Engineering TNA polymerases through iterative cycles of directed evolution

Eric J. Yik[a,1], **Victoria A. Maola**[a,1], **and John C. Chaput**[a,b,c,d,*]
[a]Department of Pharmaceutical Sciences, University of California, Irvine, CA, United States
[b]Department of Chemistry, University of California, Irvine, CA, United States
[c]Department of Molecular Biology and Biochemistry, University of California, CA, United States
[d]Department of Chemical and Biomolecular Engineering, University of California, Irvine, CA, United States
*Corresponding author at: e-mail address: jchaput@uci.edu

Contents

Abstract

DNA polymerases are important tools for biotechnology, synthetic biology, and chemical biology as they are routinely used to amplify and edit genetic information. However, natural polymerases do not recognize artificial genetic polymers (also known as xeno-nucleic acids or XNAs) with unique sugar-phosphate backbone structures. Directed evolution offers a possible solution to this problem by facilitating the discovery of engineered versions of natural polymerases that can copy genetic information back and forth between DNA and XNA. Here we report a directed

[1] Signifies co-first author status.

Methods in Enzymology
ISSN 0076-6879, https://doi.org/10.1016/bs.mie.2023.04.014

evolution strategy for discovering polymerases that can synthesize threose nucleic acid (TNA) on DNA templates. The workflow involves library generation and expression in *E. coli*, high-throughput microfluidics-based screening of uniform water-in-oil droplets, plasmid recovery, secondary screening, and library regeneration. This technique is sufficiently general that it could be applied to a wide range of problems involving DNA modifying enzymes.

1. Introduction

Directed evolution is a protein engineering technique that is commonly used to change the substrate specificity of an enzyme or to introduce new physical or chemical properties into a protein scaffold (Zeymer & Hilvert, 2018). The process mirrors natural selection by allowing a population of sequences to change over time as it adapts to increasing levels of selective pressure. During a directed evolution experiment, proteins with desired phenotypes increase in abundance, while inactive variants go extinct. As an example, efforts to increase the thermal stability of a protein often involve forcing a population of sequences to survive a thermal denaturation step that is designed to separate stably folded proteins from the more weakly folded population. Although approaches to directed evolution have changed over time due to improvements in molecular biology techniques, the process itself continues to play an important role in biotechnology where specialized enzymes are needed to catalyze chemical reactions that differ from those found in nature.

Library generation is a critical step in the directed evolution process as it defines the region of sequence space that will be surveyed in hopes of identifying new variants with desired functional properties. The most common techniques used for library construction include rational design by site-directed mutagenesis, site-saturation mutagenesis, deep mutational scanning, error-prone PCR (epPCR), and homologous recombination (Arnold & Georgiou, 2003). Most of these techniques rely on the introduction of random or controlled sequence diversity into a gene that leads to the production of nonsynonymous mutations in translated proteins. Homologous recombination, also known as DNA shuffling, offers a different route to library construction by allowing sequence diversity to accumulate through a process that recombines useful mutations found in a population of related sequences. In its original form, DNA shuffling involved the random fragmentation of a gene followed by its reassembly using the polymerase chain reaction (PCR) to produce a pool of chimeric sequences (Stemmer, 1994). In recent years, this procedure has been

refined using synthetic DNA fragments that allow researchers to precisely define the recombination sites (Ness et al., 2002).

In vitro selection technologies that link the phenotype of a protein to its genetic sequence are vital to the success of a directed evolution experiment (Packer & Liu, 2015). Library size and selection format are important parameters to consider when choosing a selection technology. Since many excellent reviews have been dedicated to this subject, we will briefly highlight a few examples. Colony screening is the simplest approach to protein engineering. It requires the least amount of technical expertise and instrumentation but is inherently low throughput due to the need to evaluate library members on an individual basis. mRNA display and phage display allow very large library diversities (10^9–10^{12}) to be queried en masse but are better suited for binding than catalysis (Levin & Weiss, 2006). Compartmentalization strategies, such as the one described in this protocol, offer attractive routes to enzyme engineering because they provide environments that support multiple turnover activity. Compartmentalized self-replication (CSR) (Ghadessy, Ong, & Holliger, 2001), compartmentalized self-tagging (CST) (Pinheiro et al., 2012), and droplet-based optical polymerase sorting (DrOPS) are examples of in vitro selection technologies that are performed in a water-in-oil (w/o) droplet (Larsen et al., 2016). These approaches have been particularly successful in the evolution of DNA polymerases for novel activities (Nikoomanzar, Chim, Yik, & Chaput, 2020).

This chapter focuses on methods that were developed to change the substrate specificity of natural DNA polymerases (Fig. 1). We were interested in engineering DNA polymerases for the ability to synthesize threose nucleic acid (TNA) on DNA templates. Although this example is highly specific, the steps taken were sufficiently general that they could be applied to many other problems in protein engineering. We examine the intricacies of library generation by homologous recombination using synthetic DNA fragments that are randomly assembled by PCR. This portion of the method is extensive, as a detailed description is required to generate a library by homologous recombination. We then describe how the library is carried through iterative cycles of selection and amplification with a special emphasis on library recovery and regeneration. The intricate details of our microfluidic-based approach of droplet-based optical polymerase sorting (DrOPS) are only lightly discussed as this technology was the subject of a previous chapter (Vallejo, Nikoomanzar, & Chaput, 2020). We suggest that the robustness of this method should enable others to broadly expand the substrate specificity of natural DNA polymerases.

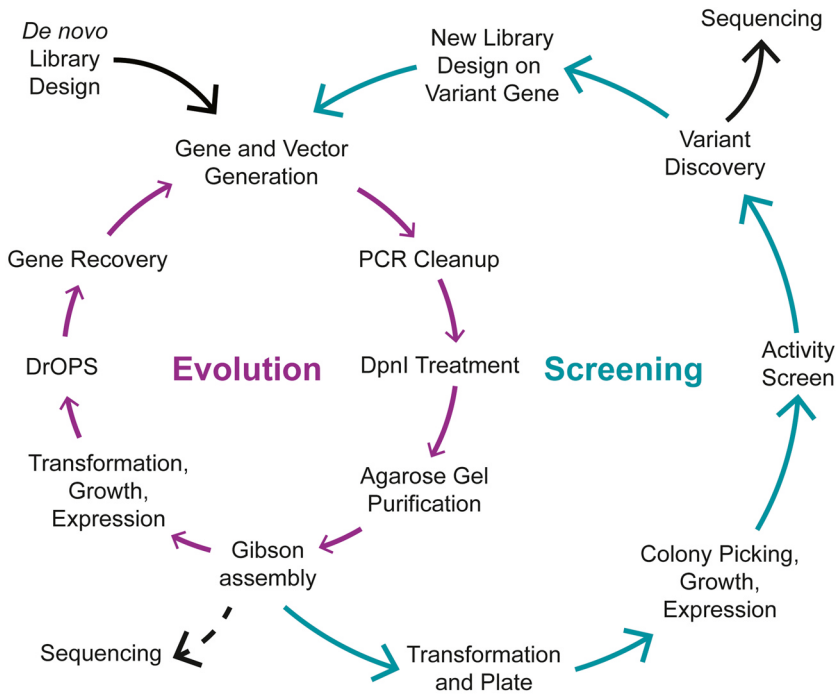

Fig. 1 Summary of directed evolution cycle for TNA polymerase discovery. A de novo library is designed through in silico modeling to verify sequence composition, primer binding sites, and proper translation. Genes are used directly from gBlocks or generated by PCR amplification with a high-fidelity polymerase, along with PCR generation of the vector backbone. The PCR amplicons are *Dpn*I treated, agarose gel purified, and ligated through Gibson assembly. The vector library is transformed into *E. coli*, grown, expressed, and encapsulated into microdroplets with a fluorescent based assay. Active droplets are sorted, recovered, gene amplified, and prepared for iterative rounds of selection or activity screening. Key variants discovered from activity screens are sequenced by Sanger sequencing and become the backbone for the next library design.

2. Library design

2.1 Designing a library for homologous recombination

In this section, we describe the in silico process of designing genes that will be used for homologous recombination. We recommend using molecular biology cloning software, such as CLC Main, when designing and annotating genes for library assembly. These programs allow users to establish electronic workflows that simplify vector design, increase

transparency, and minimize errors. Genes are typically designed at the DNA level within the framework of a vector map. Annotated vector maps should contain all the necessary information for a host cell to propagate and express the gene, including the origin of replication, antibiotic resistance, cloning sites, promoter region, and protein affinity tags. Although the choice of which expression vector is best for a given protein is beyond the scope of this protocol, users should consider such variables as desired protein expression level and the expression of endogenous proteins that may impact the selection strategy. For example, in our system, the newly expressed polymerases are challenged to synthesize TNA on a self-priming DNA hairpin, which is susceptible to nuclease degradation. To overcome this problem, we chose the vector pGDR11, which contains a T5 promoter that allows for protein expression in nuclease deficient cell lines such as XL1-Blue.

Library assembly by homologous recombination requires the user to identify homologs of the parent enzyme. This is typically done by BLAST analysis using a parent sequence downloaded from a protein sequence database (e.g., Genbank or Uniprot). Once the genes chosen for homologous recombination are identified, they can be modified with known beneficial mutations that a user may wish to include in the library, and each sequence is codon optimized to ensure efficient expression in *E. coli*. Homologous recombination sites are then chosen by identifying regions of the gene that exhibit high sequence identity between the various homologs. We recommend a minimum of 21 consecutive nucleotides to obtain T_m values in the range of 55–70 °C. As needed, these regions are modified by codon swapping to ensure 100% sequence identity at the recombination sites. Additionally, the parent gene should be modified at the 5′ and 3′ ends with short segments (20–40 nt) matching the vector insertion site. These overlapping regions are needed to insert the gene into the expression vector via Gibson assembly (Fig. 2) (Gibson et al., 2009). Finally, the computationally designed genes are ordered as linear, dsDNA gBlocks, which allows for sequence lengths up to 3 kilobases (kb). Once received, gBlocks are used directly without the need for PCR amplification or purification.

The example described in this chapter involves the directed evolution of an engineered DNA polymerase with improved DNA-dependent TNA polymerase activity. TNA is an artificial genetic polymer comprised of repeating units of α-L-threofuranosyl nucleotides that are connected by 2′,3′-phosphodiester linkages (Schöning et al., 2000). TNA is considered a

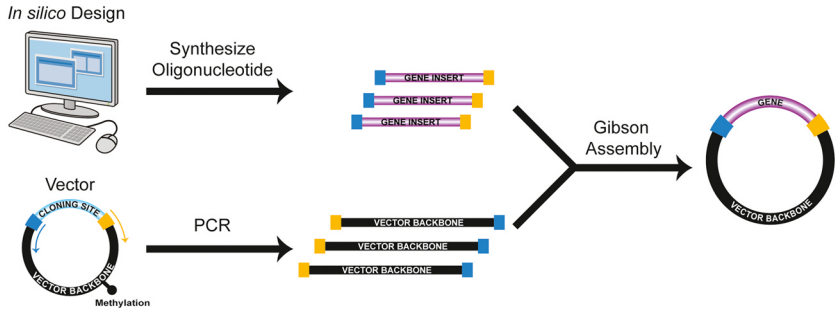

Fig. 2 From in silico design to plasmid generation. Plasmid vector maps are prepared in silico with molecular biology cloning software to ensure that each gene used for homologous recombination is properly designed and assembled. Critical to this process is the design of synthetic recombination sites that are 100% identical across the homologs. The genes are ordered as gBlocks, which prepares the material as linear dsDNA. Separately, the vector backbone is generated by PCR and prepared for gene ligation by Gibson assembly. The gene inserts and vector backbone contain overlapping regions required for vector assembly.

difficult polymer for polymerase recognition as its backbone repeat unit is one atom shorter than that of DNA and RNA. Previous work focused on an archaeal B-family DNA polymerase isolated from the hyperthermophilic species *Thermococcus kodakarensis* (Kod). High throughput screening assays led to the discovery of Kod-RSGA, an engineered form of Kod DNA polymerase carrying the exonuclease silencing mutations D141A and E143A as well as the TNA recognition mutations A485R, N491S, R606G, and T723A (Nikoomanzar, Vallejo, Yik, & Chaput, 2020). In effort to design a library for homologous recombination, BLAST analysis of Kod DNA polymerase was performed to identify other archaeal B-family DNA polymerases, including *Thermococcus gorgonarius* (Tgo), *Pyrococcus* species GB-D Deep Vent (DV), and *Thermococcus* species 9°N. The four genes were aligned, codon optimized, and modified in silico to contain the six mutations found in Kod-RSGA. Sequence alignment identified 10 recombination sites that were further modified by codon swapping to ensure 100% sequence identity for homologous recombination. The four genes (Kod-RSGA, Tgo-RSGA, DV-RSGA, and 9°N-RSGA) were reviewed for accuracy and ordered as gBlocks from IDT. Additionally, the gBlocks contained overlapping segments of 24 and 21 nt at the 5′ and 3′ ends, respectively, to facilitate their insertion into the pGDR11 vector via Gibson assembly (Gibson et al., 2009).

Materials

National Center for Biotechnology Information: BLAST and GenBank.
CLC Main Benchmark.
Custom DNA gBlock (IDT, Coralville, IA).

1. Identify the enzyme or protein to be optimized by directed evolution.
2. Perform a BLAST analysis to identify related genes that will be used for homologous recombination.
3. Confirm the sequence of each enzyme or protein by downloading the sequence from GenBank. In the current study, the following hyperthermophilic archaeal B-family DNA polymerases were used: *Thermococcus kodakaraensis* (Kod, KP682508.1), *Thermococcus gorgonarius* (Tgo, KP682507.1), *Pyrococcus* species GB-D Deep Vent (DV, KP682509.1), *Thermococcus* sp. 9°N (9°N, KP682506.1).
4. Perform a protein sequence alignment to compare the length and sequence composition of each enzyme at the DNA and protein levels.
5. As needed, modify the wild-type sequence at the DNA level to include beneficial mutations known to improve the desired activity. In our case, each polymerase sequence was modified to contain the exonuclease silencing mutations D141A and E143A and the gain-of-function mutations A485R, N491S, R606G, and T723A.
6. From the DNA sequence alignment, identify regions of high sequence identity that can be used for homologous recombination. If needed, codon swap residues in conserved regions to achieve 100% sequence identity. Recombination sites are typically ≥ 21 nt to obtain T_m values in the range of 55–70 °C.
7. Digitally insert the optimized gene sequences into the vector. In our case, the genes were inserted between a 5′ sequence encoding a His-tag and a SalI restriction site in a pGDR11 vector.
8. Identify flanking regions at the 5′ and 3′ ends of gene that can be used to insert the genes into a protein expression vector by Gibson assembly.

   Suggested primer length is at least 20–40 nt.

9. Review the sequences for accuracy before ordering as gBlocks from IDT.

2.2 Amplification of the expression vector for Gibson assembly

Here, we discuss primer design and provide technical recommendations for successful amplification of the vector backbone required for Gibson assembly (Gibson et al., 2009). As mentioned above, the parent gene is

purchased as a gBlock with short segments (20–40 nt) matching the vector insertion site. These overlapping regions are needed to insert the gene into the expression vector via Gibson assembly (Fig. 2). Considerations for primer design include nucleotide length, composition, and salt concentration, as these factors affect the melting temperature of primers in a PCR reaction. The primer sequences should be designed in the 5–3′ direction with a recommended T_m range of 55–70 °C, which can be accomplished with a 15–40 nt sequence. In addition, the T_m range of the forward and reverse primer pairs should vary by ≤3 °C. For example, the forward primer (5′-TGA GTC GAC CTG CAG CCA AGC TTA ATT AGC TGA GC-3′) is 35 nt in length with a T_m of 66.3 °C, while the reverse primer (5′-GCT GCC GCG CGG CAC CAG-3′) is 18 base pair in length with a T_m of 68.3 °C for our vector backbone generation. We found that *in silico* mapping of primers in CLC Main, in conjunction with IDT OligoAnalyzer™ Tool, will increase the chances of proper primer alignment and favorable PCR amplification.

When generating the vector backbone by PCR, we recommend using a high-fidelity polymerase, such as Q5 DNA polymerase, which produces blunt ended DNA for Gibson assembly (Gibson et al., 2009). In a standard Q5 PCR reaction, the following concentrations were used: 0.5 μM of each forward and reverse primer, 0.4 mM dNTPs, 5 ng of parent plasmid, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. For a standard PCR, we used the following program: 95 °C for 2.5 min followed by 30 cycles: (1) denature at 95 °C for 30 s, (2) anneal at 55–72 °C for 45 s, (3) extend at 72 °C, 30 s per kb, and a final polishing step at 72 °C for 1 min. The PCR amplicon lengths and purity are verified by agarose gel electrophoresis. In cases of PCR failure, we recommend troubleshooting the reaction by considering the primer sequence, length, and orientation; annealing temperature of the PCR reaction; integrity of the PCR reagents; and plasmid concentration.

Materials

Custom DNA oligonucleotides.
Nuclease free water.
Ice bucket.
Q5 DNA Polymerase.
Q5 polymerase buffer.
Deoxyribose nucleotides.
Thermal cycler.
Gel Loading Dye, Blue 6×.

General Recommendations for DNA Electrophoresis Protocol (Thermo Scientific, Waltham, MA).
Agarose powder.
Gel Doc XR+ System.
DNA Clean and Concentrator-5.
NanoDrop.

1. Design primers in the 5–3′ direction on the in silico vector map generated in Section 2.1. Primer lengths should vary between 15 and 40 nt with a T_m range of 55–70 °C.

   Design the vector forward primer at the 3′ end of the gene. Design the vector reverse primer at the 5′ end of the gene and copy sequence of the complementary strand.

2. Prepare vector forward and reverse primers as 10 μM working stocks for PCR.
3. Setup a negative control PCR reaction (50 μL) on ice with the following final concentrations: 1× Q5 buffer, 0.5 μM vector forward primer, 0.5 μM vector reverse primer, 0.4 mM dNTPs, and 0.5 μL Q5 DNA polymerase (0.02 U/μL).
4. Setup a PCR reaction (50 μL) on ice with the following final concentrations: 1x Q5 buffer, 0.5 μM vector forward primer, 0.5 μM vector reverse primer, 0.4 mM dNTPs, 0.5 μL Q5 DNA polymerase (0.02 U/μL), and 5 ng of parent plasmid from a miniprep.
5. Thermal cycle the vector amplification with the following suggestions: initial denaturation at 95 °C for 2.5 min; followed by 30 thermal cycles: (1) denature at 95 °C for 30 s, (2) anneal at 68 °C for 45 s, and (3) extend at 72 °C for 2.5 min; polishing step at 72 °C for 1 min.

   PCR reaction optimization should include the anneal temperature, primer concentration, extension times, and plasmid concentration.

6. Mix 2 μL of PCR reaction with 1 μL of Gel Loading Dye, Blue (6×) and 3 μL of nuclease free water. Load mixture into 1% ethidium-bromide agarose gel with the appropriate DNA ladder and run at 120 V for 45 min
7. Visualize agarose gel under UV in Gel Doc XR+ for proper amplicon size.

2.3 PCR cleanup and *Dpn*I treatment

Following successful amplification of the vector backbone required for Gibson assembly, the amplicon product should be purified to reduce the potential for

parental plasmid contamination in subsequent steps. We recommend performing a PCR cleanup prior to *Dpn*I treatment to remove unwanted salts, excess primers, dNTPs, and polymerase. This ensures an optimal buffering system for the *Dpn*I mediated degradation of the methylated or hemi-methylated parent plasmids. We also recommend incubating the PCR products with *Dpn*I at 37 °C for 4–5 h, compared to the suggested time of 1 h, followed by heat inactivation of the enzyme by incubating at 80 °C for 20 min.

Materials
Ice bucket.
1.5 mL centrifuge tube.
Zymo IC column.
DNA wash buffer (10 mM Tris–HCl, pH 7.4, 80% ethanol).
*Dpn*I.
rCutSmart™ (NEB).

1. Transfer PCR reactions into a 1.5 mL centrifuge tube and mix with Zymo DNA binding buffer from DNA Cleanup and Concentrator in a ratio of 300 μL of DNA binding buffer to 100 μL of PCR reaction.
2. Load mixture into Zymo IC column and centrifuge at 14,000 r.c.f for 30 s. Dispose of flow through and reuse collection tube.
3. Wash Zymo IC column with 2 × 500 μL DNA wash buffer and centrifuge at 14,000 r.c.f for 30 s. Dispose of flow through and reuse collection tube between steps.
4. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
5. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2 × 20 μL warmed (37 °C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.

   Prewarmed H_2O may help with DNA extraction.

6. Thaw the *Dpn*I reagents (rCutSmart™ buffer and *Dpn*I enzyme) on ice.
7. Transfer the eluted DNA into a 200 μL PCR tube with 5 μL of rCutSmart™ buffer and 5 μL of *Dpn*I enzyme.
8. Incubate at 37 °C for 4–5 h in an Eppendorf ThermoMixer or Thermal cycler, followed by an inactivation step at 80 °C for 20 min. The product can be held at 24 °C following the incubation until agarose gel purification.

2.4 Agarose gel purification

Agarose gel purification is an essential step for removing parent plasmids and unwanted amplicons from the reaction mixture. We suggest following the

recommendations for DNA electrophoresis (Thermo Scientific) to determine the desired gel percentage. Once the agarose gel percentage is determined, a wide lane comb can be generated by taping several wells together, best applied for loading volumes of ≥50 μL. Agarose gel thickness can affect the migration and visualization of the amplicon. For example, when using a Bio-Rad 7 × 7 cm UVTP tray, we suggest using 50 mL of agarose gel solution for casting and allowing the gel to solidify for 1 h prior to use. We suggest running the agarose gel at a low voltage of 60–70 V until the loading dye migrates through ¾ of the length of the agarose gel. It is important to track the migration of the desired oligonucleotide in reference to the gel loading dye migration. Since agarose gel percentage affects the migration of the oligonucleotide ladder and loading dye, one should empirically determine the migration pattern of the loading dye prior to running a purification gel.

When slicing the agarose gel for DNA extraction, we recommend visualizing the band with a blue light (495 nm) which is less damaging to DNA than a UV light. All surfaces, razor blades, and forceps should be sterilized with 70% ethanol prior to cutting the gel. Isolation of the DNA amplicon can be accomplished using a commercial agarose purification kit. We recommend the Zymo Agarose Dissolving Buffer (ADB) and Zymo IC columns for DNA isolation. We recommend thoroughly cleaning the interior and exterior of the Zymo IC column after loading the DNA to remove any excess salt or agarose gel matrix. Elute the purified amplicon with nuclease free water and quantify the DNA concentration by UV absorbance.

Materials

Bio-Rad comb for GT systems; 1.5 mm thick, 5.5 cm long.
Agarose powder.
Gel Loading Dye, Blue 6×.
Agarose Dissolving Buffer.
Zymo IC column.
DNA wash buffer (10 mM Tris–HCl, pH 7.4, 80% Ethanol).
Eppendorf ThermoMixer.
70% Ethanol for sterilization.

1. Pour a 2% (w/v) ethidium bromide agarose gel with 2 cm wells and solidify to room temperature.

   Comb size can be modified depending on scale of reactions.

2. Add enough Blue Loading Dye Blue 6x to the amplified DNA for a final concentration of 1× and mix by pipetting.

3. Load the DNA and appropriate DNA ladder into the 2% agarose gel. Run at 70 V for 2 h.
4. Visualize the agarose gel under blue light (495 nm) and excise the appropriate band.

Ensure the cutting tools are sterile to avoid contamination.

5. Transfer the excised gel piece into a 1.5 mL low binding centrifuge tube.
6. Add 300 μL of Agarose Dissolving Buffer (ADB) for every 100 mg of agarose gel piece, up to 400 mg.
7. Incubate at 37 °C, 400 rpm, for at least 1 h in an Eppendorf ThermoMixer.
8. Once fully dissolved, load aqueous solution into a Zymo IC DNA column and centrifuge at 14,000 r.c.f. for 30 s
9. Add 500 μL DNA wash buffer into the column. Mix by inverting the column 10 times and aspirate the buffer into waste. Repeat once more.
10. Rinse the exterior of the column with 4 × 1 mL of fresh DNA wash buffer. Place column into a clean 1.5 mL low binding centrifuge tube.
11. Wash the column with 4 × 200 μL of DNA wash buffer and centrifuge at 14,000 r.c.f. for 30 s. Discard the flow through after each spin and reuse the collection tube.
12. Remove excess ethanol by drying the column, centrifuging at 14,000 r.c.f. for 3 min.
13. Elute the DNA with 2 × 10 μL of prewarmed (37 °C) nuclease-free H_2O into a 1.5 mL centrifuge tube, for a total of 20 μL. Let stand for 1 min then centrifuge at 14,000 r.c.f. for 1 min.

2.5 Gibson assembly of the parent genes and linear vector and transformation

Gibson assembly is an advanced cloning technique commonly used for multi-fragment gene assembly in a single isothermal reaction (Gibson et al., 2009). In this technique, the gene, prepared as a gBlock, is inserted into a linearized protein expression vector in a reaction that includes a 5′ exonuclease, a DNA polymerase, and a DNA ligase. The exonuclease removes nucleotides on the 5′ ends of the dsDNA and linear vector, exposing 3′ overhangs that allow for hybridization of the insert and vector. The DNA polymerase fills in the gaps with DNA and the ligase mediates phosphodiester bond formation sealing the nicks between the insert and vector. The newly synthesized constructs are transformed into a cloning strain, such as DH5α competent *E. coli*, for repairs and propagation. We recommend

sequencing individual clones to ensure property assembly of the gene. If mutations are present, it is best to correct these positions by site-directed mutagenesis rather than ordering a new gBlock.

We recommend the following steps for successful Gibson assembly. First, we recommend resuspending the gBlock in nuclease free water at 37 °C for at least 10 min. Importantly, users should resist the temptation to vortex the sample, which can lead to DNA damage. Second, our preferred reaction conditions for Gibson assembly are: 100 ng of vector, 1 μL of gBlock, and 10 μL of 2× Gibson assembly master mix in a final 20 μL reaction. In other examples, such as gene amplification following selection, we recommend using a ratio of 100 ng of vector backbone to 100 ng of insert. Third, we prefer a reaction time of 60 min at 50 °C from the suggested 15–60 min incubation (Gibson et al., 2009). Fourth, we recommend column purifying the Gibson product, then transforming 1 μL of the product into 50 μL aliquot of DH5α competent *E. coli* with the addition of BME (1 μL, 28 μM) prior to the heat shock step to increase the transformation efficiency by degrading carbohydrates on the cell surface (Janjua, Younis, Deeba, & Naqvi, 2014).

Materials

gBlock.
vector backbone.
Ice bucket.
2× Gibson Assembly Master Mix.
Thermal cycler.
DH5α competent E. *coli* cells.
BME, β-mercaptoethanol.
Super Optimal broth with Catabolite repression (SOC).
New Brunswick Innova 44 Incubator.
50 μg/mL carbenicillin Luria-Bertani (LB) agar plates.
Nuclease free water.
10× ThermoPol.
100 μM gene flanking primers (fragment 1 forward and 11 reverse).
Taq DNA polymerase.
Deoxyribose nucleotides.
Gel Loading Dye, Blue 6x.
Agarose powder.
10 mg/mL Ethidium Bromide (EtBr).
Gel Doc XR+ System.
1× Phosphate Saline Buffer (PBS) and 50% glycerol.

50 μg/mL carbenicillin LB media.
Eppendorf Centrifuge 5804.
Plasmid Miniprep Kit.

1. Suspend gBlocks in nuclease free water according to the manufacturer recommended guidelines.

   Do not vortex the DNA. Place at 37 °C to aid reconstitution.

2. Mix 100 ng of purified vector with 1 μL of gBlock and add nuclease free water to a final volume of 10 μL in a PCR tube.
3. Add 10 μL of 2× Gibson master mix and mix by pipette for a total of 20 μL.
4. Place PCR tube into thermal cycler at 50 °C for 1 h.
5. Add 300 μL of Zymo DNA binding buffer to PCR tube and transfer mixture into Zymo IC column.
6. Spin DNA through Zymo IC column at 14,000 r.c.f for 30 s.
7. Wash Zymo IC column with 2 × 500 μL DNA wash buffer and centrifuge at 14,000 r.c.f for 30 s.
8. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
9. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2 × 10 μL warmed (37 °C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.
10. Remove a 50 μL aliquot of DH5α competent cells and place on ice for 5 min to thaw cells.
11. Add 1 μL of purified Gibson assembly product to the competent cells. Mix by flicking the tube 10 times and incubate on ice for 30 min.
12. Supplement the transformation with 1 μL of 1.4 mM BME for a final concentration of 28 μM. Mix and immediately place into 42 °C water bath for 30 s.

    BME has been shown to increase transformation efficiency (Janjua et al., 2014).

13. Place cells on ice for 2 min immediately following heat shock.
14. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37 °C, 225 rpm for 1 h.
15. Plate 250 μL of recovered cells onto 50 μg/mL carbenicillin agar plates and place into 37 °C standing incubator overnight.
16. For colony PCR, include a negative control PCR reaction (10 μL) with the following final concentration: 1× ThermoPol buffer, 0.5 μM of each gene flanking primer (fragment 1 forward primer and fragment 11 reverse primer), 0.4 mM dNTPs, and 0.2 μL Taq DNA polymerase (100 U/mL).

17. Setup colony PCR reaction (20 × 10 μL) with the following final concentration: 1× ThermoPol buffer, 0.5 μM of gene flanking primer set (fragment 1 forward and 11 reverse), 0.4 mM dNTPs, 0.2 μL Taq DNA polymerase (100 U/mL). Pick colony directly off agar plates.
18. For each individual reaction tube, swirl the single colony in the reaction mixture, remove 1 μL of PCR reaction and mix into a separate PCR tube containing 25 μL of 1× PBS, 50% glycerol (glycerol stocks).
19. Thermal cycle the colony PCR with the following suggestions: initial denaturation 95 °C for 2 min; followed by 30 thermal cycles: (1) denature at 95 °C for 15 s, (2) anneal at 58 °C for 15 s, and (3) extend at 72 °C for 2.5 min; polishing step at 72 °C for 1 min.
20. Mix 9 μL of PCR reaction with 2 μL of Gel Loading Dye, Blue (6x) and load mixture into 1% ethidium-bromide agarose gel with appropriate DNA ladder and run at 120 V for 45 min.
21. Visualize agarose gel for proper amplicon size under UV with Gel Doc XR+.
22. Transfer the glycerol stocks that amplified (26 μL containing 1 μL of PCR reaction with 25 μL of 1× PBS, 50% glycerol) into 4 mL of 50 μg/mL carbenicillin LB media. Place media into New Brunswick Innova 44 at 37 °C, at 225 rpm overnight.
23. Harvest the cells after 16–22 h of growth by centrifugation at 4000 rpm, 5 min in Eppendorf Centrifuge 5804.
24. Follow manufacturer instructions for Plasmid Miniprep II to prepare plasmids for sequencing.
25. Sequence the plasmid to verify gene assembly. If mutations are observed during sequencing, perform site-directed mutagenesis to repair genes before moving forward.

2.6 Fragment generation and homologous recombination

In this section, we discuss the generation of a homologous recombination library using a 3-step process (Fig. 3). First, the parent gene is fragmented into individual pieces by PCR using forward and reverse primers specific to each fragment. Second, the gene fragments are randomly reassembled by overlap PCR. Last, PCR is performed to amplify the full-length material from step 2.

As mentioned in Section 2.1, recombination sites were identified as short segments that are shared between the parent homologs. These sites flank the individual gene fragments and serve as primer binding sites for

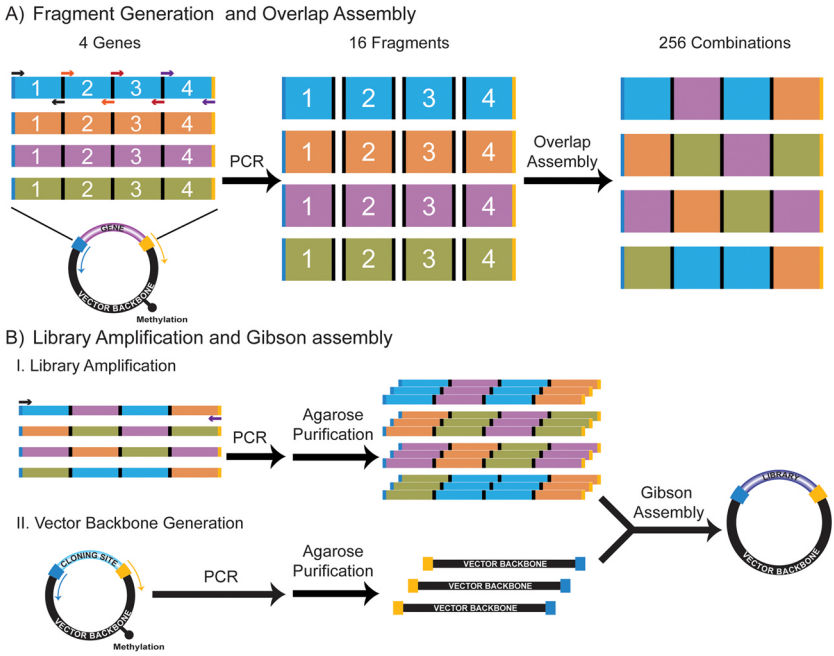

Fig. 3 Generating a homologous recombination library by fragmentation and reassembly. (A) Fragmentation and reassembly. Parent genes designed at the DNA level contain conserved regions across homologs that define the boundaries between segments of the genes that will be individually PCR amplified to create a library of DNA fragments. Fragments can vary in length and number depending on how the library was designed. The total number of gene fragments will equal the number of gene fragment regions times the number of homologs. Once the gene fragments are synthesized, the library is prepared by synthetic homologous recombination in a PCR reaction that contains all the gene fragments plus forward and reverse primer sites defining the vector overlap regions that extend past the 5′ and 3′ ends of the gene. (B) Insertion of the library into a protein expression vector. The DNA library is amplified by PCR and inserted into a protein expression vector by Gibson assembly. It is important to agarose purify both the gene insert and vector backbone to isolate the expected amplicon.

PCR amplification of the gene fragments. Refer to the suggestions of primer design in Section 2.2. In our study, the parent genes contained 11 fragmentation pieces, which required 44 individual PCR reactions to generate the 44 gene fragments as individual amplicons (Fig. 3A). PCR reactions contained final concentrations: 0.5 μM of each primer set, 0.4 mM dNTPs, 5 ng of parent plasmid, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. For a standard PCR cycling, we

recommend the following program: denature at 95 °C for 2.5 min followed by 30 cycles: (1) denature at 95 °C for 30 s, (2) anneal at 55–72 °C for 45 s, (3) extend at 72 °C for 30 s per kilobase (kb) and a final polishing step at 72 °C for 1 min. Amplicon length and purity are analyzed by agarose gel electrophoresis. Upon verification, a PCR cleanup was performed, and the amplicons were eluted in nuclease free water and UV quantified.

The second step is a fragment reassembly process that uses overlap PCR to randomly generate all possible sequence combinations defined by the fragmentation library (Fig. 3A). We recommend using 100 ng of total DNA per every 1000 bases. Since the length of our gene is 2631 bp, we used a total DNA of 263.1 ng for the overlap PCR reaction. We made 10 ng/μL working stocks by combining 25 ng of each homologous parent fragment 1 for a total of 100 ng in a final volume of 10 μL. This process was repeated for fragments 2–11 to obtain 11 individual working stocks. As an example, fragment 1 (179 bp) composes 6.8% of out of the total gene size (2631 bp). Therefore, 6.8% of the 263.1 ng of DNA required for the overlap PCR is 17.9 ng. From a 10 ng/μL working stock, we added 1.79 μL of the fragment 1 in addition to the required amount for subsequent fragments 2–11, along with 1× Q5 buffer, 0.4 mM dNTPs, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. For the overlap PCR, the recommended cycling is as follows: initial denaturation 94 °C for 30 s, followed by 9 thermal cycles of (1) denature at 94 °C for 15 s, (2) extend at 72 °C for 30 s per kb min with a decrease of 0.5 °C per cycle. Then, an additional 5 thermal cycles of (3) denature at 94 °C for 15 s, (4) anneal at 67.5° for 15 s with a decrease of 0.5 °C per cycle, (5) extend at 72 °C for 30 s per kb; with a polishing step of 72 °C for 2 min.

The final step is amplification of the shuffled gene library (Fig. 3B). The overlap PCR reaction (4 μL) from step 2, is used for a PCR with 0.5 μM of the gene flanking primers (fragment 1 forward and fragment 11 reverse), 0.4 mM dNTPs, and 0.5 μL of Q5 polymerase (0.02 U/μL) in a 50 μL reaction. We recommend 20–40 cycles of PCR with the following conditions: initial denaturation 94 °C for 30 s; followed by 17 thermal cycles (1) denature at 94 °C for 15 s, (2) anneal at 67 °C for 15 s with a decrease of 0.5 °C per cycle, (3) extend at 72 °C for 30 s per kb. Additional 23 thermal cycles of (4) denature at 94 °C for 15 s, (5) anneal at 59 °C for 15 s, and (6) extend at 72 °C for 30 s per kb; with a polishing step of 72 °C, 2 min. It is essential to verify assembly of the full gene by agarose gel electrophoresis. Successful PCR amplicons should undergo DNA cleanup and Dpn1

treatment (Section 2.3), followed by agarose gel purification (Section 2.4) in preparation for Gibson assembly (Sections 2.5 and 2.7).

Materials
Custom DNA oligonucleotides.
Ice Bucket.
Q5 polymerase buffer.
Q5 DNA polymerase.
Deoxyribose nucleotides (dNTPs).
Thermal cycler.
Deoxyribose nucleotides.
Gel Loading Dye, Blue 6×.
Agarose powder.
Gel Doc XR+ System.

1. Design custom primers based on recombination sites between all polymerases in Section 2.1.

   Primers should range in length from 15 to 25 base pairs for optimal PCR.

2. Order primers. Prepare as 100 μM stocks with nuclease free water and dilute to 10 μM working stock.
3. Setup a negative control PCR reaction (50 μL) with the following final concentrations: 1× Q5 buffer, 0.5 μM fragment 1 forward primer, 0.5 μM fragment 11 reverse primer, 0.4 mM dNTPs, and 0.5 μL Q5 DNA polymerase.
4. Setup individual PCR reactions (50 μL) with the following final concentrations: 1× Q5 buffer, 0.5 μM of each fragment primer set (fragment 1 forward and fragment 1 reverse primers), 0.4 mM dNTPs, 0.5 μL Q5 DNA polymerase, and 5 ng of parent plasmid. Repeat for each fragment from each of the 4 parents using the designated primer sets (fragment 2 forward and reverse, etc.) for a total of 44 individual PCR reactions.
5. Thermal cycle the fragment amplification with the following suggestions: initial denaturation 95 °C for 2 min; 30 thermal cycles: (1) denature at 95 °C for 30 s, (2) anneal at 68.6 °C for 30 s, and (3) extend at 72 °C for 2.5 min; polishing step at 72 °C for 1 min.
6. Mix 2 μL of PCR reaction, 1 μL Gel Loading Dye, Blue (6x), and 3 μL nuclease free water. Load samples into a 2% ethidium-bromide agarose gel, with appropriate DNA ladder, and run at 120 V for 45 min. Visualize agarose gel under UV and image.

7. Each PCR reaction was mixed with 300 μL of Zymo DNA binding buffer and transfer to Zymo IC column. Centrifuge columns at 14,000 r.c.f for 30 s and discard flow through.
8. Wash Zymo IC column with 2 × 500 μL Zymo DNA wash buffer, centrifuge at 14,000 r.c.f for 30 s and discard flow through.
9. Dry Zymo IC column by centrifugation at 14,000 r.c.f for 3 min.
10. Place Zymo IC column into a clean 1.5 centrifuge tube and elute DNA with 2 × 25 μL warmed (37 °C) nuclease free water. Centrifuge the Zymo IC column at 14,000 r.c.f for 1 min.
11. Quantify the DNA concentration by UV absorbance.
12. Mix 25 ng of each individual fragment cassette (Kod, Tgo, DV, 9°N) for fragment 1 for a total of 100 ng in 10 μL (10 ng/μL stock concentration) for assembly. Repeat process for all subsequent fragments 2–11.
13. Use mass ratio between fragment sizes and the entire gene to determine the amount of DNA to add for the overlap PCR.
14. We suggest a concentration of 100 ng of DNA per 1 kilobase. For this study, the gene size is 2631 base pair, therefore a total of 263.1 ng was used for the overlap PCR.

    Example: Fragment 1 cassette is 179 base pairs out of 2631 total gene, which is 6.8% of the gene. Therefore, 17.9 ng of fragment 1 is required. Add 1.79 μL of DNA from the 10 ng/μL stock concentration.

15. Setup a 50 μL PCR reaction with the following final concentrations: 1× Q5 buffer, 0.4 mM dNTPs, 0.5 μL Q5 DNA polymerase, and a total of 263.1 ng of DNA.
16. For overlap PCR, thermal cycle with the following conditions: initial denaturation 94 °C, 30 s, followed by 9 thermal cycles of (1) denature at 94 °C for 15 s, (2) anneal at 72 °C for 1.5 min with a decrease of 0.5 °C per cycle. Additional 5 thermal cycles of (3) denature at 94 °C for 15 s, (4) anneal at 67.5 °C for 15 s with a decrease of 0.5 °C per cycle, and (5) extend at 72 °C for 1.5 min; with a polishing step of 72 °C for 2 min
17. Setup a negative PCR reaction (50 μL) with the following final concentrations: 1× Q5 buffer, 0.4 mM dNTPs, 0.5 μL of Q5 DNA polymerase, nuclease free water in the presence of 0.5 μM fragment 1 forward, and 0.5 μM fragment 11 reverse gene flanking primers.
18. Setup a PCR reaction (50 μL) with the following final concentrations: 1× Q5 buffer, 0.4 mM dNTPs, 0.5 μL of Q5 DNA polymerase, nuclease free water in the presence of 0.5 μM gene flanking fragment 1

forward and fragment 11 reverse primers, and 4 μL of PCR product from step 16.

19. Perform amplification of PCR product with the following conditions: initial denaturation 94 °C for 30 s; followed by 17 thermal cycles of (1) denature at 94 °C for 15 s, (2) anneal at 67 °C for 15 s with a decrease of 0.5 °C per cycle, and (3) extend at 72 °C for 1.5 min. Additional 23 thermal cycles of (4) denature at 94 °C for 15 s, (5) anneal at 59 °C for 15 s, and (6) extend at 72 °C for 1.5 min; with a polishing step of 72 °C for 2 min
20. Mix 2 μL of PCR reaction with 1 μL of Gel Loading Dye, Blue (6x) and 3 μL of nuclease free water and load into 1% ethidium-bromide agarose gel, with appropriate DNA ladder, and run at 120 V for 45 min.
21. Refer to Section 2.3 for PCR cleanup and *Dpn*I treatment of the PCR amplicon.
22. Refer to Section 2.4 for Agarose gel purification.

2.7 Gibson assembly of the recombination library and linear vector

Materials

Refer to materials in Section 2.5.

1. Mix 100 ng of vector with 100 ng of gel purified gene and add nuclease free water to a final volume of 10 μL in a PCR tube.
2. Refer to Section 2.5 and follow steps 2–25.

2.8 Plasmid scale-up and purification

Plasmid scale-up is necessary to ensure that sufficient coverage of the library is available following Gibson assembly. We recommend scaling the library through five independent transformations with DH5α competent cells. The cells are transformed with manufacturer recommendations, plated, and incubated overnight at 37 °C. Cells are scrapped off the agar plates and midi prepped to recover the plasmid.

Material

Ice bucket.
DH5α competent cells.
Beta-mercaptoethanol (BME).
SOC media.
New Brunswick Innova 44 Incubator.
50 μg/mL carbenicillin LB agar plates.

50 μg/mL carbenicillin LB media.
L-shape Cell Spreader.
50 mL conical tube.
ZymoPURE Express Plasmid Midiprep Kit.

1. Remove 4 × 50 μL aliquot of DH5α competent cells and place on ice for 5 min to thaw cells.
2. Add 1 μL of purified Gibson assembly product to the competent cells. Mix by flicking the tube 10 times and incubate on ice for 30 min.
3. Supplement the transformation with 1 μL of 1.4 mM BME for a final concentration of 28 μM. Mix and immediately place into 42 °C water bath for 30 s.
4. Place cells on ice for 2 min immediately following heat shock.
5. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37 °C, 225 rpm for 1 h.
6. Plate 250 μL of recovered cells onto 50 μg/mL carbenicillin LB agar plates and place into 37 °C standing incubator overnight.
7. Scrape the colonies by adding 1 mL of 50 μg/mL carbenicillin LB media to the agar plate. Use an L-shape cell spreader to gently scrape *E. coli* cells off.
8. Pull all cells into a 50 mL conical tube and add 50 μg/mL carbenicillin LB media to a final volume of 25 mL.
9. Perform ZymoPURE Express Plasmid Midiprep Kit following the manufacturer's recommended instructions.
10. Extract DNA from column with 2 × 300 μL warmed (37 °C) nuclease free water and quantify the DNA concentration by UV absorbance.

3. Cell growth, sample preparation for Droplet-based Optical Polymerase Sorting (DrOPS) and plasmid recovery

In this section, we describe the process of expressing the polymerase library in *E. coli*, and recovery of the DNA from the droplet following droplet-base optical polymerase sorting (DrOPS) (Figs. 4 and 5). Two independent microfluidic devices are utilized to generate and sort water/oil emulsions. A complete description of the microfluidic devices and instrumentation setup was described in a previous chapter (Vallejo et al., 2020). The cell line (XL1-Blue Supercompetent Cells) used for our polymerase

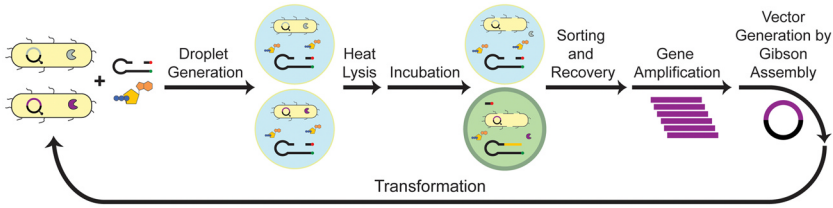

Fig. 4 Directed evolution strategy for evolving TNA polymerases by droplet-based optical polymerase sorting (DrOPS). A library of polymerase variants generated by homologous recombination or error-prone PCR is inserted into a protein expression vector by Gibson assembly. The plasmid library is transformed into *E. coli*, grown and induced to express the polymerase library. Post-induction, *E. coli* cells are encapsulated into microdroplets with the reagents necessary for performing a fluorescence-based polymerase activity assay. Droplets are collected, subjected to thermal lysis, and incubated at 55 °C. Functionally active polymerases capable of extending a labeled DNA hairpin template with TNA generate a fluorescent signal by blocking a DNA quencher strand from annealing with the template. Fluorescent droplets are sorted by a custom fluorescently activated droplet sorter (FADS) mounted on a microfluidic instrument. Fluorescent droplets are collected, and the encoding DNA is amplified for another round of directed evolution.

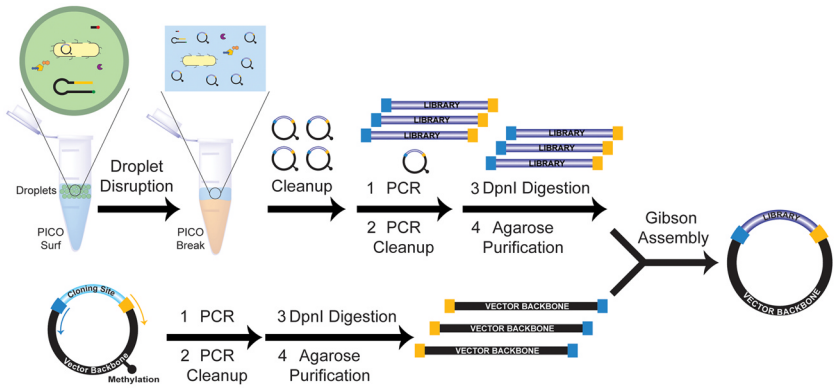

Fig. 5 Plasmid extraction, gene regeneration, and plasmid assembly. Emulsion of positively collected droplets are broken with the addition of PICO Break in a 1:1 ratio of droplet and PICO Surf, which creates a biphasic mixture. The aqueous phase (top) containing the plasmids of interest are extracted and purified. The genes from the active library are PCR amplified, purified, treated with *Dpn*I, and agarose gel purified prior to Gibson assembly with the newly generated expression vector backbone.

expression is a nuclease deficient *E. coli* strain that prevents the degradation of our fluorescently labeled DNA hairpin. We utilize the lac operon repressor system to attenuate expression level. A key step to ensure successful generation of microdroplets is buffer exchanging and cleaning of *E. coli* cells into

the reaction buffer. Once the cells are clean and quantified, an activity reaction mixture is made with the following components: 1× ThermoPol, 1 μM Cy3-fluorescently labeled DNA hairpin, 2 μM Iowa Black complementary sequence, and 100 μM of tNTPs. Excess amount of Iowa Black complementary oligonucleotide is added to ensure complete quenching. Post emulsification, droplets are subjected to a heat lysis at 95 °C for 5 min and incubated at 55 °C for various length of time depending on the desired selection stringency. A microfluidic device is utilized to sort these single layer emulsions on a custom fluorescently activated droplet sorter (FADS). A key step in recovering droplets efficiently is the washing of the collection line with excess amount of buffer used to generate the emulsions. Emulsions are disrupted by the addition of fluorocarbon oils containing surfactants and a vortex. A DNA cleanup step is performed to remove excess fluorocarbon oils prior to amplification of the gene.

Materials

Ice bucket.
XL1-Blue Supercompetent Cells.
SOC media.
50 μg/mL carbenicillin LB media.
New Brunswick Innova 44 Incubator.
Isopropyl-1-thio-β-D-galactopyranoside (IPTG).
14 mL round-bottom Falcon tube.
Eppendorf Centrifuge 5804.
10× ThermoPol.
1.5 mL centrifuge tube.
Eppendorf 5415D centrifuge in 4 °C deli.
5′-Cy3 labeled self-priming hairpin template.
3′-Iowa Black FQ quencher.
Threose nucleic acid (TNA).
Custom DNA oligonucleotides.
Q5 polymerase buffer.
Q5 DNA polymerase.
Deoxyribose nucleotides.
Thermal cycler.
Gel Loading Dye, Blue 6×.
Agarose gel.
Gel Doc XR+ System.

1. Mix 300 ng of library plasmid with 20 μL of XL1-Blue Supercompetent Cells and place on ice for 30 min.
2. Heat shock cells at 42 °C for 30 s and immediately place on ice for 2 min.
3. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37 °C, 225 rpm for 1 h.
4. Transfer the recovered cells into 500 mL baffled flask containing 50 mL of 50 μg/mL carbenicillin LB media and place into New Brunswick Innova 44 at 37 °C, 225 rpm, overnight.
5. Transfer 500 μL of overnight into a fresh 50 mL of 50 μg/mL carbenicillin LB media and place into New Brunswick Innova 44 at 37 °C, 225 rpm, until OD_{600} reaches ~0.600 au.
6. Induce protein expression by adding IPTG to a final concentration of 1 mM. Place the cells back into the New Brunswick Innova 44 at 25 °C, 225 rpm, 20 h.
7. Transfer 1 mL of cells into a 14 mL round-bottom falcon tube and centrifuge cells at 1811 r.c.f for 5 min in Eppendorf centrifuge.
8. Discard supernatant, wash cells with 1 mL of 1× ThermoPol, and centrifuge cells at 1811 r.c.f for 5 min in Eppendorf centrifuge. Repeat four times.
9. After the fourth wash, suspend cells with 2 mL of 1× ThermoPol and measure OD_{600}.
10. Transfer the appropriate volume of the suspended cell to a 1.5 mL centrifuge tube and centrifuge at 1811 r.c.f for 5 min in Eppendorf 5415D centrifuge.
11. Carefully discard supernatant as to not disturb the pellet.
12. Prepare 250 μL reaction mixture for the fluorescence-based polymerase activity assay containing the following final concentrations: 1× ThermoPol buffer, 1 μM of a 5′-Cy3 labeled self-priming hairpin template, 2 μM of a 3′-Iowa Black FQ quencher complementary to the 5′ end of the template, and 100 μM of tNTPs.
13. Follow (Vallejo et al., 2020) for microfluidic device fabrication and instrumentation for droplet generation.
14. Following droplet generation, droplets are incubated at 95 °C for 5 min, cooled to 24 °C for 5 min, and incubated at 55 °C for variable time depending on desired selection stringency (18–2 h).
15. Follow (Vallejo et al., 2020) for microfluidic device fabrication, droplet sorting, and recovery of DNA.

16. Setup a negative control PCR reaction (50 μL) with the following final concentrations: 1× Q5 buffer, 0.5 μM fragment 1 forward primer, 0.5 μM fragment 11 reverse primer, 0.4 mM dNTPs, and 0.5 μL Q5 DNA polymerase.
17. Setup PCR reaction (50 μL) for each extraction with the following final concentrations: 1× Q5 buffer, 0.5 μM fragment 1 forward primer, 0.5 μM fragment 11 reverse primer, 0.4 mM dNTPs, 0.5 μL Q5 DNA polymerase, and 4 μL of the purified recovered DNA.
18. Thermal cycle the gene amplification with the following suggestions: initial denaturation 95 °C for 2.5 min; followed by 30 thermal cycles: (1) denature at 95 °C for 30 s, (2) anneal at 70 °C for 30 s, and (3) extend at 72 °C for 2.5 min; polishing step at 72 °C for 1 min.
19. Mix 2 μL of PCR reaction with 1 μL of Gel Loading Dye, Blue (6x) and 3 μL of nuclease free water. Load mixture into 1% ethidium-bromide agarose gel, with the appropriate DNA ladder, and run at 120 V for 45 min.
20. Visualize agarose gel under UV in Gel Doc XR+ for proper amplicon size.
21. Refer to Section 2.3 for PCR cleanup and *Dpn*I treatment.
22. Refer to Section 2.4 for agarose gel purification.
23. Refer to Sections 2.5 and 2.7 for Gibson assembly.

4. Colony picking and activity screen

In this section, we discuss a simple approach for screening enriched variants for polymerase activity. Library members are transformed into *E. coli* and plated onto an antibiotic containing agar plate. Cells are diluted in a ratio of 1:20 before plating to allow for isolated colony to grow overnight at 37 °C. Single colonies are picked and placed in 4 mL of LB media with antibiotics for overnight growth at 37 °C with shaking at 225 rpm. A small amount of overnight culture (40 μL) is inoculated into 4 mL of fresh media, grown to the appropriate cell density, and induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Following induction, cells are incubated overnight at 25 °C with shaking at 225 rpm, although temperatures may vary depending on the expression vector. The following day, cells are harvested and suspended in polymerase lysis buffer. The suspended cells are subjected to a heat lysis at 70 °C for 1 h, followed by cooling on ice for 1 h, and centrifuge to pellet cell

debris and denatured proteins. The clarified lysate is estimated with 1× Bradford dye before the primer extension assay in which polymerases are challenged to extend tNTPs on a labeled DNA primer-template duplex. The reactions are quenched with stop buffer and run on a denaturing polyacrylamide gel. Gels are imaged, and analysis is performed on the banding pattern of the variant polymerase as compared to the parent polymerase.

Materials

Ice bucket.
XL1-Blue Supercompetent Cells.
SOC media.
50 μg/mL carbenicillin LB agar plates.
50 μg/mL carbenicillin LB media.
New Brunswick Innova 44 Incubator.
Isopropyl-1-thio-β-D-galactopyranoside (IPTG).
14 mL round-bottom Falcon tube.
Eppendorf Centrifuge 5804.
Polymerase lysis buffer (10 mM Tris–HCl, pH 8, 500 mM NaCl).
1.5 mL centrifuge tube.
Eppendorf 5415D centrifuge in 4 °C deli.
200 μL PCR tubes.
10× ThermoPol.
5′-IR800-DNA primer.
DNA template.
Threose nucleic acid (TNA).
Quenching buffer (25 mM EDTA, 95% formamide).
SequaGel UreaGel 29:1 Denaturing Gel System.
Oddessy Li-Cor Imager.
1× Bradford Dye.

1. Transfer 300 ng of plasmid purified from Gibson assembly of directed evolution output and transform with 20 μL XL1 Blue cells and incubate for 30 min on ice.
2. Heat shock cells at 42 °C for 30 s and immediately place back on ice for 2 min.
3. Add 1 mL of warmed SOC to cells and place into New Brunswick Innova 44 at 37 °C, 225 rpm for 1 h.
4. Plate 250 μL of recovered cells onto 50 μg/mL carbenicillin LB agar plates and place into 37 °C standing incubator overnight.

5. Transfer 4 mL of 50 μg/mL carbenicillin LB media into a 14 mL round-bottom Falcon tube. Pick single colonies for each tube (100–500 in batches of 20) and place into New Brunswick Innova 44 at 37 °C, 225 rpm, overnight.
6. Transfer 40 μL of overnight into 4 mL of fresh 50 μg/mL carbenicillin LB media in a 14 mL round-bottom Falcon tube and place into New Brunswick Innova 44 at 37 °C, 225 rpm, until OD_{600} reaches ~0.600 au.
7. Induce cells with a final concentration of 1 mM IPTG and place into New Brunswick Innova 44 at 25 °C, 225 rpm, for 20 h.
8. Centrifuge cells at 4000 rpm for 5 min in Eppendorf Centrifuge 5804 and discard media supernatant.
9. Suspend cells with 100 μL of polymerase buffer (10 mM Tris–HCl, pH 8, 500 mM NaCl, 10% glycerol) and transfer into a 1.5 mL centrifuge tube.
10. Incubate the 1.5 mL centrifuge tube at 70 °C for 1 h.
11. Place the 1.5 mL centrifuge tube on ice for 1 h.
12. Centrifuge the 1.5 mL centrifuge tube at 4 °C, 14,000 r.c.f for 1 h to pellet cell debris.
13. Mix 4 μL of the supernatant with 200 μL of Bradford dye to verify the presence of protein.
14. Prepare the activity assay mixture (20 μL) with the following final concentrations: 1× ThermoPol, 1 μM 5′-IR800-label DNA primer, 1 μM DNA template in the absence of 100 μM tNTPs, and 2 μL clarified lysate.
15. Incubate the mixture at 90 °C for 5 min and immediately place in ice for 5 min.
16. To the reaction mixture, add 100 μM tNTPs and 2 μL clarified lysate on ice.
17. Incubate the reaction at 55 °C at various time points.
18. Remove 1 μL of the reaction mixture and add with 39 μL of quenching buffer (25 mM EDTA and 95% formamide).
19. Incubate quenched reaction at 95 °C for 10 min to denature DNA and TNA heteroduplex.
20. Prepare denaturing PAGE under manufacturer's instruction for SequaGel UreaGel 29:1 Denaturing Gel System.
21. Load 10 μL of denatured sample into the denaturing PAGE and run at 10 Watts for 1.5 h.
22. Image gel on Li-Cor Odyssey to observe extension of tNTPs onto the DNA primer.

5. Mutagenic PCR

Mutagenic PCR is a molecular biology technique that is commonly used to introduce random mutations into a gene. In a typical mutagenic PCR reaction, a low fidelity polymerase, such as Taq, is used. However, Taq has been shown to incorporate nucleotides resulting in an overhanging at the 3′ termini. This is problematic for Gibson assembly, as the overhang will cause an undesired frame shift. In lieu of multiple rounds of PCR between a low fidelity and high-fidelity polymerase, we recommend utilizing a recombinantly expressed high-fidelity polymerase with the exonuclease activity silenced. The polymerase concentration will need to be empirically determined. We recommend biasing the concentration of triphosphates in favor of pyrimidines to reduce G·T wobble mismatches and G→A transitions (McCullum, Williams, Zhang, & Chaput, 2009). It is helpful to perform a concentration gradient of manganese to determine acceptable amounts to favor the generation of proper size amplicons and prevent uncontrolled amplification.

Materials

10× ThermoPol.
Manganese Chloride • 6 H_2O ($MnCl_2$).
Custom oligonucleotide primers.
1.5 mL centrifuge tube.
200 μM PCR tubes.
Deoxyribose nucleotides (dNTPs).

1. Prepare a fresh 50 mM stock of $MnCl_2$ and dilute to 1 mM.
2. Setup a negative control PCR reaction (50 μL) without plasmid, with the following final concentrations: 1× ThermoPol buffer, 1 μM each forward and reverse primers, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 μM Kod-WT (exo-) polymerase.
3. Setup PCR reactions (4 × 50 μL) with the following final concentrations: 1× ThermoPol buffer, 1 μM each fragment 1 forward and fragment 11 reverse primer, 1 mM TTP, 1 mM dCTP, 0.2 mM dGTP, 0.2 mM dATP, 5.5 mM $MgCl_2$, 0.013 μM Kod-WT (exo-) polymerase, and 100 ng plasmid. Supplement each reaction with either 250, 125, 62.5, 31.25 or 0 μM $MnCl_2$.
4. Thermal cycle the PCR according to the following suggestions: initial denaturation 95 °C for 2 min; followed by 30 cycles thermal cycling of

(1) denature at 95 °C for 1 min, (2) anneal at 60 °C for 1 min, and (3) extend at 68 °C for 2.5 min; hold at 4 °C.

To determine the optimal annealing temperature, a gradient PCR can be performed. Further optimizations can also be performed on extension time and temperature.

5. Analyze 2 μL aliquots of PCR reactions on a 1% (w/v) ethidium bromide agarose gel with the appropriate DNA ladder to validate amplicon size. Gel band intensities should be robust and clean with successful PCR.

   Troubleshooting parameters include optimization of primer length or concentration, polymerase concentration, $MnCl_2$ concentration, or extension time.

6. Refer to Section 2.3 for PCR cleanup and *Dpn*I treatment for Gibson assembly.
7. Refer to Section 2.4 for Agarose Gel Purification.
8. Refer to Sections 2.5 and 2.7 for Gibson assembly.
9. Refer to Section 2.8 for Plasmid scale-up and purification.
10. Refer to Section 3.0 for Cell growth, sample preparation for Droplet-based Optical Polymerase Sorting (DrOPS) and plasmid recovery.
11. Refer to Section 2.3 for PCR cleanup and *Dpn*I treatment.
12. Refer to Section 2.4 for Agarose Gel Purification.
13. Refer to Sections 2.5 and 2.7 for Gibson assembly.
14. Refer to Section 2.8 for Plasmid scale-up and purification.
15. Refer to Section 3.0 for Colony picking and activity screen.

6. Notes

- Problems with PCR amplification of the library can usually be fixed by gradient PCR, primer design, or cycle optimization.
- Utilize Thermo Scientific general recommendations for DNA electrophoresis to determine the appropriate agarose gel percentage for the expected amplicon size.
- We highly recommend *Dpn*I treatment and agarose gel purification of the PCR amplified expression vector before Gibson assembly.
- Increased time and enzyme concentration may improve the efficiency of *Dpn*I treatment.
- For mutagenic PCR, one may need to empirically determine the amount of polymerase and $MnCl_2$ to optimize the PCR.

7. Summary and conclusions

This chapter describes a directed evolution strategy used to engineer archaeal B-family DNA polymerases to recognize TNA, an unnatural nucleic acid analog (Schöning et al., 2000). We provide a special emphasis on library design and generation by homologous recombination using synthetic gene fragments that are randomly reassembled at engineered recombination sites. We first describe how parent scaffolds are identified, acquired, and cloned into a preferred expression vector by Gibson assembly. Next, we describe the 3-step process of homologous recombination beginning with fragment generation, where the encoding DNA for each parent scaffold is fragmented into cassettes by PCR using primers that are specific for each fragment. Second, we describe how the cassettes are randomly assembled by overlap PCR, and third, a final PCR is performed to amplify full-length material from step 2. We provide a detailed methodology for expression of the polymerase library in *E. coli* and lightly describe the preparation of single cell encapsulation into uniform water-in-oil droplets, and recovery of the DNA from the droplet following droplet-based optical polymerase sorting (DrOPS). The genes from the active library are PCR amplified, purified, and ligated to the expression vector for colony picking and activity screening. The activity screen reveals variants with desired functional activities, whose genetic information can inform the basis of the next round of directed evolution. To further diversify the sequence landscape, a protocol for mutagenic PCR is also included.

This contribution provides a systematic protocol to generate a library of mutant enzymes with various fitness levels across a diverse sequence landscape. This procedure allows users to screen extremely diverse libraries without relying on NGS sequencing. These methodologies are general and can be of value to other enzyme engineering studies on DNA modifying enzymes.

Acknowledgments

We wish to thank members of the Chaput lab for their helpful comments on the manuscript. This work was supported by grants from the National Science Foundation (CHM: 2001434 and MCB: 1946312).

References

Arnold, F. H., & Georgiou, G. (2003). *Directed evolution library creation: Methods and protocols*. Humana Press.

Ghadessy, F. J., Ong, J. L., & Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 4552–4557.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*(5), 343–345.

Janjua, S., Younis, S., Deeba, F., & Naqvi, S. M. S. (2014). High efficiency DNA transformation protocol for *Escherichia coli* using combination of physico-chemical methods. *International Journal of Agriculture and Biology, 16*(1), 132–138.

Larsen, A. C., Dunn, M. R., Hatch, A., Sau, S. P., Youngbull, C., & Chaput, J. C. (2016). A general strategy for expanding polymerase function by droplet microfluidics. *Nature Communications, 7*, 11235.

Levin, A. M., & Weiss, G. A. (2006). Optimizing the affinity and specificity of proteins with molecular display. *Molecular Biosystems, 2*(1), 49–57.

McCullum, E. O., Williams, B. A. R., Zhang, J., & Chaput, J. C. (2009). (3rd ed.). *Random mutagenesis by error-prone PCR364*. Springer.

Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V., ... Minshull, J. (2002). Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nature Biotechnology, 20*(12), 1251–1255.

Nikoomanzar, A., Chim, N., Yik, E. J., & Chaput, J. C. (2020). Engineering polymerases for applications in synthetic biology. *Quarterly Review of Biophysics, 53*, e8.

Nikoomanzar, A., Vallejo, D., Yik, E. J., & Chaput, J. C. (2020). Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synthetic Biology, 9*(7), 1873–1881.

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews. Genetics, 16*(7), 379–394.

Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., ... Holliger, P. (2012). Synthetic genetic polymers capable of heredity and evolution. *Science (New York, N. Y.), 336*(6079), 341–344.

Schöning, K. U., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R., & Eschenmoser, A. (2000). Chemical etiology of nucleic acid structure: The a-threofuranosyl-(3′–>2′) oligonucleotide system. *Science (New York, N. Y.), 290*(5495), 1347–1351.

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature, 370*, 389–391.

Vallejo, D., Nikoomanzar, A., & Chaput, J. C. (2020). Directed evolution of custom polymerases using droplet microfluidics. *Methods in Enzymology, 644*, 227–253.

Zeymer, C., & Hilvert, D. (2018). Directed evolution of protein catalysts. *Annual Review of Biochemistry, 87*, 131–157.