Published in final edited form as:

Nat Protoc. 2017 December ; 12(12): 2493–2512. doi:10.1038/nprot.2017.119.

Compartmentalized partnered replication for the directed evolution of genetic parts and circuits

Zhanar Abil¹, Jared W Ellefson², Jimmy D Gollihar², Ella Watkins², and Andrew D Ellington^{1,2,3}

¹Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas, USA.

²Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA.

³Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas, USA.

Abstract

Compartmentalized partnered replication (CPR) is an emulsion-based directed evolution method based on a robust and modular phenotype-genotype linkage. In contrast to other *in vivo* directed evolution approaches, CPR largely mitigates host fitness effects due to a relatively short expression time of the gene of interest. CPR is based on gene circuits in which the selection of a ‘partner’ function from a library leads to the production of a thermostable polymerase. After library preparation, bacteria produce partner proteins that can potentially lead to enhancement of transcription, translation, gene regulation, and other aspects of cellular metabolism that reinforce thermostable polymerase production. Individual cells are then trapped in water-in-oil emulsion droplets in the presence of primers and dNTPs, followed by the recovery of the partner genes via emulsion PCR. In this step, droplets with cells expressing partner proteins that promote polymerase production will produce higher copy numbers of the improved partner gene. The resulting partner genes can subsequently be recloned for the next round of selection. Here, we present a step-by-step guideline for the procedure by providing examples of (i) selection of T7 RNA polymerases that recognize orthogonal promoters and (ii) selection of tRNA for enhanced amber codon suppression. A single round of CPR should take ~3–5 d, whereas a whole directed evolution can be performed in 3–10 rounds, depending on selection efficiency.

INTRODUCTION

The improvement of proteins and pathways for biotechnology and biomedical applications can be achieved by directed evolution^{1,2}. Directed evolution typically uses iterative rounds of selection or screening to enrich randomized libraries for variants with altered or improved function. In this regard, it is unlike natural selection, which relies upon sequential, low-

<permission>Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to A.D.E. (ellingtonlab@gmail.com).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS Z.A. and A.D.E. wrote the manuscript. Z.A., J.W.E., J.D.G., E.W. and A.D.E. contributed technical detail to the protocol, and read, edited, and approved the final manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

frequency mutagenesis and selection. Because of this difference, it is frequently possible to more quickly evolve novel functions, because multiple mutations can be selected in parallel and the experimenter can alter the conditions and stringency of the selection in each round.

In a typical directed evolution experiment, genetic diversity is first created via either random or targeted mutagenesis, or through random recombination between a number of evolutionary related genetic elements^{3,4}. The size of the resulting library can vary based on the method, but is typically on the order of 10^6 or greater, which is much larger than the number of mutations that typically exist in a similar natural population.

Next, the created library is subjected to either rounds of screening, in which individual library members are evaluated and picked, or selection, in which active library members are enriched by survival or propagation advantage. In contrast to screening approaches, in which individual library members are physically separated, selection strategies are performed in bulk. Consequently, selection of proteins requires a means of genotype to phenotype isolation, which is achieved in a number of ways.

To date, the most common selection strategies are based on binding affinity, catalytic activity, or organismal survival. In selections for binding affinity, library variants are displayed on a cell surface^{5,6}, phage⁷, ribosome⁸, or mRNA⁹, and variants with the highest affinities are selected by exposure to an immobilized binding target while nonbinding library members are washed away. The genotype to phenotype isolation is achieved through genetic fusion of the gene of interest (GOI) to cell surface or phage coat proteins, by preserving the link between mRNA and the ribosome, or by covalently linking mRNA to its *in vitro* transcribed and translated protein¹⁰. There are also examples of using similar strategies for evolving catalytic function, although these often involve self-modification of the platform containing the GOI^{11–13}, and thus are inherently limiting.

In selections based on organismal survival, individual library members can confer survival and replication advantages upon a host organism. Here, the genotype to phenotype linkage is achieved by the cellular confinement of the live host organism. Selections of catalysis are much more facile, and can be achieved by conferring resistance to lethal conditions¹⁴ or complementing auxotrophy¹⁵. Unfortunately, selections based on organismal survival do not allow selection of proteins that are toxic or for the use of selection conditions that are harmful to the host organism. In general, the requirements for organismal survival can skew the phenotypes generated, often leading to more fit organisms, but less desirable proteins or pathways.

The advantages of emulsion-based selection methods

Another strategy for linking a genotype to its corresponding phenotype is via *in vitro* compartments. In this approach, proteinencoding genes and the corresponding gene expression products are encapsulated within a synthetic compartment, such as a water-in-oil emulsion, in which the actual selection or screening takes place. In some cases, *in vitro* transcription/translation reactions are used to create proteins, whereas in others, cells are used as transient vehicles for expression. DNA molecules in ‘successful’ compartments are amplified and passed to further rounds. This strategy improves upon conventional directed

evolution methods in two important ways: (i) it is far easier to select for a wider variety of functions, such as catalysis; and (ii) by removing the requirement for performing selections in live cells that must grow (rather than just be successful within an individual emulsion droplet), conditions can be applied and functions selected that would normally be toxic to an expression host.

Emulsion-based selections were pioneered by Tawfik and Griffiths, who demonstrated the use of water-oil emulsions for selecting methyltransferase activity¹⁶. *HaeIII*-encoding DNA and *in vitro* transcription/translation reaction components were mixed with mineral oil and surfactants to produce emulsion droplets. Active *HaeIII* methyltransferase produced in the compartments then methylated the *M.HaeIII* gene, which was in turn resistant to *HaeIII* restriction digestion. This selection strategy enriched active methyltransferases relative to a 10⁷-fold excess of genes encoding a different enzyme¹⁶.

Emulsion-based selections were adapted by Holliger and coworkers for the self-evolution of DNA polymerases via a method termed compartmentalized self-replication (CSR)¹⁷. In this method, a library of variants is expressed in *Escherichia coli*, the cells are emulsified and then lysed by heat treatment, and the best performing polymerase variant is allowed to amplify its own coding sequence in emulsion PCR (ePCR). Emulsions with billions of water-in-oil compartments are generated such that there is no more than one cell captured per emulsion droplet. Consequently, only active gene variants are carried to the next round of selection, while inactive variants are eliminated. As active variants are exponentially amplified, a functional DNA polymerase can be selected from a 10⁶-fold excess of inactive enzymes in a single round of selection¹⁸. Holliger and co-workers have used CSR to evolve DNA polymerase variants with increased thermostability¹⁷, an ability to incorporate unnatural and modified nucleotides^{19,20}, and an increased ability to amplify damaged DNA from extinct species²¹. Our group implemented CSR to engender the evolution of KOD DNA polymerase into a form that could accept RNA as a template, thus creating a high-fidelity and thermostable reverse transcriptase²². It should be noted that enzyme activities other than DNA polymerization can also be selected for by augmenting the CSR cycle; for example, an active variant of nucleoside diphosphate kinase (NDK) was selected from a mixture that contained an excess of inactive variant¹⁷. Here, NDK function provides substrates for Taq DNA polymerase by catalyzing the conversion of deoxynucleoside diphosphates (dNDPs) to dNTPs, and thereby contributes to the enrichment of its own active variant.

In vitro compartmentalization has also been implemented for engineering of enzymes with activities other than DNA modification or polymerization. As an example, Tawfik and co-workers generated variants of serum paraoxonase (PON1) by *in vitro* compartmentalization of single cells, sorted variants based on the production of a fluorescent product and FACS, and ultimately identified enzymes that had a 100-fold improvement in thiolactanase activity²³. In a similar example, Griffiths and coworkers generated Ebg enzyme variants with substantially improved β -galactosidase activity²⁴. Although this approach relies on screening rather than selection, the automated, high-throughput nature of FACS allows sorting of up to 10⁷ compartments per hour²⁵, and hence is a powerful alternative to selection systems. However, as flow cytometry requires aqueous media, the compartments

used for selection had to be emulsified twice to obtain water-oil-water double emulsions, which are highly polydisperse²⁶. The high variability between ‘double-bagged’ compartments makes it difficult to precisely discriminate incremental improvements in enzyme activities and reduces enrichment efficiency.

CPR

A further improvement in directed evolution technologies in general and emulsion-based selections in particular would be the ability to directly select for a variety of enzymatic activities or pathways, without the need for complex double-emulsification and FACS. To this end, we invented CPR, in which ePCR is used to enrich genetic elements with activities other than DNA polymerization. Gene circuits are designed in which GOI activity is linked to activation of DNA polymerase expression (Fig. 1a). Bacterial cells are transformed with plasmids encoding the diversified GOI and functionally linked DNA polymerase, and proteins are expressed *in vivo* for several hours, after which live cells are emulsified so that the majority of compartments contain no more than a single cell. During ePCR, cells are lysed, which leads to the release of DNA polymerase and a plasmid encoding the GOI variant into individual aqueous droplets containing primers for amplification of the GOI. Subsequent thermal cycling allows droplets containing the highest amounts of DNA polymerase to amplify the GOI variants with the highest activities (Fig. 1a).

As an example, we have used CPR to engineer T7 RNA polymerase (RNAP) variants with altered promoter recognition^{27,28} by using *E. coli* strains in which the Taq DNA polymerase coding sequence was placed under regulation of synthetic T7 promoter sequences (Fig. 1b). As a result, a panel of highly active T7 RNAP variants with minimal cross-reactivity were discovered²⁷. In another example, we evolved the tryptophanyl tRNA-synthetase: tRNA pair from *Saccharomyces cerevisiae* to facilitate the incorporation of an unnatural amino acid, 5-hydroxy-L-tryptophan (5HTP)²⁸. For this purpose, the Taq DNA polymerase coding sequence was interrupted with amber stop codons, which were suppressed by the orthogonal tRNA-synthetase: tRNA pair (Fig. 1b). The evolved tRNA synthetase exhibited 1,500-fold improvement in incorporation of 5HTP. Evolution of the suppressor tRNA yielded a completely orthogonal tRNA variant, which in combination with the evolved tRNA synthetase, drove 8.5-fold higher expression of a reporter gene with three amber stop codons relative to the parental tRNA²⁸. In addition, we applied a similar CPR selection strategy in combination with negative selection to evolve several new orthogonal *Methanocaldococcus jannaschii* tRNA variants with increased amber suppression activity and reduced promiscuous aminoacylation by endogenous aminoacyl-tRNA synthetases²⁹.

These examples demonstrate the potency of the method in evolving proteins as well as entire gene circuits for increased activity and specificity. There are also advantages to using CPR relative to other high-throughput screening or selection methods such as FACS³⁰. In general, larger libraries can be sampled with CPR relative to FACS. In addition, the *in vitro* replication phase allows the exponential amplification of the most active circuits. Thus, a small difference in circuit activity can potentially result in a much greater output difference and high selection efficiency. In addition, due to the short expression time, CPR minimizes

the potential fitness effects from GOI overexpression, and thus allows selection almost exclusively for the desired GOI function.

Unlike CSR, CPR can also implement negative-selection strategies to improve enzyme specificity. As an example, the enzyme activity can be linked to the expression of a conditionally toxic gene, such as modified phenylalanine tRNA synthetase (PheS-A294G) that permits incorporation of the toxic p-Cl-phenylalanine. For example, we showed that *M. jannaschii* tRNA variants could be evolved with reduced promiscuity in aminoacylation by endogenous aminoacyl-tRNA synthetases²⁹. Another way of designing a negative-selection circuit is by making an inverter circuit in which the GOI is an inhibitor or a repressor of DNA polymerase. Using this strategy, activity of GOI toward an undesired substrate inhibits ePCR and is selected against.

Although CPR is a powerful selection strategy, it still faces limitations. As with most other selection strategies, in which at least one stage of the procedure is carried out *in vivo*, CPR library size is limited by the bottleneck arising from *E. coli* transformation efficiency, which can be 10^8 - 10^{10} colony forming units. However, a well-designed library will probably still be larger and more complex than those for most other directed evolution methods. Another limitation arises from the requirement of linking enzymatic activity to DNA polymerase expression. CPR can potentially be applied for selecting genes with a variety of functions, especially enzymes and genetic elements that play a role in gene expression, but ultimately anything that enhances polymerase production can be selected for. For example, numerous reporter-based *in vivo* screening and selection strategies have already been established³¹, and could be adjusted for activating DNA polymerase expression by simply replacing the existing reporter with a thermostable DNA polymerase. However, if the impact on polymerase expression is initially small, a genetic element or the circuit as a whole may require painstaking codon, transcription, or translation optimization before the initiation of a successful selection.

Experimental design

To evolve a GOI, a gene circuit is first designed such that the desired activity of the product of the GOI activates DNA polymerase expression. There are numerous ways to establish such a circuit, but the most obvious one involves GOIs that directly impact the transcription or translation of the polymerase (Fig. 1b). As examples of the CPR procedure, we will detail the selection of (i) T7 RNAP variants with activity on an orthogonal promoter (Pcgg), which differs from the Pt₇ promoter at positions 11, 10, and 9 (ref. 28); and (ii) suppressor tryptophanyl tRNAs from *S. cerevisiae* that show enhanced incorporation of an unnatural amino acid, 5HTP. The first example application of the protocol relies on the expression of Taq DNA polymerase as a result of T7 RNAP variant binding to an orthogonal promoter directly regulating Taq DNA polymerase gene expression. The second example application of the protocol relies on amber codon suppression in the coding sequence of Taq DNA polymerase by a suppressor tRNA.

In this protocol, we will describe the development and exploitation of a gene circuit, including use of a GFP assay to confirm and optimize the proper functioning of the initial circuit, mock selections to validate and optimize the selection design, cloning and

transformation of a library, and induction of protein expression before emulsification (Fig. 2). The CPR step itself includes ePCR and the recovery of amplicons from the emulsion for the next round of selection (Figs. 2 and 3).

GFP assay (Steps 1–12).—Because polymerase expression cannot be readily monitored, establishing an initial circuit that involves GFP production is useful for validation or optimization of a gene circuit designed around a given GOI. Such an assay essentially involves replacing Taq polymerase with GFP in the circuit in which Taq polymerase would normally reside. It should be noted that because directed evolution is often used to generate novel activities in enzymes, proteins, or pathways, it can be difficult to predict if a given circuit will select for a given novel property. However, as a starting point, control positive and negative gene circuits can be designed to function with the parental—or wild-type (WT)—GOI, and can thereby be assessed for their ability to discriminate between active and inactive variants. For example, WT T7 RNAP should drive GFP or Taq production from a WT T7 (P_{T7}) promoter, whereas a catalytically dead T7 RNAP should not. As negative controls, one can also use truncated enzymes or an entirely different genetic element. In addition, Western blotting can be used to monitor the production of the thermostable polymerase.

We also recommend placing the GOI and the GFP (and ultimately DNA polymerase) coding sequences on separate, compatible plasmid vectors; this will allow individual variations in either the GOI expression plasmid or the GFP (or Taq polymerase) reporter plasmid to be more quickly assessed. *E. coli* is then transformed with the entire genetic circuit, protein expression induced, and GFP fluorescence measured in live cells using flow cytometry or microplate reading. It should be noted that no emulsification is performed in this step, and that the assay aims at analyzing the gene circuit functioning, not the emulsion quality.

Mock selection.—Before doing CPR selections, it is advisable to first perform a mock selection to confirm that the designed gene circuit functions as expected with Taq polymerase, and that selection (rather than just expression) of GOI function is possible. As above, the Taq DNA polymerase is encoded on a separate compatible plasmid vector. For mock selections, a plasmid encoding a polymerase and found that it gives greater amplicon yields as compared with those of Taq DNA polymerase (Z.A., data not shown), which might be due to increased processivity of KOD DNA polymerase.

A mock CPR selection is next carried out, in which *E. coli* expressing the DNA polymerase as well as either WT or the catalytically inactive GOI are mixed at various ratios (10^{-1} – 10^{-8} WT:inactive), the bacteria are emulsified and the GOIs eventually recovered via ePCR. Recovery of the active (WT) GOI from the mixture with the inactive variant can then be analyzed in a few ways. First, the yield of recovery PCR can be indicative of the progress of a selection, because the initial fraction of an active variant in the pool will affect ePCR yields. Thus, the WT GOI template gives the highest yield of recovery PCR, dead GOI yields only faint background amplification, and various dilutions of WT in dead GOI result in intermediate yields of recovery PCR (see ANTICIPATED RESULTS). Similarly, during directed evolution, an increasing fraction of active variants from round to round should give higher yields of recovery PCR, and thus provide a signal that the selection is proceeding.

Next, enrichment of active variants should be analyzed either by sequencing or by restriction digestion and analytical gel electrophoresis. Successful selection WT GOI, as well as another plasmid encoding a catalytically inactive or truncated GOI, will both be required. It is often convenient to construct the inactive GOI such that it differs from the wild type by the presence of a unique restriction site. Although our published examples of CPR demonstrate excellent results with Taq DNA polymerase^{27–29}, other DNA polymerases can also be used for this purpose. We have since tried using the KOD DNA will result in 10^2 – 10^5 -fold enrichment of the WT GOI, judged by the restriction pattern (see ANTICIPATED RESULTS).

Preparation of mutant libraries (Steps 13–17).—The details of library design and generation have been covered in many other places^{3,4} and are beyond the scope of this protocol. We will, therefore, only briefly describe the key points for molecular cloning of T7 RNAP and tRNA pools in appropriate plasmid vectors. Libraries can be created using error-prone PCR^{32,33} or site-saturation mutagenesis³⁴. The choice of both method and the size of the library will depend largely on the function being sought. If a function is probably distributed throughout the sequence of a gene, such as thermostability, then PCR mutagenesis maybe more appropriate. If a function, such as substrate-binding, is focused on a particular set of amino acids, then site-saturation mutagenesis may be the way to go. In particular, if a crystal structure of the gene of interest or its homolog is available, it is appropriate to randomize the residues known to interact with the specific binding partner or residues known to be involved in catalysis. That said, using these different mutational methods in combination is often most effective. For example, random mutagenesis can reveal positions or regions that most greatly affect function, and these can then be more deeply mutagenized via site-saturation mutagenesis.

Molecular cloning of libraries into a destination vector can also be performed by a multitude of approaches, the most commonly used of which are classic restriction/digestion cloning³⁵, Gibson assembly³⁶, and Golden Gate assembly³⁷. To ensure a 10^7 – 10^8 library size for the initial selection, 1–10 μ g of ligated DNA should be introduced into cells. An *E. coli* strain of choice is transformed with the generated library; the strain should preferably have good expression of the GOI, which maybe determined during the initial GFP induction experiments described above. *E. coli* competent cell preparation and transformation are likewise beyond the scope of this protocol, so the user is referred to available resources^{38,39}.

Expression and emulsification of the pool of variants (Steps 18–27).—In this step, the diversified pool of the gene of interest and the DNA polymerase are expressed *in vivo*. An *E. coli* strain of choice is used to express the proteins, usually for 1–7 h at 37 °C (the conditions should be optimized for each application). Short induction times help to ensure that there is minimal effect of selection pressure or genetic circuit function on host viability.

Induced cells are subsequently entrapped in water-oil emulsions (Fig. 4a–i). The compartments separating individual cells must satisfy two main criteria: stability to thermal cycling and impermeability to macromolecules such as DNA and proteins. There are a number of protocols for the generation of water-in-oil emulsions^{16,40–42}, one of which⁴⁰ we

found to be optimal for CPR, with slight modifications. A water-oil mixture was emulsified in a TissueLyser (for the purpose of generating a more even size distribution), and to prevent cell lysis, we replaced the carbide bead that is usually used with the TissueLyser with a rubber stopper from a 1-ml syringe. Optimization of vibration time and frequency on the TissueLyser allowed generation of stable emulsions (Fig. 4g) that had an average size ~5 μ m (Fig. 4h) and that were capable of withstanding the thermal cycles inherent to PCR.

According to our estimates, more than 10^{10} compartments are generated in a single 1.5-ml emulsion reaction, which is enough to cover a 10^8 -variant library, a number that is readily accessible by initially transforming at least 10 pg of ligated library, as described above. In our experience, emulsification of $<5 \times 10^8$ cells results in most compartments carrying no more than a single cell (Fig. 4i). Despite the variability in droplet size, both CSR and CPR have the capability of distinguishing subtle differences in enzyme activity, which is mainly due to the exponential enrichment of GOIs during the ePCR step. Although it may be possible to generate emulsion droplets that have different volumes or even greater uniformity using microfluidics platforms^{43,44}, these approaches require additional instrumentation that is not readily available in most laboratories. The emulsification approach described herein is robust enough for effective enrichment and can be carried out by almost any molecular biology lab.

Emulsion PCR and selection (Steps 28–39).—During the initial heat-denaturation step of the ePCR, the *E. coli* cells are lysed and the contents, including the expressed DNA polymerase and the plasmid encoding the variant GOI, are released into the aqueous droplet that already contains dNTPs and primers for GOI amplification. Subsequent thermal cycles lead to amplification of each library variant, and the representation of a given variant in the overall library should follow the amount of polymerase that it succeeded in producing. By using modified primers (either biotinylated primers or primers tagged with unique sequences at their 5' ends; Fig. 1a,b), the purification of amplified DNA from template (plasmid) DNA is greatly simplified. To enable subsequent cloning, the vector and amplicon must be designed so that unique restriction sites are present internal to the primerbinding sites. Introduction of restriction sites in the sequence of the primer is discouraged, as it may facilitate the carryover and accumulation of smaller nonspecific PCR amplification products in the next round of selection.

Upon completion of ePCR, the emulsion is broken with an alcohol (phenol in this protocol), and the amplified product is purified from cell debris and PCR primers. Note that because only a fraction of the emulsion droplets contain GOI variants that promote polymerase production, and thus PCR amplification, very little product is generated at this step and the product is usually not observed by gel electrophoresis. Therefore, one or two additional 'bulk' PCRs are performed using the ePCR product as template and added DNA polymerase to produce sufficient amounts of amplicon for use in the next round of CPR (see next section). Before additional PCRs can be performed, however, plasmid DNA still present in the sample should be degraded with the DpnI restriction enzyme to prevent unwanted background amplification. As DpnI cleaves only methylated DNA substrates, it will remove the plasmid DNA isolated from most *E. coli* cell lines (avoid cells lacking the Dam

methylase) while leaving ePCR products intact. This digestion step is necessary to minimize carryover of the plasmid DNA to the recovery step.

Recovery of PCR products for the next round of selection (Step 40).—After the emulsion is broken and the plasmid DNA is DpnI-digested, the amplification product should be mostly free from cell debris, plasmid and genomic DNA, Taq DNA polymerase, and primers. However, the trace amounts of plasmid DNA still present in the sample can contaminate the subsequent bulk PCRs. To minimize amplification from contaminating plasmid DNA, our group uses two strategies (Fig. 3a,b). The first strategy involves using a biotinylated primer in the ePCR step to facilitate subsequent streptavidin-bead purification of the amplified product, and then amplification by what we call recovery PCR (Fig. 3a,b) (Step 40A). The second strategy is based on introducing unique tags at the 5' ends of the ePCR product and using these tags as primer binding sites in recovery PCR (Step 40B). The recovery PCR step can be followed by an additional re-amplification PCR (re-amp PCR) step, in which nested primers are used to obtain enough product for the next round of selection. It should also be noted that the gene of interest can be further diversified by using error-prone PCR during the re-amp PCR step.

To provide broad insights into CPR, the two selection examples described in this protocol differ in their recovery procedure. T7 RNAP variants were selected using the biotinylated ePCR product, whereas tRNAs were selected using ePCR primers with unique tags. Nonetheless, these recovery strategies should be interchangeable. As the strategy that uses unique tags requires fewer reagents and gives robust results, it is the strategy that is now routinely used by the authors and is recommended to the user.

To check for primer and nonspecific template carryover into recovery PCR, the researcher may in parallel run control recovery PCR reactions, in which primers or dNTPs are omitted. These should yield noticeably less or no PCR product as compared with experimental ePCR/rescue PCR.

The recovered PCR product can be digested with the same restriction enzymes that were used for library generation, and then re-ligated into an empty vector, as was done for the initial library generation. Competent cells can then be transformed with the ligation product to initiate a subsequent round of CPR. It has been estimated that for many genes 30–50% of random single mutations are detrimental to the protein activity and/or stability, and only 0.01–1% are beneficial¹. Therefore, in libraries generated by error-prone PCR with 2–3 mutations/gene or in libraries generated by site saturation at several residues, >90% of the variants can be inactive, and will be readily eliminated from the naive library in the first round of CPR. In consequence, most CPR functional enrichments will be at least several hundred-fold. In practice, the amount of DNA actually recovered is usually even less than 1 % of the original input, and will be amplified by many thousand-fold during subsequent PCR steps. Inefficiencies in ligation and in transformation in round 2 and beyond should, therefore, not artificially reduce library size. Indeed, subcloning in subsequent rounds of CPR can be scaled down 10-fold or more without the fear of losing active library variants.

Progressing through a CPR selection (Steps 41 and 42).—A number of selection rounds (3–10) must be performed before substantial enrichment can be observed. Enrichment of improved variants can be tracked in a number of ways. First, successful selection rounds should yield higher recovery PCR yields with each increasing round. Second, it is often helpful to use a GFP assay in which GFP is substituted for the Taq polymerase gene to test the CPR selection. We suggest ligating a small fraction of the amplified pool into the circuit that expresses GFP after each round of selection, to track enrichment. Ideally, the GFP assay will show an increase in GFP fluorescence with each increasing round (see ANTICIPATED RESULTS). Finally, sequencing of the cloned library can be used to determine whether the library is converging on individual variants, which can then be characterized individually using the GFP or other relevant assays.

In addition to deciding when and how to alter library composition, more fit and functional GOIs can be obtained by gradually changing the selection pressure; for example, by decreasing the degree of induction of the GOI to force individual enzymes to work more quickly. Furthermore, to alter or improve specificity, one or more positive-selection rounds can be interspersed with negative-selection rounds. Decisions regarding how to progress through a selection are often idiosyncratic, and are best made based on familiarity with the GOI and its fitness landscape.

MATERIALS

REAGENTS

GFP assay

- B121 (DE3) gold competent cells (Agilent, cat. no. 230132) or other *E. coli* cells of choice
- SOB medium powder (Becton Dickinson, cat. no. 244310)
- D-Glucose (Fisher, cat. no. BP350500)
- 2×YT broth powder (Becton Dickinson, cat. no. 244010)
- Chloramphenicol powder (GoldBio, cat. no. C-105)
- IPTG (Sigma-Aldrich, cat. no. 15502)
- PBS (Lonza, cat. no. 17–516F)

Additional reagents for RNAP selection

- pACYC-GFPM2 plasmid or a derivative with an orthogonal promoter (for basic plasmid maps, see Supplementary Figure 1; available upon request from the Ellington group or Addgene (cat. no. 102791))
- pQE-80L-Kan-T7WT plasmid encoding T7 RNAP (available upon request from the Ellington group or Addgene (cat. no. 102790))
- Kanamycin sulfate powder (GoldBio, cat. no. K-120)
- Empty pQE vector (Qiagen, cat. no. 32943)

- Restriction enzyme BamHI-HF (New England BioLabs, cat. no. R3136)
- Restriction enzyme HindIII-HF (New England BioLabs, cat. no. R3104)
- BL21(DE3) *E. coli* strain carrying the pACYC-Taq plasmid (or its derivative harboring an orthogonal promoter of choice; for basic plasmid maps, see Supplementary Fig. 1; available upon request from the Ellington group or Addgene (cat. no. 102793))

Additional reagents for tRNA selection

- Plasmid pACYC.GFP1Amb encoding GFP with a premature amber codon (available upon request from the Ellington group or Addgene (cat. no. 102792))
- Plasmid pRST. 11B.AS3.4 encoding *S. cerevisiae* suppressor tryptophanyl tRNA or a derivative with inactive tRNA (for basic plasmid maps, see Supplementary Fig. 1; available upon request from the Ellington group or Addgene (cat. no. 102789))
- Carbenicillin disodium salt (GoldBio, cat. no. C-103)
- 5-hydroxy-L-tryptophan powder (5HTP; Sigma-Aldrich, cat. no. H9772)
- Restriction enzyme KpnI-HF (New England BioLabs, cat. no. R3142)
- Restriction enzyme BsrGI-HF (New England BioLabs, cat. no. R3575)
- BL21(DE3) *E. coli* strain carrying the pACYC.Taq. 1 Amb plasmid (for basic plasmid maps, see Supplementary Fig. 1; available upon request from the Ellington group or Addgene (cat. no. 102794))

Library transformation and expression

- DNA Clean 8c Concentrator-5 kit (Zymo Research, cat. no. D4003)
- LB medium powder (Becton Dickinson, cat. no. 240220)
- Agar (Fisher, cat. no. BP1423)
- Syringe filters (VWR, cat. no. 28145–501)
- SOB medium powder (Becton Dickinson, cat. no. 244310)
- 2×xYT broth powder (Fisher, cat. no. BP246710)
- Dextrose (D-glucose; Fisher, cat. no. D16)
- Chloramphenicol powder (Fisher, cat. no. BP904)
- Kanamycin sulfate powder (Fisher, cat. no. BP906)
- Carbenicillin disodium salt powder (Fisher, cat. no. BP26481)
- Isopropyl β -D-l-thiogalactopyranoside (Sigma-Aldrich, cat. no. 15502)

Emulsion PCR

- Tris base (Fisher, cat. no. BP152)

- Hydrochloric acid (HCl; Fisher, cat. no. A508)
- Sodium chloride (Fisher, cat. no. S271)
- Tegosoft DEC (diethylhexyl carbonate; Evonik, cat. no. 14858–73–2)
- Mineral oil (Sigma-Aldrich, cat. no. M5904)
- Abil WE 09 (polyglyceryl-4 isostearate (and) cetyl PEG/PPG-10/1 dime-thicone (and) hexyl laurate; Evonik)
- Potassium chloride (Fisher, cat. no. P217)
- Magnesium chloride (Sigma-Aldrich, cat. no. M8266)
- dNTP mix, 10 mM each (Thermo Fisher, cat. no. R0192)
- PCR primers (Integrated DNA Technologies; for sequences, see Table 1)
- Tetramethyl ammonium chloride (TMAC) solution, 5 M (Sigma-Aldrich, cat. no. T3411)
- RNase A, 1 mg/ml (Ambion, cat. no. AM2271)

Breaking emulsions

- Phenol/chloroform/isoamyl alcohol (25:24:1; vol/vol/vol; Fisher, cat. no. BP 17521) ! CAUTION Phenol and chloroform are hazardous. Avoid direct contact and inhalation. Opening and transferring the liquids should be carried out in a chemical fume hood.
- Chloroform (Fisher, cat. no. C298) ! CAUTION Chloroform is hazardous. Avoid direct contact and inhalation. Opening and transferring the liquids should be carried out in a chemical fume hood.
- 5Prime Phase Lock Gel Heavy tubes (QuantaBio, cat. no. 2302830)
- Restriction enzyme DpnI (NEB, cat. no. R0176)
- CutSmart Buffer, 10× (NEB, cat. no. B7204)
- QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
- Sodium acetate (Sigma-Aldrich, cat. no. S2889)
- Glacial acetic acid (Fisher, cat. no. A38)

Recovery PCR

- Phusion high-fidelity DNA polymerase (supplied with 5× HF buffer; NEB, cat. no. M0530)
- PCR primers (Integrated DNA Technologies; sequences listed in Table 2)
- dNTP mix, 10 mM each (Thermo Fisher, cat. no. R0192)
- TAE, 50× liquid concentrate (Apex Bio Research Products, cat. no. 20–194)
- Seakem LE Agarose (Lonza, cat. no. 50004)

- Ethidium bromide (Thermo Fisher, cat. no. 15585011) ! CAUTION Ethidium bromide is hazardous. Avoid direct contact and inhalation. Wear protective lab equipment when using this reagent.
- QIAquick Gel Extraction Kit (Qiagen, cat. no. 28704)

Additional reagents for recovery of a biotinylated ePCR product

- Dynabeads MyOne Streptavidin C1 (Thermo Fisher, cat. no. 65001)
- Tris base (Fisher, cat. no. BP 152)
- EDTA disodium salt dihydrate (Sigma-Aldrich, cat. no. E5134)
- Sodium chloride (NaCl; Fisher, cat. no. S271)
- Hydrochloric acid (HCl; Fisher, cat. no. A508)

EQUIPMENT

General equipment

- Water bath (Fisher, cat. no. FSGPD02)
- Incubator/shaker (New Brunswick, model no. Innova 44)
- Microcentrifuge (Eppendorf, model no. 5418)
- Vortex mixer (Scientific Industries, model no. SI-0236)
- Thermocycler (Bio-Rad, model no. T100)
- Microwave (LG, model no. LCS1112ST)
- 2-ml Microtubes (Eppendorf, cat. no. 022431048)
- 1.5-ml Microtubes (Eppendorf, cat. no. 022431021)
- PCR strip tubes (Genesee Scientific, cat. no. 27–125)

GFP assay

- 96-Well black microplates (Corning, cat. no. CLS3915)
- 96-Well clear microplates (Corning, cat. no. 3370)
- Microplate reader (Tecan, M200 PRO)

Library Transformation and Expression

- Electroporation cuvettes (0.2 cm gap; Bio-Rad, cat. no. 1652086)
- Electroporation apparatus (Bio-Rad, cat. no. 1652662)
- Petri dishes (Thermo Scientific, cat. no. 249964)
- Rattler plating beads (Zymo Research, cat. no. SI001)
- Spectrophotometer (Biochrom WPA, cat. no. C08000)
- Falcon 50-ml conical centrifuge tubes (Corning, cat. no. 352070)

- AirPore Tape Sheets (Qiagen, cat. no. 19571)
- Round-bottom polystyrene tubes (Corning, cat. no. 14-959-IB)

Emulsion PCR

- Spectrophotometer (Biochrom WPA, model no. C08000)
- 1 ml Syringes (Covidien-Medtronic, cat. no. 1180100555)
- TissueLyser LT (Qiagen, model no. 69980)
- Microman pipette for viscous liquids (Gilson, cat. no. F148504)
- Capillary pistons (Gilson, cat. no. CP 100)
- Inverted Epi-Fluorescence Phase Contrast Microscope (Olympus, model no. 1X51)
- FITC/GFP filter cube (Chroma, cat. no. 41001)
- Cellometer cell counting chambers (Nexcelom, cat. no. SD100)
- High-performance near-infrared charge-coupled device (CCD) camera, IEEE 1394 FireWire (QImaging, model RoleraXR)
- Microscope camera calibration slide (0.01-mm stage micrometer; OMAX, cat. no. CS-A36CALM1)
- ImageJ (<https://imagej.nih.gov>)

Recovery PCR

- Gel electrophoresis equipment (Bio-Rad, cat. nos. 1704486, 1645050)
- Razor blades (VWR, cat. no. 55411-050)
- Blue light transilluminator (Syngene, cat. no. FG-05)
- UV-Vis spectrophotometer (NanoDrop, model no. 2000c)

Additional equipment for recovery of a biotinylated ePCR product

- DynaMag-2 magnet (Thermo Fisher, cat. no. 1232ID)
- Sample mixer (HulaMixer Sample Mixer; Thermo Fisher, cat. no. 15920D)

REAGENT SETUP

TAE, 1x—Dilute 50× TAE concentrate 50-fold with dH₂O. The stock can be stored at room temperature (20–25 °C) for months.

Agarose gel, 1% (wt/vol)—Mix 0.75 g of agarose in 75 ml of TAE buffer and microwave, shaking occasionally, until the agarose dissolves completely. Add 5 µl of ethidium bromide and pour into a gel-casting tray. Incubate until it solidifies (–30 min). If necessary, the solidified gels can be stored in a plastic wrap at 4 °C for a few days.

LB medium—Dissolve 25 g of LB powder in 1 liter of ultrapure H₂O and sterilize by autoclaving. The sterilized stock solution can be stored at room temperature for up to 1 year.

2xYT broth supplemented with 1% (wt/vol) glucose—Dissolve 31 g of 2xYT powder in 1 liter of ultrapure H₂O and sterilize by autoclaving. Prepare a solution of 20% (wt/vol) D-glucose by dissolving 20 g of glucose in 100 ml of ultrapure water and sterilize by filtration (0.2-µm-pore filter). Add 20% (wt/vol) glucose to room temperature 2xYT broth at a ratio of 1:20. The sterilized stock solution can be stored at room temperature for up to 1 year.

SOC medium—To prepare SOB medium, dissolve 28 g of SOB powder in 980 ml of ultrapure water and sterilize by autoclaving. Prepare a solution of 20% (wt/vol) D-glucose by dissolving 20 g of glucose in 100 ml of ultrapure water and sterilize by filtration (0.2-µm-pore filter). To prepare SOC, add 20 ml of 20% (wt/vol) glucose to 980 ml of SOB. The stock can be stored at room temperature or at 4 °C for up to 1 year.

Chloramphenicol, 25 mg/ml, 1,000x—Prepare a 25 mg/ml solution of chloramphenicol in 100% (vol/vol) ethanol. The stock can be stored at −20 °C for years.

Kanamycin, 50 mg/ml, 1,000x—Prepare a 50 mg/ml solution of kanamycin in ultrapure water and sterilize by filtration (0.2-µm-pore filter). The stock can be stored at −20 °C for years.

Carbenicillin, 100 mg/ml, 1,000x—Prepare 100 mg/ml solution of carbenicillin disodium salt in ultrapure water and sterilize by filtration (0.2-µm-pore filter). The stock can be stored at −20 °C for years.

Agar plates—Dissolve 25 g of LB medium powder and 15 g of agar in 1 liter of ultrapure water and sterilize by autoclaving. Cool the solution to −50 °C in a temperature-controlled water bath. Add 1 ml each of appropriate antibiotics. Pour into plates and let solidify. The plates can be stored at 4 °C for up to 1 month.

100 mM 5HTP, 10x—Dissolve 22 mg of 5HTP powder in 10 ml of ultrapure water. The stock can be stored at 4 °C for several months.

IPTG, 1 M—Dissolve 2.38 g of IPTG powder in 10 ml of ultrapure water. Filter sterilize with a 0.2-µm syringe filter. The solution can be stored at −20 °C for years.

TMAC, 100x—Dilute 5 M of TMAC solution 1:500 to prepare 10 mM solution in water. The solution can be stored at −20 °C for years.

Tris-buffered saline 1x—Prepare a 50 mM Tris, 150 mM NaCl solution in ultrapure water and adjust the pH to 7.6 with concentrated HCl. The stock can be stored at room temperature for up to 1 year.

Taq buffer, 10×—Prepare a. 500 mM KCl/100 mM Tris/15 mM MgCl₂ solution in ultrapure water and adjust the pH to 8.3 with concentrated HCl. The stock can be stored at −20 °C for years.

3 M Sodium acetate pH 5.2—Prepare 3 M sodium acetate in ultrapure water. Adjust the pH to 5.2 with glacial acetic acid. The solution can be stored at room temperature for up to 1 year.

2× Binding and washing buffer (2× B&W buffer)—Mix the following components in water and adjust the pH to 7.5 with concentrated HCl. The solution can be stored at room temperature for up to 1 year.

Component	Amount	Final concentration
Tris base	12.1 mg	10 mM
EDTA	2.92 mg	1 mM
NaCl	1.17 g	2 M
H ₂ O	To 10 ml	
Total	10 ml	

1× Binding and washing buffer (1 × B&W buffer)—Dilute 2× B&W buffer in an equal volume of dH₂O. The solution can be stored at room temperature for up to 1 year.

PROCEDURE

GFP assay · TIMING 1 d

1| To assay the newly designed gene circuit, transform B121 (DE3) with an appropriate GFP reporter plasmid (such as pACYC-GFPM2) as well as either a WT (e.g., plasmid pQE-80L-Kan-T7WT) or inactive GOI-encoding plasmid (see ‘Experimental design’). To transform, thaw chemically competent cells (100 µl) on ice and mix with 1–2 µl of each plasmid. Incubate the mixture on ice for 30 min.

▲ CRITICAL STEP The efficiency of transforming two plasmids together is notably lower than single-plasmid transformation. As transformation of freshly ligated circular plasmids is of low efficiency as well, it is advisable to transform the reporter plasmid and the GOI-encoding library pool sequentially via two transformations, as is described in later steps (see ‘Cloning mutant libraries’). However, when transforming two supercoiled plasmids (plasmids prepared by standard miniprep protocols), as in the current step, co-transformation gives adequate results.

▲ CRITICAL STEP For the RNAP selection, use the pACYC-GFP plasmid or a derivative with an orthogonal promoter, as well as either the WT RNAP-encoding plasmid (pQE-WT) or a derivative with an inactive RNAP. For the tRNA selection, use the pACYC. GFPIAmb plasmid as well as a plasmid encoding either a suppressor tryptophanyl tRNA (pRST.UB.AS3.4) or a nonfunctional mutant tRNA-encoding plasmid (see Reagents).

- 2| Heat-shock the cell suspension at 42 °C for 40 s and return the cells to ice for 2–3 min.
- 3| Recover the transformations by mixing them with 400 µl of room-temperature SOC and incubating them at 37 °C with shaking (250 r.p.m.) for 1 h.
- 4| Grow each of the transformations in triplicate in liquid culture as follows: transfer 10–50 µl of recovered cells to 3 ml of 2×YT supplemented with 1% (wt/vol) glucose and appropriate antibiotics, and grow the cells at 37 °C with shaking (250 r.p.m.) overnight.

▲ **CRITICAL STEP** For the T7 RNAP selection, grow the cells in the presence of chloramphenicol and kanamycin. For the tRNA selection, grow the cells in the presence of chloramphenicol and carbenicillin, and 1 mM 5HTP as the unnatural amino acid supplement.

- 5| The next morning, inoculate 20 µl of the overnight culture in 2 ml of fresh 2×YT supplemented with 1% (wt/vol) glucose and appropriate antibiotics, and grow the cultures at 37 °C with constant shaking (250 r.p.m.) until the culture reaches an OD₆₀₀ value of 0.4–0.6 (this takes ~2 h).

? TROUBLESHOOTING

- 6| Induce the cultures with 0.4 mM IPTG (add 0.8 µl of the 1 M IPTG stock) and grow at 37 °C for additional 3 h.
- 7| Transfer 500 µl of the induced culture to a 1.5 ml Eppendorf tube and centrifuge the samples at 3,000g for 1 min at room temperature.
- 8| Gently resuspend the cell pellets in 1 ml PBS and centrifuge at 3,000g for 1 min at room temperature.
- 9| Repeat the washing step and resuspend the cell pellets in 1 ml of PBS.
- 10| Transfer 200 µl of the cell suspension to a black 96-well microplate and measure GFP fluorescence (excitation/emission 481/507 nm) using a microplate reader.
- 11| Transfer another 200 µl of the cell suspension to a clear 96-well microplate and load 3 empty wells with 200 µl of pure PBS for a blank reading. Measure absorption at 600 nm using the Tecan fluorescence microplate reader.
- 12| Analyze the GFP fluorescence data as follows. Average the OD₆₀₀ readings from the 3 blank wells and subtract the average from the readings of cell culture wells. Normalize each of the cell culture fluorescence readings with its corresponding OD₆₀₀ measurement to obtain a relative fluorescence units to OD₆₀₀ value ratio (RFU/OD₆₀₀). Calculate the averages and standard errors of RFU/OD₆₀₀ for each of the triplicates and normalize all the sample values to that of a negative control, as appropriate. For WT T7 RNAP in combination with a GFP polymerase expressed under the WT promoter, we typically obtain signal-to-noise ratios in the range of 100:1 (see ANTICIPATED RESULTS).

? TROUBLESHOOTING

Library transformation and expression • TIMING 4–7d

13| Use a library generation method of choice^{3,4} to prepare a randomized GOI pool and subclone the pool into an appropriate destination vector (at least 1 μ p vector).

▲ CRITICAL STEP For generation of the RNAP library, we advise randomizing the T7 RNAP at 6 positions in the T7 RNAP specificity loop (R746, L747, N748, R756, L757, Q758). Subclone the library (the full gene length) into the empty pQE plasmid using BamHI and HindIII restriction sites.

▲ CRITICAL STEP For the tRNA selection, we advise randomizing the AS3.4 tRNA gene at the anticodon stem, acceptor stem, and the loop regions. Subclone the library into pRST.HB.AS3.4 using KpnI and BsrGI restriction enzymes (replace the existing ScW tRNA AS3.4 with the library).

14| Using a standard protocol^{38,39}, prepare 200 μ l of fresh electrocompetent BL21(DE3) cells carrying an appropriate reporter plasmid from 100 ml of log-phase cells grown in media supplemented with chloramphenicol.

▲ CRITICAL STEP For the RNAP selection, use a BL21(DE3) strain carrying the pACYC-Tag plasmid (or its derivative harboring a mutant promoter of choice). For the tRNA selection, use a strain carrying the pACYC.Tag.lAmb plasmid (or a derivative with an increased number of amber codons, for a more stringent selection).

15| Distribute the freshly prepared competent cells equally into two (0.2-cm gap) electroporation cuvettes and transform each sample with 500 ng (1,000 ng total) of the column-purified library ligation in 5 μ l of ultrapure water. Use the electro-porator to apply 2.5 kV of electrical potential (time constant range = 4.8–5.4) to each cuvette. Immediately after application of the potential, mix the content of each cuvette with 1 ml of room temperature SOC and transfer the cells to a round-bottom polystyrene tube. Recover the transformations in two separate tubes for 1 h at 37 °C, with shaking at 250 r.p.m.

16| Pool the transformants together and plate serial dilutions to determine the library size (Box 1).

▲ CRITICAL STEP For the T7 RNAP selection, use LB agar plates supplemented with 25 μ g/ml chloramphenicol and 50 μ g/ml kanamycin. For the tRNA selection, use LB agar plates supplemented with 25 μ g/ml chloramphenicol and 100 μ g/ml carbenicillin.

17| Transfer the remaining volume of the transformed culture to 20 ml (10 $\times$ volume of recovered cells) 2 $\times$ YT broth (in a 100-ml flask) supplemented with 1% (wt/vol) glucose and appropriate antibiotics and grow the cultures at 37 °C overnight with shaking (250 r.p.m.).

▲ CRITICAL STEP For the T7 RNAP selection, use 25 μ g/ml chloramphenicol and 50 μ g/ml kanamycin for outgrowth. For the tRNA selection, use 25 μ g/ml chloramphenicol and 100 μ g/ml carbenicillin as for outgrowth.

18| The next morning, inoculate 100 μ l of the overnight culture in 10 ml fresh 2 $\times$ YT broth supplemented with 1% (wt/vol) glucose and appropriate antibiotics (and 1 mM 5HTP for the tRNA selection) and seal the opening with an AirPore Tape Sheet (using these sheets instead of the tight caps promotes gas exchange during growth). Grow at 37 °C with shaking (250 r.p.m.) for ~2 h until the exponential phase is reached (OD₆₀₀ value = 0.4–0.6).

19| Induce the culture with 0.4–1 mM IPTG (final concentration) and grow for an additional 1–7 h at 37 °C. The amount of inducer and the length of induction can be optimized to regulate the selection stringency. We recommend starting at 1 mM IPTG and growing for 7 h in the first round of selection.

Emulsion PCR - TIMING 3–5 h

20| Prepare the oil/surfactant mixture in a 2-ml tube as described below and add a rubber tip from a 1 ml syringe to the mixture. Vortex the mixture for 2 s, and let it pre-chill on ice.

Component	Amount (μ l)	Final concentration
Tegosoftware	876	73% (vol/vol)
Mineral oil	240	20% (vol/vol)
Abil WE 09	84	7% (vol/vol)
Total	1200	

▲ CRITICAL STEP Omission of the syringe tip results in inefficient emulsification.

21| Prepare the aqueous phase of the ePCR mix as described below and let it pre-chill on ice.

Component	Amount (μ l)	Final concentration
ddH ₂ O	242.85	
IO $\times$ Taq buffer	30	1x
10 mM dNTP	6	200 μ M
10 μ M Forward ePCR primer	9	300 nM
10 μ M Reverse ePCR primer	9	300 nM
10 mM TMAC	3	100 μ M
RNaseA	0.15	0.5 μ g/ml
Total	300	

▲ CRITICAL STEP Use the biotinylated primer pair AJM.9 and AJM.10 for the selection of T7 RNAP, and the primer pair JWE.44 and JWE.45 for the selection of tRNA (Table 1). Note that these two primer sets carry different modifications to illustrate different purification and recovery strategies later in the protocol (see ‘Recovery’ section).

▲ CRITICAL STEP Note that primers AJM.9 and AJM.10 amplify only the relevant fragment of the T7 RNAP gene, and not the entire gene sequence.

▲ **CRITICAL STEP** Note that this buffer composition was optimized for Tag DNA polymerase, and a different enzyme might require a different optimal buffer composition.

22| Measure the OD₆₀₀ value of the induced culture from Step 19 and centrifuge 100–500 µl of the culture at 1,500# for 15 min at room temperature.

▲ **CRITICAL STEP** Too many cells during emulsification can yield more than one cell per compartment and thereby result in background amplification. Too few cells can give a very low yield of emulsion PCR product, as judged by the recovery PCR yield. We observed that in our experience, the equivalent of 0.5 ml of culture at an OD₆₀₀ value of 1 (roughly 5×10^8 cells) often gives the best results (e.g., 125 µl of a culture at an OD₆₀₀ value = 4).

23| Resuspend the cells in 1,000 µl of Tris-buffered saline (TBS) and centrifuge as in Step 22.

24| Repeat the washing step (Step 23) once.

25| Resuspend the cells in 300 µl of ice-cold 1× ePCR mix from Step 21.

26| Pipette the resuspended cells into the prechilled oil/surfactant mixture from Step 20. At this point, the aqueous phase with cells sinks to the bottom of the tube and remains as a separate phase (Fig. 4a).

27| Mix by flipping the suspension back and forth and proceed to emulsification by vibrating the tubes in the TissueLyser at 42 Hz for 4 min at 4 °C. We routinely use this method and recommend it to the reader.

▲ **CRITICAL STEP** The emulsion should appear viscous, with a milky white color, and no phase separation should be seen (Fig. 4b).

▲ **CRITICAL STEP** The duration and frequency of the vibrations are important parameters that will probably require optimization. The distribution of bacterial cells in emulsion droplets can be checked under a microscope (Box 2, Fig. 4h,i). See ANTICIPATED RESULTS for acceptable emulsion quality.

▲ **CRITICAL STEP** Emulsification can in principle also be performed by vibrating the mixture on a conventional vortex mixer for 3–5 min⁴⁵. However, the authors warn that they have not used this method, and therefore do not have comparative data.

28| Using the Gilson pipette, aliquot 12× 100 µl of the emulsion mixture into PCR tubes. If trapped air bubbles are present, remove them by flicking the tubes gently.

▲ **CRITICAL STEP** Using the Gilson pipette is advisable because it exhibits superior pipetting accuracy and precision with viscous liquids.

29| Perform ePCR with the following thermal cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95 °C, 5 min		
2–6	95 °C, 30 s	55 °C, 30 s	68 °C, 1.5 min/kb
7–26	95 °C, 30 s		68 °C, 1.5 min/kb
27			68 °C, 8 min

▲ **CRITICAL STEP** Cycle numbers 2–6 have a lower annealing temperature to ensure the annealing of the complementary portion of the primers to the primer binding sites on the template. The subsequent 20 cycles (7–26) should favor annealing by the entire primer to correctly polymerized amplicons. We have found that this division minimizes nonspecific primer binding and, therefore, maximizes the yield of the expected product.

■ **PAUSE POINT** The emulsions can be kept at 4 °C for at least a week after ePCR.

Breaking of the emulsions • TIMING 3–4 h

30| After the thermal cycling program is finished, pool the emulsion PCRs together (Fig. 4c) and centrifuge in a 1.5-ml Ep-endorf tube (10 min at 12,000# at room temperature). This step results in the mixture separating into two layers: a top oil layer and a bottom milky white, amorphous solid-like layer that contains the remaining intact emulsion droplets (Fig. 4d). The top oil layer often appears cloudy, as it may contain some small emulsion droplets. These are typically emulsion droplets that contain no cells. They remain in the top oil layer, as they are lower in density.

31| Remove and discard the top oil layer, leaving the white emulsion in the bottom.

32| To break the emulsion, add an equal volume of chloroform/isoamyl alcohol/phenol (24:1:25) and vortex for ~10 s. At this step, the mixture still appears cloudy but exhibits a greatly reduced viscosity. The bottom amorphous solid-like layer should no longer be present (Fig. 4e).

▲ **CRITICAL STEP** The chloroform/isoamyl alcohol/phenol mixture is typically buffered with an aqueous phase to prevent the phenol from oxidizing. Take care to pipette from the bottom organic phase, and not from the aqueous phase.

33| Transfer the mixture to a Phase Lock gel tube (optional) and centrifuge at 16,000# for 2 min at room temperature to separate the phases. If Phase Lock gel tubes are used, the mixture separates into three layers. As the middle gel layer is mostly immobile, it is relatively easy to pipette away the desired top aqueous layer in the next step. If regular tubes are used, the mixture separates into two mobile layers (Fig. 4f), and pipetting the top aqueous layer takes some more effort and skill.

34| Collect the top aqueous phase, transfer to a new tube, and mix with an equal volume of chloroform.

35| Briefly vortex the mixture and transfer to a new Phase Lock tube (optional). Centrifuge the tube again at 16,000# for 2 min at room temperature to separate the phases.

36| Collect the top aqueous phase and transfer to a new Eppendorf tube.

37| Purify the aqueous phase using the QIAquick PCR Purification Kit. ELute the DNA twice with 50 µl of ddH₂O pre-warmed to 60 °C.

▲ CRITICAL STEP Adsorption of DNA on the QIAquick membrane depends on pH: adsorption is 95% efficient if the pH is <7.5, but is markedly reduced at higher pH values (QIAquick Spin Handbook, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjVrczqsXWAhWE0SYKHf_uCUEQFggrMAE&url=https%3A%2F%2Fwww.qiagen.com%2Fus%2Fresources%2Fdownload.aspx%3Fid%3D3987caa6-ef28-4abd-927e-d5759d986658%26lang%3Den&usg=A0vVawl69dkR9Zyza7u3nPqZ6kxs). Binding buffer PBI in the QIAquick kit contains a pH indicator that turns violet at pH >7.5 and yellow at pH <7.5. At this step, we typically observe the mixture turn violet, which is an indication that the pH must be reduced to ensure proper binding of DNA to the membrane. The pH of the mixture can be reduced through the addition of 10 µl of 3 M sodium acetate (pH 5.2).

▲ CRITICAL STEP There is very little DNA amplification that occurs as a result of emulsion PCR alone, and the amplification product is almost never visible by gel electrophoresis. Quantification of DNA concentration using the NanoDrop spectrophotometer at this step also typically gives unreliable results. The user is advised to simply proceed to Step 38 without worrying about low or undetectable DNA yields. However, despite no immediately observed effect, maximizing the recovery of DNA (such as eluting with pre-warmed water) at this step is important.

38| Digest the purified DNA with the DpnI restriction enzyme. For this purpose, mix the eluate with 10 µl of 10× CutSmart Buffer and 2 µl (40 units) of DpnI enzyme. Incubate the reaction for at least 1 h at 37 °C.

▲ CRITICAL STEP This step eliminates carryover of genomic DNA and (more importantly) plasmid DNA. Omission of this step results in nonspecific amplicons accumulating during the recovery PCR step, and lowers the selection efficiency.

■ PAUSE POINT DpnI digestion can also be carried out overnight at 37 °C.

39| Column-purify the DpnI-digested DNA using the QIAquick PCR Purification Kit one more time and elute in 60 µl of ddH₂O pre-warmed to 60 °C.

▲ CRITICAL STEP Two column purifications reduce carryover of emulsion-PCR primers, and reduce background amplification.

■ PAUSE POINT The purified DpnI-digested DNA can be stored at 4 °C for up to 1 week or at -20 °C for years.

ePCR product recovery • TIMING 6–10 h

40 Purify the ePCR product of contaminating plasmid DNA by either recovery of biotinylated products (option A) or recovery of tagged products (option B).

(A) Recovery of biotinylated ePCR product • TIMING 8–10 h—(i) Resuspend settled Dynabeads and transfer 100 µl of the suspension to a new tube.

(ii) Mix the beads with 1,000 µl of 1× B&W buffer.

(iii) Place the tube on a magnet for 1 min and discard the supernatant.

(iv) Resuspend the beads in 500 µl of 1× B&W buffer.

(v) Repeat Step 40A(iii and iv) for a total of 3 washes.

(vi) Place the tube on a magnet for 1 min and discard the supernatant. Resuspend the Dynabeads in 100 µl 2× B&W buffer and mix with 100 µl of DpnI-digested DNA from Step 39 and 100 µl of distilled water. Incubate the mixture for 15 min at room temperature with gentle rotation.

(vii) Place the tube on a magnet for 2–3 min and discard the supernatant.

(viii) Wash the beads with 1× B&W buffer as in Step 40A(iii and iv) for a total of 5 times.

(ix) Set up the following 50-µl recovery PCR mix using 0.5–5 µl of bead-bound DNA as the template and primer pair AJM.15 and AJM.16 (Table 2):

Component	Amount (µl)	Final concentration
ddH ₂ O	To 100	
5× HF buffer	20	1×
10 mM dNTP	2	200 µM
10 µM Forward primer	5	500 nM
10 µM Reverse primer	5	500 nM
Template (ePCR)	1–10	
2,000 U/ml Phusion	1	20 U/ml
Total	100	

▲ CRITICAL STEP We recommend optimizing the volume of added template,

(x) Perform a PCR using the thermal cycling conditions shown in the table below:

Cycle number	Denature	Extend
1	98 °C, 30 s	
2–26	98 °C, 5 s	72 °C, 1 min/kb

Cycle number	Denature	Extend
27		72 °C, 8 min

▲ **CRITICAL STEP** We recommend optimizing the number of thermal cycles.

(xi) Separate the recovery PCR products by gel electrophoresis on a 1% (wt/vol) agarose gel and excise the band of the correct size using a fresh razor blade. Extract DNA using the QIAquick Gel Extraction Kit.

▲ **CRITICAL STEP** A small nonspecific amplification product is sometimes observed in addition to the band of the correct size. Whether the additional band is observed or not, gel purification of the correct size is important, as it minimizes the carryover of these smaller elements to the next amplification step and/or CPR round.

? TROUBLESHOOTING

■ **PAUSE POINT** The purified DNA can be stored at 4 °C for up to a week or at –20 °C for years.

(xii) Set up overlap extension PCR using appropriate DNA fragments and primers to obtain a full T7 RNAP gene sequence (Supplementary Note; Supplementary Fig. 2; Supplementary Table 1). This step serves two purposes: (i) it assembles the full-length T7 RNAP coding sequence for digestion and ligation into the empty vector; and (ii) it yields sufficient PCR product for continuing to the next round of CPR (Step 41; 100–1,000 ng is needed). Use 1–5 µl of the purified recovery PCR product.

(xiii) Check the size of the amplicon on a 1% (wt/vol) agarose gel. Purify the DNA using a PCR-purification kit per manufacturer's instructions. Alternatively, if additional PCR bands are observed, perform gel purification as previously described.

■ **PAUSE POINT** The purified DNA can be stored at 4 °C for up to a week or at –20 °C for years.

(B) Recovery of tagged ePCR product • TIMING 6–9 h—(i) Set up the following 50-µl recovery PCR mix using 0.5–5 µl of DpnI-digested, column-purified sample from Step 39 as the template and primer pair JWE.46 and JWE.47 (Table 2):

Component	Amount (µl)	Final concentration
ddH ₂ O	To 100	
5× HF buffer	20	1x
10 mM dNTP	2	200 µM
10 µM Forward primer	5	500 nM
10 µM Reverse primer	5	500 nM
Template (ePCR)	1–10	

Component	Amount (μl)	Final concentration
2,000 U/ml Phusion	1	20 U/ml
Total	100	

▲ **CRITICAL STEP** We recommend optimizing the volume of added template.

(ii) Perform a PCR using the PCR conditions shown in the table below:

Cycle number	Denature	Extend
1	98 °C, 30 s	
2–26	98 °C, 5 s	72 °C, 1 min/kb
27		72 °C, 8 min

▲ **CRITICAL STEP** We recommend optimizing the number of thermal cycles.

(iii) Separate the recovery PCR products by gel electrophoresis on a 1% (wt/vol) agarose gel and cut out the band of the correct size. Extract the DNA using the QIAquick Gel Extraction Kit.

▲ **CRITICAL STEP** A small nonspecific amplification product is sometimes observed in addition to the band of the correct size. Whether the additional band is observed or not, gel purification of the correct size is important, as it minimizes the carryover of these smaller parasitic elements to the next amplification step or CPR round.

? TROUBLESHOOTING

(iv) Quantify the concentration of the amplified DNA product using the NanoDrop spectrophotometer.

▲ **CRITICAL STEP** Depending on the desired scale of molecular cloning, 100–1,000 ng of purified insert DNA may be required for the subsequent recloning (Step 41). If a sufficient amount of DNA product is obtained, continue to Step 41. If not enough DNA product is obtained, continue to Step 40B(v).

■ **PAUSE POINT** The purified DNA can be stored at 4 °C for up to 1 week or at –20 °C for years.

(v) (Optional) Perform a third, re-amp PCR using the recovery PCR product as the template and nested primers of choice to obtain enough DNA product to continue on to the next round of selection. Set up a reaction as in Step 40(i) and perform the thermal cycling program below. Column-purify the product.

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
2–26	98 °C, 5 s	(T _m + 3) °C, 30 s	72 °C, 1 min/kb
27			72 °C, 8 min

T_m, primer melting temperature.

■ **PAUSE POINT** The purified DNA can be stored at 4 °C for up to 1 week or at –20 °C for years.

Repetition of CPR rounds • TIMING 5–10 weeks

41| Continue to the next round of selection by digesting the purified PCR product with appropriate restriction enzymes and ligating as in Step 13. The molecular cloning reactions can be scaled down 10× or more in later rounds of a CPR selection, as compared with the initial library generation.

42| Repeat the CPR cycles with increasing selection stringency until enrichment levels off or (as observed by GFP assay) it is determined that further randomization or alternative selection conditions are necessary (see ANTICIPATED RESULTS). To increase selection pressure, it is easiest to modulate the time of induction (decrease time to as short as 1 h), temperature of induction (increase), IPTG concentration (decrease down to 0.1 mM), or other aspects of the genetic circuit that impact Tag polymerase production. For example, in the case of amber suppression, the number of amber codons in the Tag DNA polymerase open reading frame can be increased. To minimize accumulation of nonspecific amplification products in later rounds of CPR, change the sequences of the portions of primers that do not bind to the primer binding sites on the initial template.

This can be done for ePCR, recovery PCR, and/or re-amp PCR as necessary.

? TROUBLESHOOTING

? TROUBLESHOOTING

In certain cases, the user may encounter complications in some steps of the PROCEDURE, such as poor *coli* growth, poor dynamic range in GOI or DNA polymerase expression, poor ePCR or recovery PCR yields, or poor enrichment efficiencies. Probable causes of unsuccessful results and proposed solutions are summarized in Table 3. In brief, if the user encounters poor bacterial growth upon induction (Step 5), and/or poor dynamic range in GFP expression (with the GFP reporter; Step 12), then optimization of GOI expression might be required. To this end, promoter, terminator, or ribosome binding site strengths can be varied, as can GOI induction conditions such as induction temperature and/or inducer concentration. Poor GOI amplification (for example, of a positive control; Step 40) in ePCR is another common problem in unoptimized ePCR experiments. This may be due to poor primer design (inefficient hybridization to the GOI) and/or to the accumulation of smaller, nonspecific PCR amplification products. Such PCR ‘parasites’ often arise from random

recombination, deletion, and off-target amplification and can unfortunately carry over and preferentially accumulate from round to round, as smaller DNA molecules are generally preferred in both PCR amplification and in *E. coli* transformation. To minimize carryover and accumulation, it is advisable to design primers to amplify the smallest GOI fragment possible, to use primers with minimal cross-reactivity with other portions of the plasmid or genomic DNA, and to use PCR conditions that favor target amplification (low primer concentrations, few temperature cycles). If smaller products begin to be observed on gels at the numerous quality control checks laid out, then it is advisable to change the sequence of the primer and thereby provide an evolutionary barrier to parasitic amplicons.

The other major issue that may arise in CPR selections is selection coefficients that are too low to efficiently winnow the GOI library (Step 42). Low selection coefficients should initially be observed during the recommended mock selection, but may also become manifest once the GOI library is introduced. Inefficient selection may be due to poor Taq polymerase expression, which would again require optimization of either GOI expression, as above, or better coupling of GOI and Taq polymerase expression through circuit redesign. It is also possible that selection coefficients are too low because too much Taq polymerase is being expressed in the absence of the GOI; for example, in the tRNA selection that we describe it proved useful to ensure that background suppression by endogenous tRNAs did not lead to Taq expression.

It is also possible that selections fail to proceed (Step 42) because of poor library design. It is tempting to create extremely large libraries with the expectation that deep scans of sequence space are more likely to yield novel variants. Although this may be true, the fact that most mutants are deleterious or neutral means that the larger the library, the more difficult the amplification, to the point that recovery of active variants against a background of inactive variants becomes extremely challenging. Conversely, if a library is too conservative, then there may be few highly active variants to be selected. A dearth of active variants may be observed by sequencing but can often be more simply monitored by comparing the band intensity of the secondary PCR between rounds of selection. If the band observed after a secondary PCR does not increase in subsequent rounds, it may prove useful to impart diversity between rounds of selection, for example, by error-prone amplification. As users gain skill with CPR the balance between selection and diversity can be tuned on the fly. For example, if a particular selection pressure results in poor secondary amplification and a lack of enrichment, it may be necessary to add diversity before increasing selection pressure.

• TIMING

Steps 1–12, GFP assay: 1 d

Steps 13–19, Library transformation and expression: 4–7 d

Steps 20–29, emulsion PCR: 3–5 h

Steps 30–39, breaking of the emulsions: 3–4 h

Step 40A, recovery of biotinylated ePCR product: 8–10 h

Step 40B, recovery of tagged ePCR product: 6–9 h

Steps 41 and 42, repetition of CPR rounds: 5–10 weeks

Box 1, dilution plating to estimate Library size: 1 d

Box 2, direct visualization of emulsion for rough estimation of droplet size: 1–2 h

Box 2, direct visualization of emulsified GFP-expressing *E. coli* cells for estimation of cell distribution in emulsion droplets: 1–2 h

ANTICIPATED RESULTS

The emulsions generated using this protocol should be able to withstand thermal cycles of PCR. As a reference, we provide a photograph of emulsions (Fig. 4g) that differ in stability due to various emulsification conditions, such as various ratios of oil/aqueous phase volumes. The stability can be checked visually after 30 cycles of PCR: the emulsion should appear milky and not separated into phases. Emulsion droplets should be relatively uniform in size and ~2–7 μm in diameter (Fig. 4h). Finally, no more than one bacterial cell should be entrapped in a single emulsion droplet. To this end, make sure that no more than 20% of emulsion droplets contain a GFP-expressing bacterial cell, as in the example shown (Fig. 4i).

A mock selection result should demonstrate 10^2 – 10^5 -fold enrichment of active GOI from a mixture with inactive GOI, as judged by restriction digestion or sequencing. As an example, we provide a gel image of a restriction digestion of amplicons that emerged during the selection of an active tRNA synthetase (uncleaved) mixed with its inactive variant (cleaved; Fig. 4j). Besides estimating enrichment by the mock selection experiment, the user may also roughly track enrichment during the directed evolution campaign by estimating the yields of recovery PCR with each consecutive round, as a higher fraction of active variants in the pool usually provides higher yields of recovery PCR (Fig. 4k).

To further provide the reader with a practical example of what to expect during the course of a CPR selection, Figure 4l shows results of one of the experiments described herein, the selection of T7 RNAP variants that recognized an orthogonal promoter over multiple cycles²⁸. Starting with 2×10^7 variants of T7 RNAP randomized at 6 amino acid residues in the specificity loop²⁸, 7 rounds of CPR yielded an overall 2.4-fold increase over the naive library in GFP expression from the P_{CGG} promoter, ultimately producing a single variant that had only 2% of the activity of WT T7 RNAP on the WT P_{T7} promoter. Per our Troubleshooting instructions, we reasoned that additional genetic diversity was required to obtain higher activity at the P_{CGG} promoter. Accordingly, we used the single variant from R7 to introduce additional diversity in a larger region (500 bp) of the specificity loop by means of error-prone PCR and subjected the pool to 5 more rounds of CPR. Introduction of additional genetic diversity via error-prone PCR at round 7 initially decreased the fitness of the pool, but the additional mutations then led to continued improvement in function through round 12. *E. coli* cells were transformed with the pool from round 12 carrying P_{CGG} -driven GFP plasmid, and the 8 brightest clones were sequenced. The most active variant from this subset exhibited 20% of the activity of WT T7 RNAP on the WT P_{T7} promoter. Three other

mutations that were frequently observed in the R12 pool were added to the most active variant in combinations and assayed for GFP expression. The variant with highest activity demonstrated 40–60% of the activity of WT T7 RNAP on the WT P_{T7} promoter²⁸.

These results extend the admonition provided in the Troubleshooting section, that a combination of focused and random mutagenesis might be required throughout the course of a selection to achieve the best results. Although we have advised that the initial library design should rely on structural data that identify, say, the active site or substrate-binding residues, traversing a GOI's fitness landscape may ultimately involve many additional residues, and it is possible that a focused pool can become trapped in a local maximum, as in the example above. Opening up the library to additional random mutagenesis or to recombination via DNA shuffling might, therefore, be required to navigate away from the local maxima to find a higher maximum. Conversely, higher mutation rates can lead to the accumulation of deleterious mutations in parallel with advantageous ones and it may be difficult to deconvolute the mutational 'load.' Therefore, at the conclusion of a selection, it is often worthwhile to use sequencing data to generate a consensus clone or to rationally add and subtract substitutions from the most active variant of the pool, as was also observed during the T7 RNAP selection.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the Welch Foundation (F-1654 to A.D.E.), the DOD Air Force Research Laboratory (FA9550-14-1-0089), Firebird Biomolecular Sciences (1R41GM119434-01A1), and the John Templeton Foundation (54466). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

References

- 1Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution Nat. Rev. Mol. Cell Biol. 2009; 10:866–876. [PubMed: 19935669]
- 2Packer MS, Liu DR. Methods for the directed evolution of proteins Nat. Rev. Genet. 2015; 16:379–394. [PubMed: 26055155]
- 3Gillam EMJ, Copp JN, & Ackerley DF. Directed Evolution Library Creation: Methods and Protocols 2nd. edn (Humana Press , 2014).
- 4Tee KL, Wong TS. Polishing the craft of genetic diversity creation in directed evolution Biotechnol. Adv. 2013; 31:1707–1721. [PubMed: 24012599]
- 5Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries Nat. Biotechnol. 1997; 15:553–557. [PubMed: 9181578]
- 6Bessette PH, Rice JJ, Daugherty PS. Rapid isolation of high-affinity protein binding peptides using bacterial display Protein Eng. Des. Sel. 2004; 17:731–739. [PubMed: 15531628]
- 7McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains Nature. 1990; 348:552–554. [PubMed: 2247164]
- 8Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display Proc. Natl. Acad. Sci. USA. 1997; 94:4937–4942. [PubMed: 9144168]
- 9Wilson DS, Keefe AD, Szostak JW. The use of mRNA display to select high-affinity protein-binding peptides Proc. Natl. Acad. Sci. USA. 2001; 98:3750–3755. [PubMed: 11274392]

- 10Leemhuis H, Stein V, Griffiths AD, Hollfelder F. New genotype-phenotype linkages for directed evolution of functional proteins *Curr. Opin. Struct. Biol.* 2005; 15:472–478. [PubMed: 16043338]
- 11Amstutz P. In vitro selection for catalytic activity with ribosome display *J. Am. Chem. Soc.* 2002; 124:9396–9403. [PubMed: 12167034]
- 12Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold *Nature*. 2007; 448:828–831. [PubMed: 17700701]
- 13Cesaro-Tadic S. Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library *Nat. Biotechnol.* 2003; 21:679–685. [PubMed: 12754520]
- 14Palmer AC, Kishony R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance *Nat. Rev. Genet.* 2013; 14:243–248. [PubMed: 23419278]
- 15Dietrich JA, McKee AE, Keasling JD. High-throughput metabolic engineering: advances in small-molecule screening and selection *Annu. Rev. Biochem.* 2010; 79:563–590. [PubMed: 20367033]
- 16Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution *Nat. Biotechnol.* 1998; 16:652–656. [PubMed: 9661199]
- 17Ghadessy FJ, Ong JL, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication *Proc. Natl. Acad. Sci. USA.* 2001; 98:4552–4557. [PubMed: 11274352]
- 18Ghadessy FJ, Holliger P. Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes *Methods Mol. Biol.* 2007; 352:237–248. [PubMed: 17041269]
- 19Loakes D, Gallego J, Pinheiro VB, Kool ET, Holliger P. Evolving a 33. polymerase for hydrophobic base analogues *J. Am. Chem. Soc.* 2009; 131:14827–14837. [PubMed: 19778048]
- 20Ramsay N. CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase *J. Am. Chem. Soc.* 2010; 132:5096–5104. [PubMed: 20235594]
- 21d'Abbadie M. Molecular breeding of polymerases for amplification of 36. ancient DNA *Nat. Biotechnol.* 2007; 25:939–943. [PubMed: 17632524]
- 22Ellefson JW. Synthetic evolutionary origin of a proofreading reverse 37. transcriptase *Science*. 2016; 352:1590–1593. [PubMed: 27339990]
- 23Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik DS. High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments *Chem. Biol.* 2005; 12:1281–1289. [PubMed: 16356845]
- 24Mastrobattista E. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting 40. of double emulsions *Chem. Biol.* 2005; 12:1291–1300. [PubMed: 16356846]
- 25Davies D. Cell separations by flow cytometry *Methods Mol. Biol.* 2012; 878:185–199. [PubMed: 22674134]
- 26Colin PY, Zinchenko A, Hollfelder F. Enzyme engineering in biomimetic compartments *Curr. Opin. Struct. Biol.* 2015; 33:42–51. [PubMed: 26311177]
- 27Meyer AJ, Ellefson JW, Ellington AD. Directed evolution of a panel of orthogonal T7 RNA polymerase variants for in vivo or in vitro synthetic 44. circuitry *ACS Synth. Biol.* 2015; 4:1070–1076. [PubMed: 25279711]
- 28Ellefson JW. Directed evolution of genetic parts and circuits by compartmentalized partnered replication *Nat. Biotechnol.* 2014; 32:97–101. [PubMed: 24185096]
- 29Maranhao AC, Ellington AD. Evolving orthogonal suppressor tRNAs to incorporate modified amino acids *ACS Synth. Biol.* 2017; 6:108–119. [PubMed: 27600875]
- 30Yang G, Withers SG. Ultrahigh-throughput FACS-based screening for directed enzyme evolution *Chembiochem.* 2009; 10:2704–2715. [PubMed: 19780076]
- 31van Rossum T, Kengen SW, van der Oost J. Reporter-based screening 49. and selection of enzymes *FEBS J.* 2013; 280:2979–2996. [PubMed: 23578225]
- 32Leung DW, Chen E, Goeddel DV. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction Technique. 1989; 1:11–15.
- 33Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis *PCR Methods Appl.* 1992; 2:28–33. [PubMed: 1490172]

- 34Hutchison CA III. Mutagenesis at a specific position in a DNA sequence J. Biol. Chem. 1978; 253:6551–6560. [PubMed: 681366]
- 35Green MR, Sambrook J, & Sambrook J. Molecular Cloning: A Laboratory Manual 4th. edn (Cold Spring Harbor Laboratory Press , 2012).
- 36Gibson DG. Enzymatic assembly of DNA molecules up to several hundred kilobases Nat. Methods. 2009; 6:343–345. [PubMed: 19363495]
- 37Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability PLoS One. 2008; 3:e3647. [PubMed: 18985154]
- 38Miller EM, Nickoloff JA. Escherichia coli electrotransformation Methods Mol. Biol. 1995; 47:105–113. [PubMed: 7550724]
- 39Sambrook J, Russell DW. Transformation of E. coli by electroporation CSH Protoc. 2006
- 40Diehl F. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions Nat. Methods. 2006; 3:551–559. [PubMed: 16791214]
- 41Williams R. Amplification of complex gene libraries by emulsion PCR Nat. Methods. 2006; 3:545–550. [PubMed: 16791213]
- 42Nakano M. Single-molecule PCR using water-in-oil emulsion J. Biotechnol. 2003; 102:117–124. [PubMed: 12697388]
- 43McDonald JC. Fabrication of microfluidic systems in poly(dimethylsiloxane) Electrophoresis. 2000; 21:27–40. [PubMed: 10634468]
- 44Mazutis L. Single-cell analysis and sorting using droplet-based microfluidics Nat. Protoc. 2013; 8:870–891. [PubMed: 23558786]
- 45Schutze T. A streamlined protocol for emulsion polymerase chain reaction and subsequent purification Anal. Biochem. 2011; 410:155–157. [PubMed: 21111698]
- 46Kelly JR. Measuring the activity of BioBrick promoters using an in vivo reference standard J. Biol. Eng. 2009; 3:4. [PubMed: 19298678]
- 47Davis JH, Rubin AJ, Sauer RT. Design, construction and characterization of a set of insulated bacterial promoters Nucleic Acids Res. 2011; 39:1131–1141. [PubMed: 20843779]
- 48Bonde MT. Predictable tuning of protein expression in bacteria Nat. Methods. 2016; 13:233–236. [PubMed: 26752768]
- 49Mikula RJ. Emulsion characterization Adv. Chem. Ser. 1992; 231:79–129.

Box 1 |**Dilution plating to estimate library size · TIMING 1 d**

This box provides a step-by-step description of the procedures for dilution plating in order to estimate the library size.

1. After outgrowth in Step 16, transfer 1/1,000 volume of recovered culture to a new tube with SOC to a final volume of 100 μ l (this initial dilution is labeled as the 10^3 -fold dilution).
2. From here, make 100- μ l 10-fold serial dilutions in SOC to make 10^4 -, 10^5 -, 10^6 -, and 10^7 -fold dilutions.
3. On prewarmed LB plates with appropriate antibiotics, spread the entire volume (90–100 μ l) of each dilution. Incubate the plates in a 37 °C incubator overnight.
4. The next morning, count the colonies on a plate that has 50–500 colonies, and multiply the number by the dilution factor to get the library size. For example, imagine you get 300 colonies on the 10^5 -fold dilution plate. This means the library size is $300 \times 10^5 = 3 \times 10^7$ colony-forming units.

Box 2 |**Analysis of emulsion droplet quality using microscopy**

Microscopy is a commonly used emulsion characterization tool and is widely used for estimation of emulsion droplet sizes⁴⁹. Optical microscopy involves imaging a thin layer of the emulsion via the use of transmitted light, reflected light, polarized light, fluorescence, or confocal microscopy. Although sample preparation is relatively simple (the emulsion is simply smeared on a slide), care must be taken to avoid perturbing the already unstable emulsion. Electron microscopy (EM) techniques, such as scanning electron microscopy, are superior to optical microscopy techniques in that they offer increased imaging depth and resolution. In addition, EM involves fast freezing of the sample, thus preservation of the morphology of the emulsion droplets. However, sample handling and preparation are more difficult compared with those of optical microscopy, and much more sophisticated instrumentation is required. We will therefore outline a brief procedure for visualization of emulsion droplets via optical microscopy, which is sufficient for a rough estimation of emulsion quality.

Direct visualization of emulsion for rough estimation of droplet size • TIMING 1–2 h

1. Gently dilute the emulsion from Step 27 1:20–1:50 in mineral oil and transfer 20 µl of the dilution to a Cellometer cell counting chamber.
2. Visualize the sample under an epi-fluorescence phase-contrast microscope using 40× magnification and a phase-contrast imaging setting. Take snapshot images using a CCD camera.

▲ **CRITICAL STEP** As the emulsion is unstable upon contact with the cover glass material, images should be taken only in the initial 5 min after sample addition to the chamber. With time, emulsion droplets will coalesce to form larger droplets and the sample will no longer be representative of the original emulsion.

3. Using the same imaging conditions, visualize a 0.01-mm stage micrometer for size calibration. Take snapshot images using a CCD camera.
4. Analyze the images using Image! (<https://imagej.nih.gov>). For guidelines on acceptable emulsion quality, see ANTICIPATED RESULTS.

Direct visualization of emulsified GFP-expressing *E. coli* cells for estimation of cell distribution in emulsion droplets**• TIMING 1–2 h**

1. Transfer 20 µl of undiluted emulsion from Step 27 (given that GFP-expressing cells were emulsified) to a Cellometer cell counting chamber.
2. Visualize the sample under an epi-fluorescence phase-contrast microscope using 40× magnification and the FITC/GFP excitation-emission filter. Take snapshot images using a CCD camera.
3. Analyze the images using Image! (<https://imagej.nih.gov>).

▲ **CRITICAL STEP** To avoid emulsification of multiple *E. coli* cells per emulsion droplet, it is important that no more than 20% of emulsion droplets contain GFP-expressing *E. coli* cells, as observed by optical microscopy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1 |.

General CPR concept, **(a)** Schematic of CPR principle. A gene circuit is generated in which a partner-gene activity allows the expression of a DNA polymerase in bacterial cells.

Inactive gene variants lead to no expression of the DNA polymerase. The genetic circuit containing the diversified partner-gene pool is expressed *in vivo*, allowing DNA polymerase production only in cells with active partner-gene variants. The live cells are subsequently emulsified to produce no more than a single cell per emulsion droplet. The initial boiling step of the ePCR lyses the cells, releasing the produced DNA polymerase protein as well as the partner-gene-encoding plasmid into the aqueous solution of the emulsion droplet.

Ensuing thermal cycling amplifies only the active partner-gene variant, which is recovered and used in the next round of CPR selection, **(b)** Examples of genetic circuits for CPR. Partner-gene function can be linked to expression of a DNA polymerase in a number of ways, as is demonstrated by the example of T7 RNAP, pol, polymerase; tRNA synthetase, and tRNA engineering.

Figure 2 |.

Overview and time line of experiments. A CPR procedure encompasses bacterial expression of the DNA polymerase and partner genes for several hours, emulsification of the bacterial cells in water-oil droplets, *in vitro* amplification of active partner genes via ePCR, breaking of the emulsions, purification of amplified DNA from cell debris and PCR primers, purification of amplified DNA from background plasmid DNA by DpnI digestion, and final recovery and bulk amplification of the partner gene using an additional recovery PCR. A single round of CPR typically takes 3–5 d, and the procedure can be repeated 3–10 times until no more enrichment is observed.

Figure 3 |.

Schematic of the two Recovery strategies, **(a)** Recovery of biotinylated ePCR product. 5'-biotinylated PCR primers are used for amplification by ePCR. The biotinylated ePCR product is later recovered from contaminating plasmid DNA using streptavidin-coated magnetic beads. The ePCR product is further amplified via nested PCR to obtain enough product to continue with the next round of selection, **(b)** Recovery of tagged ePCR product. PCR primers with unique 5'-sequence tags are used for amplification by ePCR. The tagged ePCR product is later recovered from contaminating plasmid DNA by amplification using primers that bind to the unique tag. Finally, internal primers can be used in a final nested PCR to obtain enough product to continue with the next round of CPR selection.

emulsion. Scale bar, 10 μm . **(i)** Superimposed GFP fluorescence/phase-contrast microscopy image of emulsified GFP-expressing DH10B(DE3) *E. coli* cells under 40 $\times$ magnification. Scale bar, 10 μm . **(j,k)** Example of mock selection data. *E. coli* expressing either wild-type tyrosyl-tRNA synthetase from *Methanococcus jannaschii* (MjYRS) or its nonfunctional variant (containing a stop codon and a *NotI* restriction site) were mixed at the indicated ratios and subjected to a single round of ePCR. **(j)** After the mock selection samples were amplified by re-amp PCR and equal amounts of DNA were restriction-digested by *NotI*, the DNA fragments were analyzed by gel electrophoresis to distinguish active (uncut) from inactive (cut) variants of MjYRS. Several thousandfold enrichment of active enzyme variant is observed. Star, active variant fragment size; arrows, inactive variant fragment sizes. Adapted with permission from ref. 29, American Chemical Society, **(k)** Gel-electrophoresis image of recovery PCR. 1: pure active MjYRS amplicon; 0: pure inactive MjYRS amplicon; 10^{-1} - 10^{-4} : amplicons of active/inactive MjYRS dilutions. **(l)** Monitoring enrichment progress by GFP assay. BL21 *E. coli* cells carrying pACYC-GFPmut2 plasmid (in which P_{T7} drives GFP expression) and plasmid ligations from the initial T7 RNAP selection rounds were assayed in a microplate reader for GFP fluorescence. XX, negative control T7 RNAP with two premature stop codons; WT, parental T7-RSS plasmid reported; R0, naive library; R1–R12, the output for each subsequent round during the selections for use of PT7; CGG-R7–8, a single clone from round 7 (this mutant was subject to error-prone PCR, yielding CGG-R7 epPCR); CGG-R12-KI, a single clone from R12; other CGG-R12 variants are selected combinations of mutations seen in the round 12 population. Data represent averages of three independently grown samples. Error bars represent 1 s.d. Adapted with permission from ref. 28, Nature Publishing Group.

TABLE 1 |

Primers for ePCR.

Primer	Sequence 5'–3'	Purpose
AJM.9 (F)	/5 BioSG/GACGAGAATACAGG AGAAATCAGTGAAAAAGTTA AACTGGGAA CAAAAGCCCTAGCGG	ePCR T7 RNAP
AJM.10 (R)	/5 BioSG/GTATCGTTCCGAAGCTATCA TGTATCAGCGCGAATGATTCAATAC CATACTTTTCATGTG	ePCR T7 RNAP
JWE.44 (F)	CCAAATCAGAAAGAACTCAACATCG TCACGTAGGCGCCGCCCTATACCTTGTCT	ePCRtRNA
JWE.45 (R)	CGTCTACCCGAAGCCTCTACTATGAT CAGTTGTTCCACAGGGTAGCCAGCAG CATC	ePCRtRNA

F, forward; R, reverse.

TABLE 2 |

Primers for Recovery PCR.

Primer	Sequence 5'–3'	Purpose
Rec-For B (AJM.15)	GATACAATACAACCAG CGATAGACTCAGG	Recovery T7 RNAP
Rec-Rev B (AJM.16)	CAATACCATACTTTTCA TGTGCCCAGACCACTG	Recovery T7 RNAP
Rec-For A (JWE.46)	CCAAATCAGAAAGAA ACTCAACATCGTCAC	Recovery tRNA
Rec-Rev A (JWE.47)	CGTCTACCCGAAGCCTC TACTATGATCAGT	Recovery tRNA

TABLE 3 |

Troubleshooting table.

Step	Problem	Possible reasons	Solution
5	No or poor growth	Protein is toxic; examine by growth assay	<ul style="list-style-type: none"> • Reduce IPTG concentration • Optimize codon usage • Reduce promoter and/or ribosome-binding site (RBS) strength
12	No or low GFP signal for positive control (WT GOI with cognate substrate)	Gene of interest is poorly expressed (examine by SDS-PAGE or western blot-ting) or gene of interest is toxic (examine by growth assay)	<p>Regardless of the possible reason, consider modulating the following parameters:</p> <ul style="list-style-type: none"> • Optimize the codon sequence of the gene • Optimize IPTG concentration • Optimize temperature of expression • Optimize promoter and/or RBS strength
40A(xi) and 40B(iii)	No amplification	(i) Poor protein expression (ii) Too few cells used for emulsion PCR (iii) Too little ePCR amplicon was used as a template for recovery PCR (iv) Too few cycles/too short extension time used in emulsion/recovery PCR (v) DNA polymerase is trapped in emulsion droplets	(i) Optimize induction conditions: inducer concentration, induction temperature/time (ii) Increase the number of cells used for emulsion PCR (iii) Increase the amount of template used for recovery PCR (iv) Increase number of cycles/extension time for PCR (v) Try supplementing the emulsion PCR reaction with 1 mg/ml BSA
	Amplification of multiple bands	(i) Partner gene plasmid copy number is too high or protein expression is too high (ii) Primer annealing temperature was set too low (iii) Too many PCR cycles were used (iv) Primer concentration was too high (v) Nonspecific primer annealing	<p>For i-v, gel-purify the correct sized band. Additional solutions:</p> (i) Reduce the copy number of the plasmid; optimize the promoter strength ^{46,47} , RBS strength ⁴⁸ , or induction conditions (ii) Increase annealing temperature (iii) Reduce the number of thermal cycles (iv) Reduce primer concentration (v) Check primer sequences for homo/heterodimer formation and repetitions in the template. Try a few different primer pairs
	Amplification of a wrong size (usually smaller size, and especially at later rounds)	(i) Partner gene plasmid copy number is too high or protein expression is too high (ii) Too many cells used in ePCR (iii) Primer annealing temperature set too low (iv) Elongation step of PCR was too short (v) The size of amplified GOI fragment is long (>1 kb) (vi) Accumulation of nonspecific amplification products in later rounds due to nonspecific recombinations/deletions	(i) Use a lower-copy-number plasmid or optimize protein expression (ii) Use fewer cells (<5 × 10 ⁸) (iii) Use a higher annealing temperature. For templates with high GC content, use Phusion GC buffer and 3–6% (vol/vol) DMSO. Check primer sequences for repetitions in the template; optimizing primer pairs can help (iv) Use at least 1.5 min/kb in ePCR and 1 min/kb in recovery PCR steps (v) If possible, design amplicon size to be shorter (vi) Switch PCR primers, if possible. Amplify a smaller GOI fragment, if possible
	Mock selection results in no/poor WT variant enrichment	(i) DNA polymerase expression is leaky (ii) Too many cells were used in ePCR step (iii) DpnI digestion failed (iv) Emulsion quality is suboptimal	(i) Optimize DNA polymerase expression by modulating promoter/RBS strength. Use a strong transcription terminator after each gene in the plasmid. Do not omit 1% glucose in the 2×YT broth. Use a weak protein degradation tag at the C-terminus of the DNA polymerase coding sequence (ii) Use fewer cells (<5×10 ⁸) (iii) Use a fresh tube of DpnI enzyme (iv) Optimize emulsification (vibration frequency, time) for stability (emulsion must remain intact after PCR) or optimize emulsification for droplet size (average size must be 2–7 μlτl)
42	No enrichment of functional variants	(i) Selection pressure was not strong enough or the selection pressure was too strong (ii) Library quality is suboptimal	(i) Start with lower selection pressure and increase gradually by modulating time of expression (reduce) and/or concentration of IPTG (decrease) (ii) Redesign library to enrich for functional mutants ³