation Induced cytidine Deaminase (AID), which deaminates cytidine (C) to generate uridine (U). This triggers various mismatch repair mechanisms resulting in a mutation rate of $\sim 10^{-3}$ s.p.b. at Ig loci [65]. Several researchers have successfully hijacked this natural mechanism for diversifying and evolving non-antibody proteins. In 2001, Bachl et al. set the stage for SHM-based protein directed evolution [66]. They demonstrated a high rate of reversion of a premature stop codon in a green fluorescent protein (GFP) cloned into a hypermutator B cell line (18–81) that expresses endogenous AID. They concluded that elevated reversion rates depended on AID and were rate limited by transcriptional levels of the target gene, in agreement with previous findings on SHM mechanisms [67, 68]. In 2004, Wang et al. applied SHM to the directed evolution of an entire open reading frame [69] by integrating a single copy of red fluorescent protein (RFP) into Ramos cells, which express endogenous AID, using a lentivirus. Through iterative SHM and FACS, RFP mutants with enhanced photostability and far-red emissions were evolved. The study was conducted in the pre-CRISPR era, and thus the RFP GOI was not targeted to an Ig locus but was rather integrated at various genomic locations within their cell population. However, the authors noted that the most evolved RFP variant, which they called mPlum, was located in the Ig heavy chain locus of chromosome 14, indicating that there is indeed a target locus where mutagenesis rates are highest, and that SHM is responsible for high levels of mutagenesis at the GOI. Yet it is expected that this targeting is incomplete, as mutation rates readily occur outside the Ig domain in cell lines that express endogenous AID [70].

Recently, a major development that avoids the use of hypermutator cell lines that express endogenous AID to mutate GOIs was independently published by

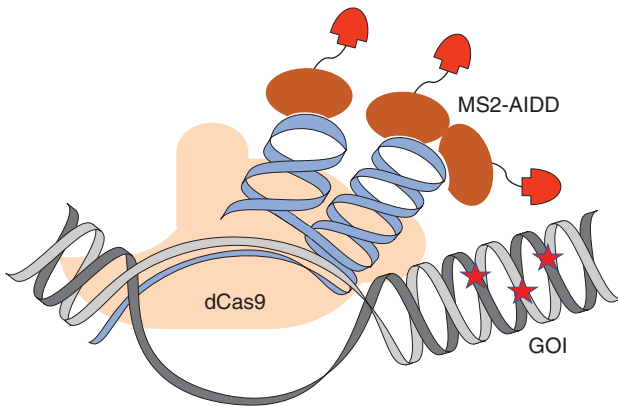


Figure 1.5 Targeted mutagenesis by somatic hypermutation (CRISPRx). A region of ~100 bp is targeted for mutagenesis by a dCas9 complexed with sgRNA-MS2 to a hyperactive cytidine deaminase (AID). Source: Hess et al. [71]; Ma et al. [72].

two groups (Figure 1.5) [71, 72]. Hess et al. linked AID to a catalytically inactive dCas9 using MS2-modified sgRNAs, which achieved precise targeting of SHM to defined loci in HEK293 cells [71]. The system, which they called CRISPRx, allowed targeted mutagenesis of multiple genomic locations simultaneously. Their reported mutation rate was $\sim 5 \times 10^{-4}$ s.b.p., which is similar to that observed for SHM [65]. In their first application, Hess et al. evolved GFP (excitation, 395 nm; emission, 509 nm) into enhanced green fluorescent protein (EGFP) (490/509 nm) by selecting for spectrum-shifted variants. Later, they mutated the target of the cancer therapeutic bortezomib, PSMB5, and identified known and novel mutations that confer bortezomib resistance. At the same time, Ma et al. developed a dCas9-AID fusion and targeted BCR-ABL for mutagenesis to efficiently identify known and new mutations conferring imatinib resistance mutations in chronic myeloid leukemia cells [72].

In both of these CRISPR-guided AID strategies, induction of mutagenesis at the GOI was followed directly by a single round of enrichment for the selected phenotype. Therefore, these studies do not directly demonstrate continuous evolution. However, multi-generation continuous directed evolution could be carried out using cell lines stably transcribing sgRNAs that tile the GOI. Although Hess et al. observed some limited off-target mutagenesis, owing both to off-target activity of AID and off-target binding of sgRNAs [71], the durability of this system, while untested, may be reasonably high, as the positions that are most prone to mutagenesis are outside of the spacer and PAM needed for sgRNA binding and multiple sgRNAs targeted to the same locus can be used. In addition, these methods are capable of introducing diversity at endogenous genomic loci, since CRISPR targeting is programmable.

Another strategy for targeting AID to GOIs is based on fusing AID to T7 RNAP [73]. The main advantage is that T7 RNAP induction could be precisely controlled in *E. coli*, and although not demonstrated by More et al., could be largely transferred between various organisms. Mutations accumulate during induction of

transcription, as the T7 RNAP carries AID over large stretches of DNA. Indeed, due to its high processivity, T7 RNAP can direct mutagenesis over several kb.

1.4.4 Orthogonal DNA Replication (OrthoRep)

Our lab recently developed a system for ICE, termed OrthoRep, based on orthogonal DNA replication [74, 75] (Figure 1.6). Fundamentally, OrthoRep can be described as a cell harboring a synthetic DNA replication system that propagates without affecting endogenous replication of the host genome. We implemented this additional replication system in the form of an orthogonal DNAP/plasmid pair, where orthogonality means that the DNAP is dedicated to the cognate plasmid and does not participate in genomic replication (unlike Pol I in the Pol I/ColE1 systems). This property allows us, broadly, to engineer DNA replication *in vivo* for user-defined purposes without harming the host. For the purpose of ICE, we can make the orthogonal DNAP as error-prone as desired, since the genome is completely spared from mutation. Then, GOIs can simply be encoded on the orthogonal plasmid and rapidly and continuously mutated by the orthogonal ep DNAP during evolution.

To create OrthoRep, we developed an orthogonal DNAP/plasmid pair in *S. cerevisiae* by leveraging the unique pGKL1 and pGKL2 (or p1 and p2) selfish elements [76, 77]. p1 and p2 are linear, high copy, DNA plasmids that can replicate autonomously in the cytoplasm of yeast. P1 and p2 each encode dedicated DNAPs,

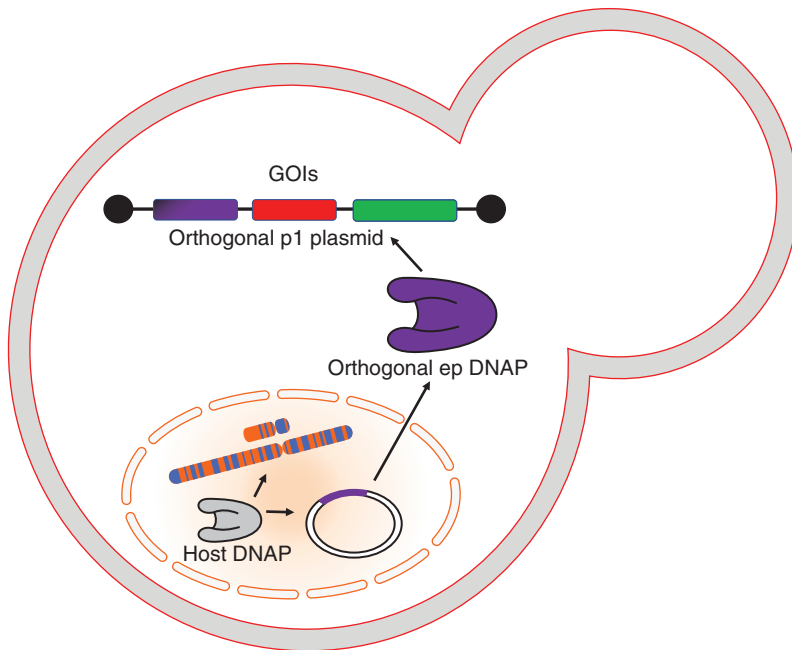


Figure 1.6 OrthoRep. In OrthoRep, GOIs encoded on the orthogonal p1 plasmid are replicated by the orthogonal ep DNAP. The genome is fully spared from mutation by the orthogonal ep DNAP. Source: Ravikumar et al. [74]; Ravikumar et al. [75].

TP-DNAP1, and TP-DNAP2, respectively, and rely on additional p2-encoded replication and transcription factors to propagate. For p1 and p2 replication, TP-DNAP1 and TP-DNAP2 recognize terminal proteins, TP1 and TP2, which are covalently linked to the 5' termini of p1 and p2. These terminal proteins act as origins of replication and serve as primers for initiation (in contrast to canonical RNA primers). This unique mechanism of protein-priming combined with the compartmental isolation of cytoplasmic p1 and p2 replication from nuclear DNA make these elements orthogonal to genomic replication. In fact, we demonstrated that TP-DNAP1/p1 and TP-DNAP2/p2 are mutually orthogonal DNAP/plasmid pairs, showing that TP-DNAPs are highly specific for their cognate TP-bound plasmid [78]. For OrthoRep, we repurposed the TP-DNAP1-p1 pair, while leaving the TP-DNAP2-p2 pair intact.

Since the native p1 mutation rate of $\sim 1 \times 10^{-9}$ s.p.b. was far too low for continuous evolution experiments, a large protein engineering effort was undertaken to reduce the fidelity of TP-DNAP1. First, the *TP-DNAP1* gene was deleted from p1 and TP-DNAP1 was expressed in trans from a nuclear expression vector. This enabled facile characterization of TP-DNAP1 variants and prevents ep TP-DNAP1s from mutating their own gene. Then, after mixed successes in predicting TP-DNAP1 mutators from related DNAPs, we cloned a scanning mutagenesis library of *TP-DNAP1* covering all single amino acid variants, and from this, screened $\sim 14\,000$ clones (\sim onefold theoretical coverage) for elevated mutation rates. This effort yielded a set of moderate mutators, which were then used to clone and screen combinatorial libraries, eventually leading to the discovery of a variant (TP-DNAP1-4-2) that mutates p1 at $\sim 1 \times 10^{-5}$ s.p.b. At this rate, OrthoRep mutates GOIs 100 000-fold faster than the *S. cerevisiae* genome.

The high p1 mutation rate driven by TP-DNAP1-4-2 showed no sign of erosion over extensive serial culturing (at least 90 generations and in unpublished experiments, at least 300 generations) and we found that the genomic rate ($\sim 10^{-10}$ s.p.b.) remained unchanged in the presence of TP-DNAP1-4-2, demonstrating continuous mutagenesis with complete orthogonality. Notably, p1 replication with TP-DNAP1-4-2 was experimentally determined to exceed the error-induced extinction threshold of the host genome (at most, 4.7×10^{-6} s.p.b.). Even moderate genomic rates of 1.6×10^{-7} – 5.2×10^{-7} s.p.b. were observed to be unstable over short durations, confirming that targeted mutagenesis with OrthoRep bypasses genomic error thresholds in a sustainable manner.

The utility of OrthoRep as a scalable directed evolution platform was demonstrated in an experiment that repeated the evolution of *Plasmodium falciparum* DHFR (*PfDHFR*) resistance to the antimalarial drug pyrimethamine, in 90 independent lines. By encoding *PfDHFR* on p1 in a *S. cerevisiae* strain lacking the endogenous dihydrofolate reductase (DHFR) gene (*DFR1*), we were able to evolve *PfDHFR* for strong resistance to pyrimethamine inhibition. This was done simply by serially passaging small volume (0.5 ml) cultures a few times under drug selection, allowing for the experiment to be done at the scale of 90 replicates. Adapted populations primarily converged on a previously undiscovered, but highly fit, region of the *PfDHFR* resistance landscape. Moreover, by repeating evolution many times, we were able to

capture rare stochastic events that steered populations away from the common path towards alternative high- and moderate-fitness outcomes. In fact, some of the events occurred in just 1/90 replicates, providing insight into the (ir)reproducibility of molecular adaptation that would not have been captured with smaller experiments. We also showed that evolution is highly dependent on the starting genotype by repeating evolution with a variant of *PfDHFR* containing a synonymous codon change, and finding a drastically different set of evolutionary outcomes.

Moving forward, we believe that the properties of OrthoRep are uniquely positioned to address the long-term technological challenges associated with ICE (Table 1.1). Although many approaches are capable of elevating mutagenesis of GOIs over genomic genes, the level of targeting in OrthoRep (at least ~100 000-fold) is currently unmatched. Furthermore, continuous mutagenesis in OrthoRep should be extraordinarily durable. Because replication of GOIs occurs exclusively through the action of the ep DNAP, inheritance of GOIs is intrinsically coupled to mutagenesis. Put in more specific terms, cells may acquire a disabling mutation in p1's origin of replication that ceases its mutagenesis, but these mutant p1s will no longer get replicated, and are immediately removed from the population. In addition, p1's origins of replication are mostly proteinaceous (the TPs) and cannot be mutated by the orthogonal DNAP, so the chance of a disabling mutation at the origin is low to begin with. In short, continuity of mutagenesis is enforced in the orthogonal DNA replication architecture. This durability will become increasingly important as the field makes headway towards more and more difficult protein functions that require many mutations to access. Such problems include the *de novo* evolution of enzymes, evolution of protein–protein interactions, and evolution of high-affinity therapeutic antibodies against difficult targets, all of which are at the forefront of the protein engineering field.

1.5 Conclusion

While directed evolution has been an extraordinarily powerful approach to protein engineering, the predominant approach of subjecting a GOI to rounds of *in vitro* PCR-based mutagenesis, transformation into host cells, and selection limits both the scale and depth of protein functions that can be evolved. By achieving *in vivo* continuous mutation at high rates targeted to only GOIs, the emerging field of ICE promises a fundamental transformation in the power and accessibility of directed evolution. We look forward to the continued application and improvement of the systems described in this chapter to ambitious protein engineering problems for years to come.

References

- 1 Turner, N.J. (2009). Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 5 (8): 567–573.

- 2 Packer, M.S. and Liu, D.R. (2015). Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16 (7): nrg3927.
- 3 Cobb, R.E., Chao, R., and Zhao, H. (2013). Directed evolution: past, present, and future. *AlChE J.* 59 (5): 1432–1440.
- 4 Bloom, J.D. and Arnold, F.H. (2009). In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Nat. Acad. Sci.* 106 (Suppl. 1): 9995–10000.
- 5 Wellner, A., Gurevich, M.R., and Tawfik, D.S. (2013). Mechanisms of protein sequence divergence and incompatibility. *PLoS Genet.* 9 (7): e1003665.
- 6 Soskine, M. and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11 (8): 572–582.
- 7 Jespers, L.S., Roberts, A., Mahler, S.M. et al. (1994). Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. *Nat. Biotechnol.* 12 (9): 899–903.
- 8 Coelho, P.S., Brustad, E.M., Kannan, A., and Arnold, F.H. (2013). Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* 339 (6117): 307–310.
- 9 McIsaac, R.S., Engqvist, M.K.M., Wannier, T. et al. (2014). Directed evolution of a far-red fluorescent rhodopsin. *Proc. Nat. Acad. Sci.* 111 (36): 13034–13039.
- 10 Goldsmith, M., Eckstein, S., Ashani, Y. et al. (2016). Catalytic efficiencies of directly evolved phosphotriesterase variants with structurally different organophosphorus compounds in vitro. *Arch. Toxicol.* 90 (11): 2711–2724.
- 11 Joyce, G.F. (2004). Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.* 73 (1): 791–836.
- 12 Badran, A.H. and Liu, D.R. (2015). In vivo continuous directed evolution. *Curr. Opin. Chem. Biol.* 24: 1–10.
- 13 d’Oelsnitz, S. and Ellington, A. (2018). Continuous directed evolution for strain and protein engineering. *Curr. Opin. Biotechnol.* 53: 158–163.
- 14 Carlson, J.C., Badran, A.H., Guggiana-Nilo, D.A., and Liu, D.R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* 10 (3): 216–222.
- 15 Dickinson, B.C., Leconte, A.M., Allen, B. et al. (2013). Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Nat. Acad. Sci.* 110 (22): 9007–9012.
- 16 Wong, B.G., Mancuso, C.P., Kiriakov, S. et al. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat. Biotechnol.* 36 (7): 614.
- 17 Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Nat. Acad. Sci.* 88 (16): 7160–7164.
- 18 Herr, A.J., Ogawa, M., Lawrence, N.A. et al. (2011). Mutator suppression and escape from replication error-induced extinction in yeast. *PLoS Genet.* 7 (10): e1002282.
- 19 Wilke, C.O., Wang, J.L., Ofria, C. et al. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412 (6844): 331.

- 20 Nowak, M. and Schuster, P. (1989). Error thresholds of replication in finite populations mutation frequencies and the onset of muller's ratchet. *J. Theor. Biol.* 137 (4): 375–395.
- 21 Lang, G.I. and Murray, A.W. (2008). Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178 (1): 67–82.
- 22 Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Nat. Acad. Sci.* 104 (41): E2774–E2783.
- 23 Milholland, B., Dong, X., Zhang, L. et al. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* 8: 15183–15183.
- 24 Doming, E. (2006). *Quasispecies: Concept and Implications for Virology*, 299.
- 25 Giraud, A., Matic, I., Tenaillon, O. et al. (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* 291 (5513): 2606–2608.
- 26 Notley-McRobb, L., Seeto, S., and Ferenci, T. (2002). Enrichment and elimination of mutY mutators in *Escherichia coli* populations. *Genetics* 162 (3): 1055–1062.
- 27 Wang, H.H., Isaacs, F.J., Carr, P.A. et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460 (7257): 894.
- 28 Liang, L., Liu, R., Garst, A.D. et al. (2017). CRISPR Enabled trackable genome engineering for isopropanol production in *Escherichia coli*. *Metab. Eng.* 41: 1–10.
- 29 Nyerges, Á., Csörgő, B., Draskovits, G. et al. (2018). Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc. Nat. Acad. Sci.* 115 (25): 201801646.
- 30 Abil, Z., Ellefson, J.W., Gollihar, J.D. et al. (2017). Compartmentalized partnered replication for the directed evolution of genetic parts and circuits. *Nat. Protoc.* 12 (12): 2493–2512, nprot.2017.2119.
- 31 Badran, A.H., Guzov, V.M., Huai, Q. et al. (2016). Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* 533 (7601): 58.
- 32 Dickinson, B.C., Packer, M.S., Badran, A.H., and Liu, D.R. (2014). A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* 5 (1): 5352.
- 33 Esvelt, K.M., Carlson, J.C., and Liu, D.R. (2011). A system for the continuous directed evolution of biomolecules. *Nature* 472 (7344): 499.
- 34 Hubbard, B.P., Badran, A.H., Zuris, J.A. et al. (2015). Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* 12 (10): 939–942.
- 35 Leconte, A.M., Dickinson, B.C., Yang, D.D. et al. (2013). A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* 52 (8): 1490–1499.
- 36 Pu, J., Zinkus-Boltz, J., and Dickinson, B.C. (2017). Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol.* 13 (4): 432–438.
- 37 Pu, J., Kentala, K., and Dickinson, B.C. (2017). Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chem. Biol.* 13 (2): 431–437.
- 38 Badran, A.H. and Liu, D.R. (2015). Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* 6 (1): 8425.

- 39 Deng, D., Yan, C., Pan, X. et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335 (6069): 720–723.
- 40 Hu, J.H., Miller, S.M., Geurts, M.H. et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556 (7699): 57.
- 41 Packer, M.S., Rees, H.A., and Liu, D.R. (2017). Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nat. Commun.* 8 (1): 956.
- 42 Bryson, D.I., Fan, C., Guo, L.-T. et al. (2017). Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* 13 (12): 1253.
- 43 Wang, T., Badran, A.H., Huang, T.P., and Liu, D.R. (2018). Continuous directed evolution of proteins with improved soluble expression. *Nat. Chem. Biol.* 13: 1253–1260.
- 44 Brödel, A.K., Isalan, M., and Jaramillo, A. (2018). Engineering of biomolecules by bacteriophage directed evolution. *Curr. Opin. Biotechnol.* 51: 32–38.
- 45 Brödel, A.K., Jaramillo, A., and Isalan, M. (2016). Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat. Commun.* 7: 13858.
- 46 Camps, M., Naukkarinen, J., Johnson, B.P., and Loeb, L.A. (2003). Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Nat. Acad. Sci.* 100 (17): 9727–9732.
- 47 Fabret, C., Poncet, S., Danielsen, S. et al. (2000). Efficient gene targeted random mutagenesis in genetically stable *Escherichia coli* strains. *Nucleic Acids Res.* 28 (21): e95.
- 48 Troll, C., Yoder, J., Alexander, D. et al. (2014). The mutagenic footprint of low-fidelity Pol I ColE1 plasmid replication in *E. coli* reveals an extensive interplay between Pol I and Pol III. *Curr. Genet.* 60 (3): 123–134.
- 49 Alexander, D.L., Lilly, J., Hernandez, J. et al. (2014). Random mutagenesis by error-prone Pol I plasmid replication in *Escherichia coli*. *Methods Mol. Biol.* 1179: 31–44.
- 50 Shinkai, A. and Loeb, L.A. (2001). *In vivo* mutagenesis by *Escherichia coli* DNA polymerase I. Ile(709) in motif A functions in base selection. *J. Biol. Chem.* 276 (50): 46759–46764.
- 51 Koch, D.J., Chen, M.M., Beilen, J.B.v., and Arnold, F.H. (2009). *In vivo* evolution of butane oxidation by terminal alkane hydroxylases AlkB and CYP153A6. *Appl. Environ. Microbiol.* 75 (2): 337–344.
- 52 Na, D., Lee, S., Yi, G.-S., and Lee, D. (2011). Synthetic inter-species cooperation of host and virus for targeted genetic evolution. *J. Biotechnol.* 153 (1–2): 35–41.
- 53 Allen, J.M., Simcha, D.M., Ericson, N.G. et al. (2011). Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Res.* 39 (16): 7020–7033.
- 54 Halperin, S.O., Tou, C.J., Wong, E.B. et al. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* 560 (7717): 248–252.
- 55 Finney-Manchester, S.P. and Maheshri, N. (2013). Harnessing mutagenic homologous recombination for targeted mutagenesis *in vivo* by TaGTEAM. *Nucleic Acids Res.* 41 (9): e99.

- 56 Nilsen, H. and Krokan, H.E. (2001). Base excision repair in a network of defence and tolerance. *Carcinogenesis* 22 (7): 987–998.
- 57 Boiteux, S. and Guillet, M. (2004). Abasic sites in DNA: repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair* 3 (1): 1–12.
- 58 Crook, N., Abatemarco, J., Sun, J. et al. (2016). *In vivo* continuous evolution of genes and pathways in yeast. *Nat. Commun.* 7: 13051.
- 59 Wilhelm, F.X., Wilhelm, M., and Gabriel, A. (2005). Reverse transcriptase and integrase of the *Saccharomyces cerevisiae* Ty1 element. *Cytogenet. Genome Res.* 110 (1–4): 269–287.
- 60 Boutabout, M., Wilhelm, M., and Wilhelm, F.-X. (2001). DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1. *Nucleic Acids Res.* 29 (11): 2217–2222.
- 61 Carr, M., Bensasson, D., and Bergman, C.M. (2012). Evolutionary genomics of transposable elements in *Saccharomyces cerevisiae*. *PLoS One* 7 (11): e50978.
- 62 Devine, S.E. and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev.* 10 (5): 620–633.
- 63 Simon, A.J., Morrow, B.R., and Ellington, A.D. (2018). Retroelement-based genome editing and evolution. *ACS Synth. Biol.* 11.
- 64 Wu, X., Feng, J., Komori, A. et al. (2003). Immunoglobulin somatic hypermutation: double-strand DNA breaks, AID and error-prone DNA repair. *J. Clin. Immunol.* 23 (4): 235–246.
- 65 Rajewsky, K., Forster, I., and Cumano, A. (1987). Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* 238 (4830): 1088–1094.
- 66 Bachl, J., Carlson, C., Gray-Schopfer, V. et al. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.* 166 (8): 5051–5057.
- 67 Chaudhuri, J., Tian, M., Khuong, C. et al. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422 (6933): 726–730.
- 68 Yu, K., Huang, F.-T., and Lieber, M.R. (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J. Biol. Chem.* 279 (8): 6496–6500.
- 69 Wang, L., Jackson, W.C., Steinbach, P.A., and Tsien, R.Y. (2004). Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* 101 (48): 16745–16749.
- 70 Wang, C.L., Harper, R.A., and Wabl, M. (2004). Genome-wide somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* 101 (19): 7352–7356.
- 71 Hess, G.T., Frésard, L., Han, K. et al. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13 (12): 1036–1042.
- 72 Ma, Y., Zhang, J., Yin, W. et al. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13 (12): 1029–1035.

- 73 Moore, C.L., Papa, L.J. 3rd, and Shoulders, M.D. (2018). A processive protein chimera introduces mutations across defined DNA regions *in vivo*. *J. Am. Chem. Soc.* 140: 11560–11564.
- 74 Ravikumar, A., Arrieta, A., and Liu, C.C. (2014). An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* 10 (3): 175–177.
- 75 Ravikumar, A., Arzumanyan, G.A., Obadi, M.K.A. et al. (2018). Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* 175 (7): 1946–1957.
- 76 Klassen, R. and Meinhardt, F. (2007). Linear protein-primed replicating plasmids in eukaryotic microbes. *Microb. Linear Plasmids* 7: 187–226.
- 77 Gunge, N. and Sakaguchi, K. (1981). Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGK11 and pGK12, from *Kluyveromyces lactis* into *Saccharomyces cerevisiae* by cell fusion. *J. Bacteriol.* 147 (1): 155–160.
- 78 Arzumanyan, G.A., Gabriel, K.N., Ravikumar, A. et al. (2018). Mutually orthogonal DNA replication systems *in vivo*. *ACS Synth. Biol.* 7 (7): 1722–1729.

