Phage Assisted Continuous Evolution (PACE): a How-to Guide for Directed Evolution

Serban C. Popa,[1] Ichiro Inamoto,[1] Benjamin W. Thuronyi,[2] and Jumi A. Shin[1,*]

[1] Department of Chemistry, University of Toronto, 3359 Mississauga Road, Mississauga, ON, L5L 1C6, Canada.

[2] Department of Chemistry, Williams College, 47 Lab Campus Drive, Williamstown, MA, 01267, USA.

* To whom correspondence should be addressed: jumi.shin@utoronto.ca, (905) 828-5355

Keywords: continuous evolution, protein design, protein engineering, phage, bacterial one-hybrid, plaque assay, mutational analysis, DNA sequencing.

ABSTRACT

Directed evolution methods are becoming increasingly popular, as they are extremely powerful toward developing new biomolecules with altered/novel activities, e.g., proteins with new catalytic functions or substrate specificities, and nucleic acids that recognize an intended target. Especially useful are systems that have incorporated *continuous* evolution, where the protein to be evolved undergoes continuous mutagenesis to evolve a desired trait with little to no input from the researcher once the system is started. However, continuous evolution methods can be challenging to implement in the lab and daunting for researchers to invest time and resources. Our intent is to provide basic information and helpful suggestions that we have gained from our experience with bacterial phage-assisted continuous evolution (PACE). Specifically, we review factors to consider before adopting PACE for a given evolution scheme, different types of selection circuits that can be utilized with particular focus on the PACE-B1H selection system, what optimization of a PACE selection circuit may look like using directed evolution of ME47 as a case study, and additional techniques that may be incorporated into a PACE experiment. With this information, researchers will be better equipped to determine if PACE is a valid strategy to use to evolve their proteins and how to set up a valid selection circuit.

INTRODUCTION

Although directed evolution can be applied toward generating novel nucleic acids and nucleic-acid-encoded biomolecules, researchers have primarily focused on the continuous evolution of proteins. Proteins are the vehicle through which biological processes take place that allow for life to occur as we know it. For decades, scientists have taken advantage of proteins to advance science: examples include restriction enzymes, DNA polymerases, antibodies, fluorescent proteins, and CRISPR-associated proteins. A wealth of technologies has been developed to engineer and modify these proteins toward a specific need. One example of protein design success is the variety of available fluorescent proteins with different spectral properties, made by modifying the original green fluorescent protein from *Aequoria victoria* (reviewed in refs. [1, 2]), as well as fluorescent proteins from other marine organisms; this demonstrates the utility and function that protein engineers have achieved by manipulating the protein scaffold.

Protein engineering & principles of directed evolution

Methods for protein engineering can be classified into two major categories: rational design and directed evolution. Rational design refers to the use of literature, modeling, and knowledge of the protein scaffold to generate novel proteins with desired traits[3, 4]. For example, we can alter the specificity of an enzyme by mutating amino acids in the enzyme's active site. Although the rational approach is powerful and rapid, a limitation is the availability of useful information: e.g., we must know an enzyme's structure and function—in particular, the structure of its active site—in order to mutate amino acids that can impact ligand-binding and catalysis. Rationally designed libraries tend to be small as they are laborious to create, which limits their utility. In addition, because proteins are complex and dynamic scaffolds, we cannot reliably predict that these rationally

designed modifications will give the intended outcome.

Directed evolution harnesses genetic systems toward improving the protein by subjecting it to multiple rounds of mutagenesis and selection[5, 6]. By selecting proteins with improved attributes while introducing random mutations into their sequences, directed evolution can circumvent the limitations of rational design. During directed evolution, mutations are typically incorporated into the protein's DNA-coding sequence (CDS) through methods such as error-prone PCR and chemical mutagenesis, thereby generating a DNA library for the protein of interest. The resulting library can then be screened; library members are sorted to find those rare individuals with the desired change. Alternatively, a selection can be carried out where only library members with a level of beneficial activity above a specified threshold are observed. Screens and selections exist on a continuum, where the trait that is meant to be evolved determines the appropriate strategy to be adopted[7]. The CDSs for the desired proteins are then isolated and inserted into the next cycle of mutagenesis and selection. This process can be repeated indefinitely, and it is similar to natural evolution: the mutant proteins represent individuals of a population competing with each other to pass their genes into the next generation, and the selective assay represents the selective pressure that defines the "fitness" required for the organism to survive. After multiple iterations of mutagenesis and selection, advantageous mutations can accumulate in the protein CDS, thereby allowing the protein to acquire the desired property, as defined by the selective assay.

In contrast to rational design, directed evolution can be performed with less knowledge of the protein. A solid understanding of the protein's activity and mode of action is required, as these are vital toward building an appropriate selective assay for the protein. However, structural information is typically unnecessary, as directed evolution allows you to work with larger libraries that can be mutated randomly and provide good coverage of all possibilities[5]. This also means

that directed evolution can identify mutations in unexpected regions of the protein that would have been impossible to foresee by using high-resolution structural information. However, the use of directed evolution brings new challenges to the table, one of which is the design of a selective assay that appropriately guides the protein's evolution in the desired direction. Other challenges include the generation of large, unbiased DNA libraries, avoidance of false positive and negative signals, and the significant effort involved in manually performing multiple generations of mutagenesis and selection[8, 9].

PACE remedies some of these challenges by automating the maintenance of the mutagenesis and selection cycles. PACE accelerates the evolutionary process by exploiting the rapid, replicative lifecycle of the M13 phage and the ease of recovery of the phage genome from the host cell. *Although PACE has been successfully applied to protein evolution, this system can be used toward evolution of any genetically encoded molecule, including nucleic-acid aptamers*.

Directed Evolution & the M13 bacteriophage

Phage-assisted continuous evolution (PACE) is a ***true* evolution system**, in which evolving genes are subjected to continuous, seamless cycles of mutagenesis and selection[10-12]. Liu and coworkers developed PACE to provide an environment like natural evolution, where random mutations in DNA are produced in every generation at a rate that is much higher than what occurs naturally, and expressed proteins are selected for their fitness *in situ*; this contrasts with conventional methods of directed evolution, in which mutagenesis and selection are performed in discrete steps that require mutations to be introduced by scientists at every iteration of evolution[13].

To accomplish directed evolution without constant human intervention, PACE utilizes the

continuous infection of *Escherichia coli* host cells by a modified version of M13 bacteriophage. The mature M13 bacteriophage particle features a rod-shaped protein shell carrying a circular single-stranded phage DNA (Fig. 1). The protein shell contains five different phage coat proteins. The majority of the coat is built from >2000 copies of phage protein pVIII, while smaller numbers of proteins pIII, pVI, pVII, and pIX are found at the ends of the rod-shaped shell. All coat proteins are essential for the maturation of M13 phage[14].

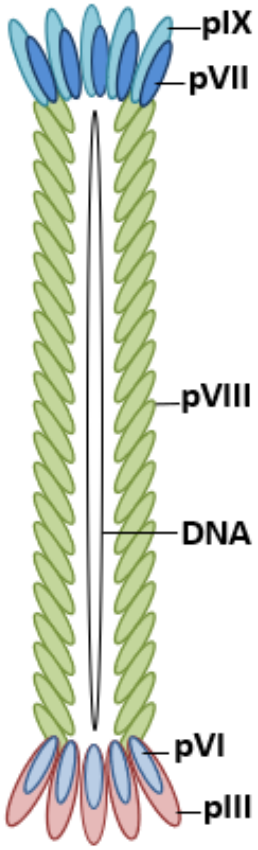


Figure 1. Schematic of M13 bacteriophage. The five phage coat proteins (pIII, pVI, pVII, pVIII, pIX) and single-stranded circular DNA are labeled[14].

During infection, the M13 particle attaches to F^+ *E. coli* via the F pilus using the N2 domain of the gIII protein, which "reels" in the phage and allows the N1 domain of pIII to interact with the TolQRA complex, after which the C domain of gIII mediates virion uncoating and infection

(Fig. 2A)[14, 15]. Once inside the cytoplasm of host *E. coli*, phage DNA uses the host's DNA replication machinery to produce a circular double-stranded version of itself and starts to express various phage proteins required for production of mature M13 particles. 30 minutes after phage infection, dozens of double-stranded, transcriptionally active phage molecules are present within the host cell[14]. Phage are made via rolling-circle replication from the initial viral single-stranded genome—note this is an important factor when considering gene dosage effects for PACE (more on this later)[16]. The phage coat proteins assemble on the surface membrane of the host *E. coli*, and a copy of the circular single-stranded phage DNA is packaged inside. Once assembly is complete, the M13 particle detaches from the bacterial membrane and exits to the environment to start another cycle of infection. Unlike other bacteriophage, M13 is not lytic: the host *E. coli* is kept alive, while mature M13 particles release from the cell membrane. The host growth rate is significantly diminished, however, as a result of producing phage particles, but host cells can continue to grow, divide, and further produce M13 phage indefinitely.

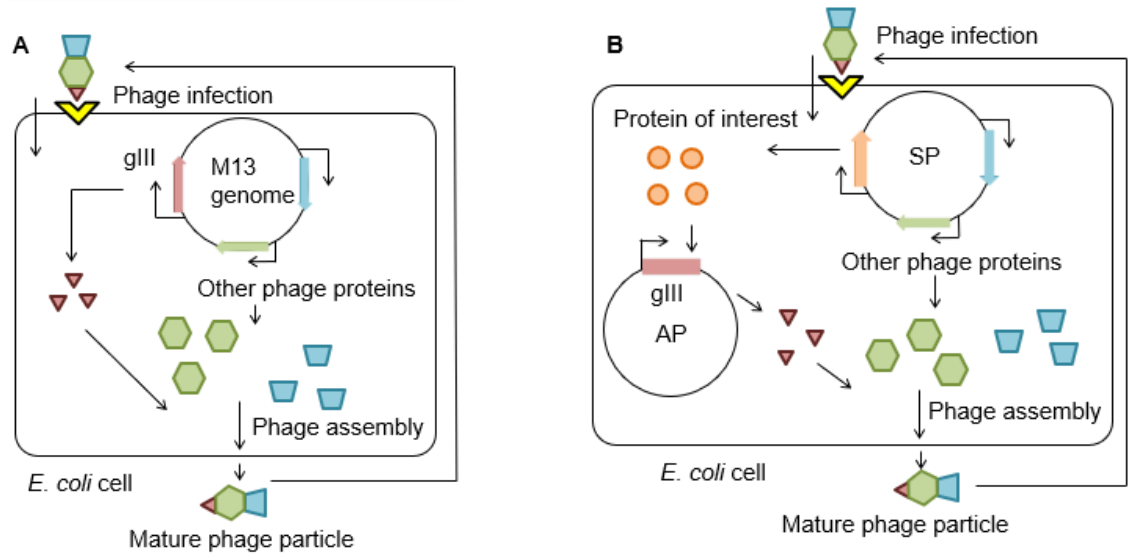


Figure 2. Simplified schematic of the replication cycle: **A)** M13 bacteriophage, and **B)** selection phage (SP). For both diagrams, gIII and its product pIII are shown as brown triangles; other phage proteins are represented in green and blue. The protein of interest, expressed from the SP that mediates expression of gIII, is represented as orange circles. M13 infection of host *E.coli* is

mediated by interaction of the gIII coat protein with the TolQRA complex (yellow)[14, 15]. Upon entry into the host, the single-stranded viral genome is converted to a double-stranded genome by host DNA polymerases and expression of the M13 genome can begin by utilizing host machinery[16].

Phage protein pIII, which is encoded by phage gene gIII, is essential for phage maturation and infectivity[17]. The infectivity of M13 phage scales with increasing levels of pIII over a range of two orders of magnitude[18]. PACE utilizes a mutant M13 bacteriophage whose gIII gene is replaced by that for the protein of interest (the mutant phage is called Selection Phage, SP, Fig. 3)[12]. Thus, the SP expresses the protein instead of pIII in host *E. coli*; the SP cannot produce mature phage particles by itself. To complement the SP, the gIII gene is supplied on a separate plasmid in the host *E. coli* (Accessory Plasmid, AP) as part of a selection system that activates pIII production (the "gIII selection system") in response to the activity of the protein of interest. SP can only propagate by expressing the protein from phage DNA, followed by expression of gIII that is mediated by the protein's activity (Fig. 2B). Thus, *successful SP propagation is linked to the activity of the protein of interest.* SP carrying a mutant protein with enhanced activity will have a fitness advantage over other SP particles, because the enhanced protein activity allows for increased pIII production, thereby increasing offspring production. Over time, SP harboring the coding sequences expressing improved proteins will outcompete others in the population. If phage-dependent activity produces sufficient pIII product, then there will not be a fitness advantage gained from producing additional pIII, which makes fine-tuning the stringency of the selection circuit a key step in developing PACE.

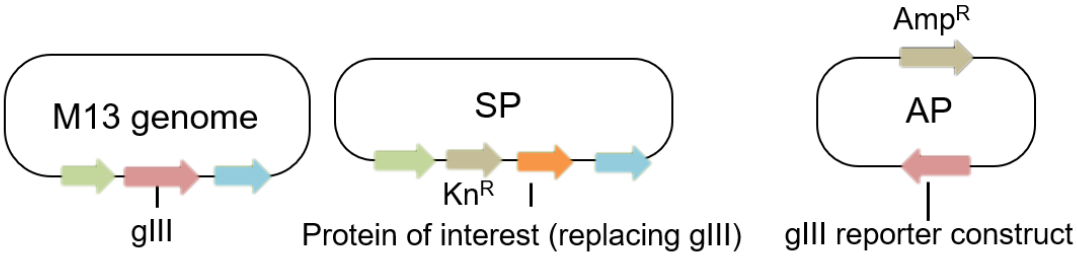


Figure 3. Schematics of the M13 genome, SP, and AP displaying their critical genetic elements. The gIII and protein of interest are shown in brown and orange, respectively. Other phage genes are represented in green and blue. Antibiotic resistance genes are shown in yellow. Swapping the native gIII protein from the M13 genome with the CDS of the protein (and, optionally, Kan^R) renders the SP dependent on the AP in order to be capable of producing pIII to assemble mature M13 bacteriophage[10-13].

Compared to the directed evolution system described previously, the PACE gIII selection system is the equivalent of the "selective assay" that establishes the direction of evolution, and the SP is the vector that expresses the protein of interest. The competition between SP particles carrying mutations in their proteins and selection for their maximum fitness as defined by the gIII selection system is the basis of evolution in PACE.

Continuous evolution in PACE

A defining feature of PACE is the continuous evolution of the SP that can be achieved by the set-up shown in Figure 4. The simplest PACE system consists of two flasks: the Chemostat and Lagoon are connected by tubes driven by two separate pumps. The Chemostat contains a culture of host *E. coli*, incubated at 37 ºC. The first pump continuously delivers fresh media into the Chemostat, while transporting excess media and cells out of the Chemostat to the waste container (Fig. 4, red lines). Thus, the bacterial culture in the Chemostat is maintained under healthy growth conditions, while its volume remains constant. The Lagoon is filled with a culture of host *E. coli* infected with SP; the second pump continuously delivers Chemostat culture into the Lagoon, while the same pump transports excess media, cells, and phage out of the Lagoon to waste (Fig. 4, blue lines). Therefore, the Lagoon is constantly inoculated with fresh, uninfected host *E. coli* for SP particles to infect, while the already-infected cells in the Lagoon exit to waste. Again, the volume of the Lagoon is kept constant throughout the experiment. This system allows SP to propagate in

the Lagoon for hundreds of generations without interruption.

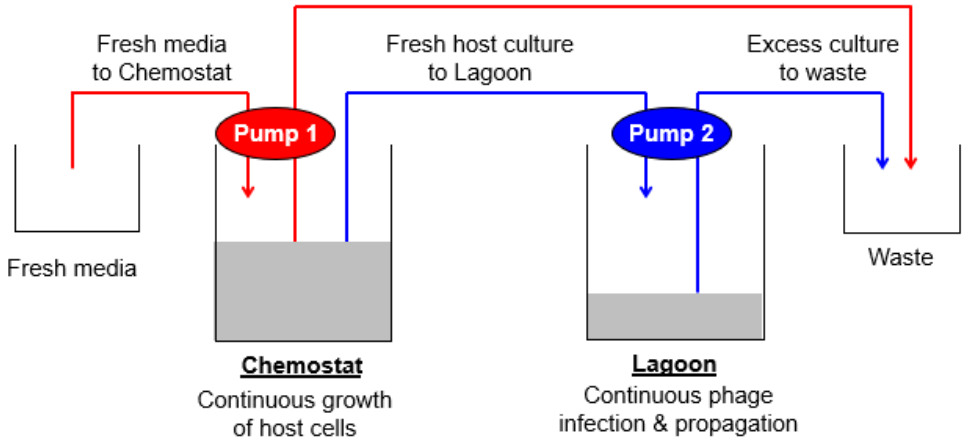


Figure 4. Schematic diagram of the PACE pump system. Host cells transformed with the AP construct are maintained in the Chemostat and continuously pumped into the Lagoon at the optimal flow rate to permit M13 propagation and minimize host replication. By utilizing pumps to regulate the flow of media into/out of the Chemostat and Lagoon, PACE can proceed uninterrupted[10, 13].

The flow rate of the Chemostat culture into the Lagoon plays a key role in PACE, mainly by preventing false positive mutations (mutations that decouple gIII expression from the activity of interest to be evolved) from accumulating in the *E. coli* genome[10-12]. When established properly, the Lagoon culture only allows SP to propagate while preventing host *E. coli* cells from dividing, thereby preventing *E. coli* offspring from lingering in the Lagoon. Typically, the flow rate into the Lagoon is set at more than double the Lagoon volume per hour; e.g., the flowrate of a 30 mL Lagoon may be 80 mL/hr[10]. At this rate, the culture in the Lagoon is refreshed in less than 30 minutes, allowing *E. coli* cells to linger in the Lagoon for less than a half hour. Since the fastest doubling time for *E. coli* is approximately 20 minutes, an average uninfected cell in the Lagoon has barely enough time to divide; however, host cells infected with M13 phage will experience a significant increase in their doubling times[14]. This limits the ability of the host *E. coli* to produce progeny that persist in the Lagoon, which prevents mutations from accumulating

in their genome that might remove the selective pressure established by the gIII selection system by allowing for pIII expression without the desired protein activity.

The challenge for the phage is to infect a host, replicate, and detach from the host within a predetermined time frame to avoid being washed out of the Lagoon. Increasing the flow rate through the Lagoon is one way to increase the stringency of the PACE system, but it is recommended to switch to a more stringent selection circuit rather than to rely on flow rates into the Lagoon to alter the stringency of the selection[19]. As a result of all these factors, PACE allows only mutations in the phage DNA (SP genome) to accumulate, since *phage DNA is the only genetic material that can pass to future generations in the Lagoon*. This limits the amount of "cheating" that can occur within the Lagoon as a result of activity-independent expression of pIII; however, cheater phage can still arise that have recombined gIII or some other component of the selection circuit into its genome that will express gIII outside of the host circuit. Minimizing the number of recombination hotspots in the phage genome can decrease the likelihood of this occurring, and having phage that robustly activate the selection circuit through the intended mechanism may help to avoid enriching the recombinant phage within the Lagoon.

During PACE, the mutagenesis rate during DNA replication is greatly increased using the MP (Mutagenesis Plasmid) that increases the rate of protein evolution[12]. This is accomplished through the arabinose-dependent expression of a dominant-negative variant of the *E. coli* DNA Pol III proofreading domain on the MP, alongside other genes that facilitate mutagenesis[12, 20]. Currently, the most effective variant of the MP is MP6 (available from Addgene), which increases the mutagenesis rate to >300,000-fold over the basal level of mutation in *E. coli*[20]. This high mutagenesis rate allows for the rapid accumulation of beneficial mutations in the protein after arabinose induction of MP. In theory, PACE evolutionary experiments yield mutations from the

very first generation, with enrichment of fitness-improving mutations within 1-3 days of continuous propagation. PACE experiments usually run for several days and multi-experiment evolutionary trajectories have exceeded 20 days. PACE duration depends on the nature of the desired changes and the evolutionary trajectories the protein takes to achieve the desired function[21].

Workflow for building the PACE system

This protocol aims to be a guide for establishing a basic, economical PACE system on a standard lab bench. Our protocol is written for researchers with strong foundations in molecular biology and microbiology. A solid understanding of evolutionary concepts, such as selective pressure and fitness is a useful asset while running and troubleshooting a PACE evolution.

The workflow for developing a PACE system is shown in Figure 5. We describe this development in two separate steps: *1)* developing the gIII selection system, and *2)* installing the Chemostat-Lagoon pump system for PACE evolution. We roughly estimate that it takes 1-2 years for a lab to complete this workflow from scratch, although the timeline will vary depending on the lab's experience with PACE, the complexities of the protein of interest, and the existence of selection assays that are valid for the protein of interest. Adapting an existing selection scheme to a moderately different evolutionary goal can be completed in a much shorter time.

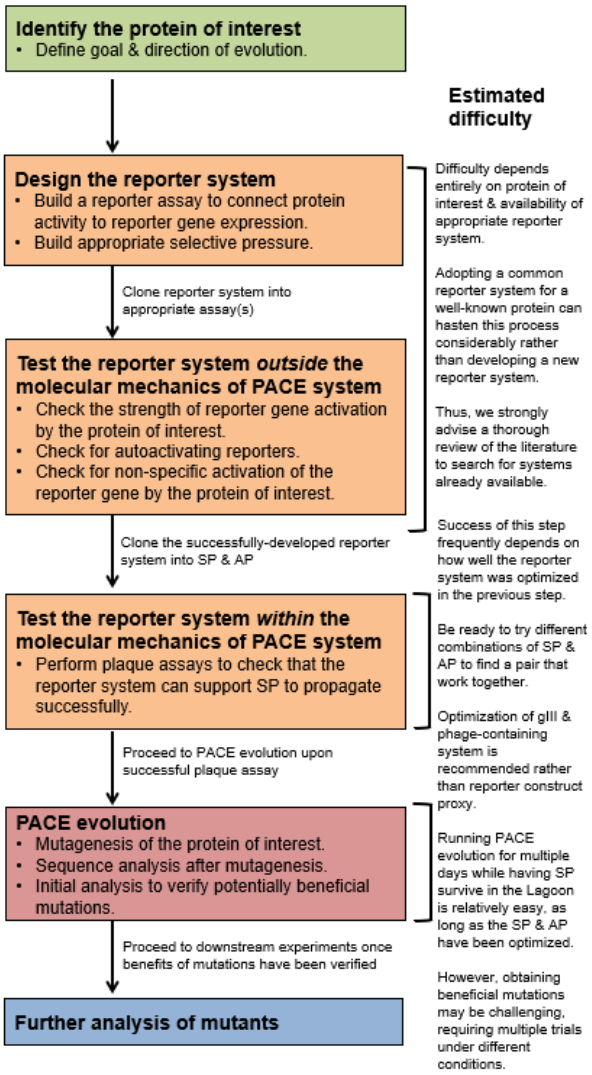


Figure 5. Typical workflow for the development of a PACE selection circuit. Before embarking on adopting PACE to evolve a protein, considerations must be made as to what is the goal of the evolution, the availability of a well-characterized selection circuit, the nature of the selective pressure, and whether it is worth it the effort invested to carry out refinement of various systems before obtaining a successful PACE circuit. Designing and constructing the selection could take anywhere from 0.5-1.5 years to achieve depending on how much troubleshooting needs to take place. Evolving the protein of interest in PACE once the system is set up could take up to a year to carry out the actual experiment, interpret the results, and examine potentially beneficial mutations.

The two major steps of PACE development will be described in separate sections. In the first section, the development of a PACE selection circuit will be described. It is impractical to describe the development of all possible circuits, because the molecular mechanism for a circuit is

different for each protein of interest. Instead, we will use the development of the PACE-B1H gIII selection circuit as an example to demonstrate the key concepts during this step. In the second section, we will discuss the development of the Chemostat-Lagoon pump system in detail, followed by a brief section describing the expected results from PACE evolution. This protocol will focus on the basic PACE pump system. Protocols for more complex PACE will not be provided, although they will be discussed at the end of this paper.

The Liu lab has also developed an intermediate technique between plate-based selections and PACE called phage assisted non-continuous evolution (PANCE) that can be developed and utilized in tandem with PACE[19, 21]. PANCE can use the same selection circuit that PACE utilizes, but instead of media continuously flowing from a Chemostat into the Lagoon, we subculture the contents of the Lagoon into fresh media[22, 23]. This makes PANCE slower than PACE at developing mutations, but PANCE has the advantage of not requiring a specialized set-up. PANCE is useful for preliminary evolution of a protein of interest to obtain variants with high enough activity to support PACE, which requires a certain minimum propagation efficiency to sustain phage titers. PANCE also demonstrates high utility as it can easily be parallelized, and it can be used to assess the feasibility of developing a selection circuit before committing significant time and resources to adopting full-blown PACE[23].

Protocol #1: Developing a PACE selection system

Establishing a proper PACE selection circuit is key to successful PACE. Here, we discuss guidelines for the construction of a PACE selection circuit by using the development of our PACE-B1H as an example. The PACE selection circuit comprises a gIII cassette (located on the AP) that is activated specifically by the protein of interest (expressed from the SP genome)[12]. The desired

protein activity leads to pIII production, which allows for production of next-generation phage particles (Fig. 2B). When developing the selection circuit, three major aspects must be considered: *1)* tuning the amplitude of gIII activation to yield sufficient pIII expression required for phage propagation; *2)* confirming that the gIII cassette is not auto-activated by endogenous molecules, i.e., that the system has low background; *3)* verifying that the gIII cassette is only activated through specific interaction with the protein of interest, i.e., that the system has high specificity. Typically, development of the PACE selection circuit begins by identifying both the gIII cassette and protein of interest that together produce a strong signal. Then, this promising pair is investigated to verify specificity of interaction. The goal is to produce a gIII cassette-protein pair that produces a strong signal (high phage propagation characterized by plaque forming units, PFU) or high signal-to-noise ratio (strong signal relative to background from an indirect reporter like luciferase).

Initial development of the selection circuit may be done independently from the phage-*E. coli* infrastructure. The phage-*E. coli* interaction is an already complicated system on its own, making this system unsuitable as an environment for the construction of a unique selection system. Instead, the selection circuit should be developed using a reporter gene in lieu of gIII that is more straightforward: these reporter genes include LacZ, luciferase, and HIS3. Once the reporter system has been developed "outside" of PACE, its components will be cloned into the AP and SP to test whether those components can properly control gIII expression and phage propagation.

Alternatively, you could develop an AP construct that has a reporter cassette where the stop codon of gIII overlaps with the start codon of the reporter; e.g., the Liu lab utilizes LuxA/B as the reporter, and the gIII stop codon overlaps with luxAB by 1 base (TAATG)[19, 21, 24]. Given the overlapping nature of this construct, expression of luxAB depends on gIII transcription

and translation such that luciferase activity can be used as an indirect means to monitor changes in gIII expression.

Goal of SP-ME47 PACE

In our PACE-B1H system, our protein of interest is ME47, a rationally designed basic helix-loop-helix (bHLH) protein comprising 68 amino acids that specifically targets the E-box motif (5'-CACGTG)[25]. Our goal was to use the molecular mechanisms of the B1H system to make ME47 control expression of an E-box-regulated gIII gene to generate ME47 mutants that bind to the E-box with higher affinity and specificity. A mutation in the ME47 sequence that improves its E-box binding capability should increase gIII expression and give the SP carrying the mutated ME47 a selective advantage over other SPs in the population. Over the course of PACE, SPs carrying beneficial mutations would then outcompete other SPs and dominate the Lagoon population. These mutations can then be isolated for further analysis.

Developing the B1H for ME47: initial screening

The bacterial-one-hybrid (B1H) system was developed by the Wolfe lab to assay the affinity and specificity of a given protein-DNA interaction in *E. coli*[26-28]. At the molecular level, the B1H functions similarly to related systems in yeast (Fig. 6A)[29].[29] Briefly, in the B1H, the protein of interest is expressed from the pB1H2w2 vector as a fusion protein to the RNA polymerase omega subunit: this fusion protein, therefore, behaves as a transcriptional activation domain (AD)[26-28, 30]. The system uses HIS3 as the reporter gene, which is located on a second vector, pH3U3[26-28, 31]. HIS3 is a yeast gene that encodes an essential enzyme in the histidine biosynthesis pathway; activation of this gene allows the host cell, *E. coli* US0 (*ΔhisBΔpyrF*) whose native *hisB*

gene (*hisB* is the bacterial analog of the yeast HIS3 gene and will be used to refer to the HIS3 reporter gene) was deleted, to grow in histidine-deficient media[26-28]. To test how strongly the AD construct interacts with the cognate DNA to mediate *hisB* expression, the doubly transformed US0 cells are grown in histidine-deficient media supplemented with various concentrations of 3-Amino-1,2,4-triazole (3-AT), an inhibitor of the *hisB* gene[32]. Thus, in the B1H, the strength of the protein-DNA interaction is represented by the amount of bacterial growth when challenged to grow in histidine-deficient media.

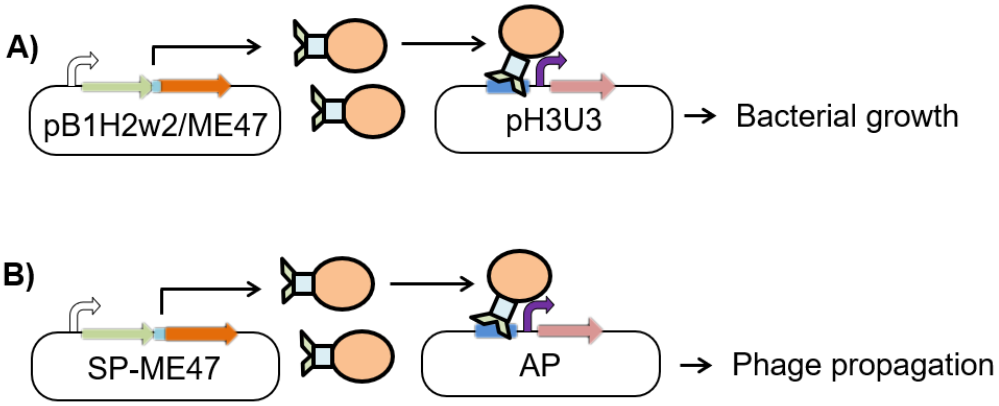


Figure 6. Schematic diagram illustrating **A)** B1H system and **B)** PACE-B1H system. The omega subunit of RNAP fused to ME47 (green and orange, respectively) acts as an activator domain (AD). ME47 binds to the E-box (blue) present on the pH3U3/AP vector and positions the omega subunit of RNAP so that it can recognize the -10/-35 elements of the weak *lac* promoter (purple). This leads to expression of the downstream gene (brown) that is needed for *E. coli* growth and phage propagation[26-28].

To construct the B1H reporter system in *E. coli*, we cloned the sequence expressing the ME47 bHLH into the pB1H2w2 vector; the expressed protein contains ME47 fused to the N-terminus of the RNA polymerase omega subunit separated by a 12 amino-acid linker (Fig. S1). Similarly, we cloned a variant of the E-box sequence, the AT-E-box (5' **A**C<u>CACGTG</u>G**T**, core E-box underlined, flanking AT base pair in bold), upstream of the *hisB* gene in the pH3U3 vector, with 13 nucleotides separating the 3' edge of the AT-E-box and the 5' edge of the -35 weak

promotor. We chose these parameters as a starting point and used materials already available at the time of cloning. Subsequently, we doubly transformed E. *coli* US0 with pB1Hw2w/ME47 and pH3U3/-13 AT-E-box-containing plasmids. Unfortunately, these doubly transformed host cells did not grow in histidine-deficient media, indicating that this B1H pair (ME47 vs. -13 AT-E-box) was incapable of reporter gene activation (Fig. S2).

In our search for a protein-DNA pair capable of producing a B1H signal, we adjusted two features in the *hisB* reporter gene. First, we used an alternate E-box variant as the target sequence for ME47. The new sequence, the GC-E-box (**G**CCACGTG**C**), has flanking GC nucleotides in place of the AT nucleotides in the AT-E-box. Nucleotides in this position were shown previously to influence the protein-DNA interaction for this class of DNA binding proteins[33], and therefore, we surmised that this alteration might achieve a successful B1H pair. Second, it is known that the relative position of the target DNA sequence (in our case, the E-box) within the rest of the *hisB* promoter influences the strength of reporter gene activation. To account for this, we generated variants of *hisB* reporters containing different numbers of nucleotides between the edges of the E-box and the -35 promoter. Thus, we generated a library of E-box reporters that contained either the AT- or GC-E-box, 7 to 25 bp away from the -35 weak promoter (Fig. S2).

We tested ME47 against the various GC- and AT-E-box reporter constructs (Fig. S2). As before, we saw no growth when ME47 was paired against the -13 AT-E-box. In contrast, other reporters, such as -7 GC-E-box and -11 AT-E-box, gave varying degrees of bacterial growth. Therefore, these reporters do activate the *hisB* gene and were candidates for the PACE-B1H reporter.

Developing the B1H for ME47: testing for autoactivation

Following their identification, we screened the candidate reporters for specificity. First, we checked for autoactivation that can result from endogenous cellular molecules activating the reporter. The candidate reporters were singly transformed into *E. coli* US0, and the B1H assay was repeated. Cells transformed with non-autoactivating reporter constructs will not grow, because these cells lack ME47 to activate the reporter. In contrast, cells transformed with an autoactivating reporter will still grow. The results are presented in Figure S3A. As expected, many cells were incapable of growing: e.g., cells harboring the -7 GC-E-box or -11 AT-E-box. However, other cells—such as those carrying the -9 GC-E-box—showed strong growth in the absence of ME47. Reporters that allow cells to grow without ME47 were eliminated from the candidate list, as they were most likely autoactivating.

Developing the B1H for ME47: testing for specificity

We tested the remaining candidates for specificity of the protein-DNA interaction to ensure that reporter gene activation was controlled by the specific molecular interaction between ME47 and E-box in the reporter gene. To do so, we synthesized a "nonspecific" version of each candidate reporter gene by replacing the central 8 bp sequence with a nonspecific sequence known not to interact with ME47 (NS-DNA, 5' TTCCAAGG[25]). We doubly transformed *E. coli* US0 cells with ME47 and the NS-DNA reporter constructs and then performed the B1H. Figure S3B clearly shows that all candidates exhibit minimal cell growth, even at the highest serial dilution. Compared to the cell growth shown by their E-box-containing counterparts, growths resulting from nonspecific reporters were at least six orders of magnitude lower: we note that although the B1H is not a quantitatively linear assay, this is a strong signal-to-noise ratio, confirming the specificity of interaction between ME47 and E-box reporters in the B1H.

Testing the selection circuit in PACE-B1H

As the final step in developing the PACE selection circuit, we cloned ME47—still fused to the RNA-polymerase omega subunit—into the SP to give SP-ME47. Also, we cloned the four candidate E-box constructs (-7 and -11 GC-E-boxes, -9 and -11 AT-E-boxes) into the AP, upstream of the gIII gene. We obtained mature phage particles of SP-ME47 after transforming the SP-ME47 genome into *E. coli* 1059 and purifying the supernatant from its overnight culture. The SP genome can be constructed and manipulated like a regular plasmid and does not require any special handling[34]. We then transformed the four AP/E-box vectors separately into *E. coli* 1030[35]. There are two strains of host cells used in PACE: S1059 and S1030. Both strains have the F pilus to allow for M13 infection via gIII. The S1059 strain has a constitutive AP that produces gIII regardless of activity from the desired protein from the SP and is a permissive host. The S1030 cell lacks the constitutive AP of S1059 and needs to be transformed with an AP, whose design is based on the initial screening process outlined earlier. Recently, S2060 cells were developed as an updated version of the S1030 host. S2060 contains a lacZ gene under the control of a phage shock promoter that produces blue plaques when X-gal/Bluo-Gal are used in the plaque assay. This facilitates the identification of plaques and is recommended for use in the plaque assay. S1059 cells can be replaced with "S2208" cells that are S2060 cells transformed with pJC175e (the constitutive AP, available from Addgene)[19].

In theory, SP-ME47 should be able to activate the gIII gene on the AP/E-boxes via the omega-ME47 construct (Fig. 6B). To test this, we performed a plaque assay using the purified SP-ME47 phage particles against *E. coli* S1030 transformed with the AP/E-boxes. Activity-dependent plaquing is a stringent test for propagation, but it is not quantitative. However, overnight propagation tests will give a quantitative measure of how fit the phage are on the developed selection circuit[35]. Typically, the propagation tests are done with wild-type gIII-containing

phage, active phage (contain the protein of interest cloned into the SP), and empty phage with gIII deleted. For this discrete assay, cultures containing the appropriate AP are grown to mid-log phase in Davis rich medium and infected with a sub-saturating amount of SP-ME47 for 10 min at 37 °C (typically 10^6 PFU/mL is adequate). It is important that a sub-saturating amount of SP be used, as not all the host cells will be initially infected, which allows the phage to propagate and enrich overnight. The culture is then spun down and the supernatant is saved for titering on a plaque assay on S1030 *E. coli* cells carrying the desired AP. To estimate the instantaneous propagation rate of the phage, cells post-infection can be reinoculated into fresh Davis rich medium and grown to OD_{600} ~0.8 rather than overnight, then pelleted, and the supernatant can be saved for titering as described before. The expected results should show $>10^5$-fold propagation for wild-type phage and $\geq 10^3$-fold difference between active and inactive/empty phage, ideally as large a difference as possible. Active phage should enrich at least 10-fold before considering PACE, and 100-fold is safer; otherwise, the continuous flow of PACE will lead to phage washout before fitness-improving mutations have a chance to accumulate and be enriched.

The results are shown in Figure S4. For the -7 GC E-box AP, we observed numerous plaques after the plaque assay, whereas the -11 AT/GC E-boxes did not produce any plaques. This indicates that the -7 GC E-box B1H selection circuit—now harboring the SP and AP—could activate the gIII gene, thereby allowing phage propagation. The successful plaque assay established that our selection circuit was fit for use in the Chemostat-Lagoon pump system. However, if the starting phage propagates too robustly on the selection circuit, there will not be a selective pressure on the protein to improve. We would advise not doing PACE with a circuit that gives $>10^3$-fold overnight propagation on a given circuit, though it may possibly still give useful results depending on the selection that is being carried out.

Protocol #2: PACE & related techniques

Handling of sterile materials

• Autoclaving should be done under standard sterilization conditions. Adjust the sterilization time as appropriate to account for large volumes of media.

• Use 0.22 µm filters for filter sterilization.

- Note that these filters are used to remove bacterial cells (and larger particles) from the solution. Phage particles will pass through these filters.

• Use standard aseptic techniques to handle sterile materials. We regularly use a Bunsen burner or laminar flow cabinet for this purpose.

Recipes for general media & stock solutions

2x YT Top Agar

3.1 g 2x YT media
0.75 g Agar
100 ml ddH_2O
Autoclave to sterilize

Davis rich medium (DRM) base solution

35 g Anhydrous potassium phosphate, dibasic
10 g Potassium phosphate, monobasic
5 g Ammonium sulfate
5 L ddH_2O
Autoclave to sterilize

Alternatively, pre-mixed DRM is available from US Biological, CS050H-001 and CS050H-003.

DRM supplement solution

90 g D-Glucose
10 g Sodium citrate, tribasic
0.52 g Magnesium sulfate (heptahydrate)
10 g Casein hydrolysate
0.5 g L-leucine
500 mL ddH_2O
Filter sterilized

Note: An updated recipe for DRM is currently available for use, and we invite others to explore it[17]. However, we have routinely using this older recipe for its simplicity and consistency. For large quantities of DRM supplement, it is feasible to dissolve chloramphenicol/other antibiotics directly in the supplement before filter sterilizing.

Arabinose solution (1.5 M)

11.25 g L-Arabinose
Dissolve in 50 mL ddH_2O

Note: Introducing 1.5 M arabinose solution at 0.5 mL/hr to a Lagoon receiving bacterial culture at 100 mL/hr will result in a final arabinose concentration of approximately 7 mM in the Lagoon.

LB

10 g Tryptone
10 g NaCl
5 g Yeast extract
1 L ddH_2O
Autoclave to sterilize

Ampicillin (1000x stock)

50 mg/mL solution in H_2O, filter sterilized

Kanamycin (1000x stock)

30 mg/mL solution in H_2O, filter sterilized

Chloramphenicol (1000x stock)

25 mg/mL solution in anhydrous EtOH

Description of bacteria, phage and plasmids

Bacteria

Escherichia coli DH5α (F^-)

- Used for general molecular biology purposes, such as cloning.
- Does not contain the F plasmid (F-); will not be infected by M13 phage and derivatives.
- Can be replaced with other generic lab strains of *E. coli*.

E. coli S1030 (F^+, no AP, no MP)

- Host strain for PACE.
- Carries the F plasmid, allowing infection by M13 phage and derivatives.
- Lacks AP; by itself, does not support propagation of SP.
- Lacks MP; by itself, does not introduce increased mutagenesis rate.
- Amp^S, Kan^S, Cm^S, Tet^R.

E. coli S2060 (F^+, no AP, no MP)

- Host strain for PACE.
- Carries the F plasmid, allowing infection by M13 phage and derivatives.
- Lacks AP; by itself, does not support propagation of SP.
- Lacks MP; by itself, does not introduce increased mutagenesis rate.
- Amp^S, Kan^S, Cm^S, Tet^R.
- *E. coli* S2060 possess a LacZ gene under the control of a phage shock promoter that will make visualizing plaques easier when grown on Xgal-containing media[19].

E. coli S1059 (F^+, + $AP_{constitutive}$, no MP)

- Host strain used for constitutive propagation of SP.

- Carries the F plasmid, allowing infection by M13 phage and derivatives.
- Carries AP with constitutively active gIII gene, thereby allowing SP propagation regardless of the SP genotype.
- Lacks MP; by itself, does not introduce increased mutagenesis rate.
- Amp^R, Kan^S, Cm^S, Tet^R.

E. coli S2208 (F^+, + $AP_{constitutive}$, no MP)

- Host strain used for constitutive propagation of SP.
- Carries the F plasmid, allowing infection by M13 phage and derivatives.
- Carries AP with constitutively active gIII gene, thereby allowing SP propagation regardless of the SP genotype.
- Lacks MP; by itself, does not introduce increased mutagenesis rate.
- Amp^R, Kan^S, Cm^S, Tet^R.
- *E. coli* S2208 possess a LacZ gene under the control of a phage shock promoter that will make visualizing plaques easier when grown on Xgal-containing media[19].
- Can be made by transforming S2060 cells with pJC175e.

Phage

Selection Phage (SP)

- Mutated derivative of the M13 bacteriophage.
- Will infect any F^+ *E. coli.*
- Lacks one gene essential for bacteriophage propagation (gIII) that has been replaced by the CDS of the protein of interest.
- gIII needs to be provided by the host *E. coli* strain, typically on the AP as a gIII reporter gene activatable by the protein of interest.
- Phage propagation depends on the protein (expressed from SP genome) successfully activating the gIII reporter gene found on AP.
- Carries a Kan^R gene upstream of the cloning site for the gene of interest. SP without KanR can be used to increase speed of phage propagation by decreasing the size of the phage genome.
- Techniques to assemble the SP genome via Golden Gate cloning are outlined in the Supplementary Information of Thuronyi *et al*., 2019 and the genome fragments used for this purpose are available from Addgene[34].

Plasmid vectors

Accessory plasmid (AP)

- Carries gIII reporter gene.
- Amp^R (typically).
- Low-copy SC101 origin (typically).
- Plasmids with other characteristics, or multiple plasmids, may be needed to implement a given selection circuit.

Mutagenesis plasmid (MP)[20]

- Carries the genes necessary for increasing the mutagenesis rate during PACE.
- Mutagenesis is induced by arabinose and suppressed by 25 mM glucose.

- Cm.
- **VERY IMPORTANT:** Cell strains hosting MP should always be handled carefully and grown in a media containing 25 mM glucose to suppress the mutagenic activity of MP. Avoid serially propagating these strains as much as possible, as this may cause accumulation of mutations in the host genome.

Handling bacteria

Bacterial cultures

Overnight cultures of *E. coli* strains are prepared by inoculating a single bacterial colony in LB containing appropriate antibiotics and incubating for 16 hr, 37 °C, with shaking at 200 RPM. A loopful of frozen bacterial glycerol stock can also be used to start an overnight culture. Similarly, an overnight culture can be started from a plaque of SP-infected *E. coli*; a sterile pipette tip can be used to stab the plaque to transfer material into the culture tube.

Transformation

All *E. coli* strains in this protocol can be transformed using common transformation methods, such as chemical transformation and electroporation.

Note: Attempts to doubly transform *E. coli* S1030 with AP and MP using chemical transformation may suffer from very low transformation efficiency. We advise to first transform *E. coli* 1030 with AP alone and to make competent cells from the AP-transformed *E. coli* S1030. MP can be easily transformed into chemically competent *E. coli* S1030 + AP.

Storing bacterial glycerol stocks

1. Mix an overnight culture of the bacterial strain 1:1 (v:v) with sterile, 50% glycerol. The final concentration of glycerol will be 25%. Mix well.
2. Store this mixture at -80 °C. The glycerol stock must always be kept frozen to retain maximum viability.

E. coli can be stored this way for years without completely losing viability, provided the stock is kept frozen. Bacteria infected with M13 phage (or SP) can be stored in the same way without losing their capacity to produce and release phage particles.

Handling phage

Note on disinfection

2% bleach solution should be used to disinfect equipment and surfaces that come into contact with SP. 70% EtOH is *not* recommended for this purpose.

"Plasmid-like" properties of SP genome

The SP genome is a circular, single-stranded DNA inside the bacteriophage protein shell. When SP infects *E. coli*, the SP injects its genome into the bacterial cytoplasm, and the genome is synthesized into a circular double-stranded DNA. This circular-double stranded form of the SP genome can be handled similarly to a regular plasmid. (From here, the acronym "SP genome" will indicate the circular, double-stranded form of the phage genome.)

Thus, it is possible to

- Extract and purify the SP genome from a culture of SP-infected *E. coli* host using

common miniprep protocols and commercially available kits.

- Modify the SP genome by conventional methods, such as restriction cloning, USER cloning, Golden Gate cloning, and site-directed mutagenesis.
- Transform the SP genome back into *E. coli* using common transformation methods.
    - When the SP genome is introduced this way into an *E. coli* host suitable for phage propagation, the SP genome will function normally, producing the phage proteins and protein of interest, eventually leading to production and secretion of mature phage particles to the environment.
    - Transformation efficiency of this procedure is usually very low. While this is not a problem when transforming with a uniform clone of the SP genome, it may cause difficulties when developing libraries of SP *in vitro*.

Note on isolating plasmids from E. coli 1030 or E. coli 1059

E. coli strains S1030 and S1059 infected with the SP genome carry multiple plasmids, such as the AP, SP genome, MP, and F plasmid, which will not be differentiated during plasmid purification. Therefore, the extracted plasmid solution will contain a mix of these plasmids, and further processing (such as transforming *E. coli* DH5α with this solution followed by replica plating) is necessary to obtain these plasmids in pure form. Alternatively, the SP can be isolated using commercially available kits that are specifically made for this purpose.

Handling cultures of E. coli host infected with SP

Since M13 (and its derivative SP) do not kill the cells in order to propagate, SP-infected *E. coli* can be used to start an overnight culture to produce more SP-infected cells. The supernatant of this culture will contain mature phage particles released from the cells. Typical phage titers from overnight growth of activity-independent propagation host cells are 10^8-10^{12} depending on the phage propagation efficiency.

Crude isolation of phage particles from an infected E. coli host culture

1. Start an overnight culture of the appropriate, SP-infected *E. coli* strain in LB + Amp.
2. Centrifuge the overnight culture at 3000 x g, 5 min, to isolate the *E. coli* cells.
3. Pass the culture supernatant through a 0.2 µm syringe filter to completely remove any *E. coli* cells remaining in the supernatant.
4. The supernatant is ready to be used as a source of purified phage.

Storage of purified phage particles

1. Mix the solution of purified phage with sterile glycerol to a final glycerol concentration of 50%. Typically, we add 1.7 mL purified phage solution to 3.3 mL sterile 75% glycerol to obtain ~50 % final glycerol concentration.
2. Store the phage solution in 50% glycerol at -20 °C. Ensure that the solution does not freeze.

The phage solution can be stored this way for at least a week without affecting the viability of the SP particles. Others have reported that SP particles can be stored this way for years. Alternatively, purified phage solutions can be stored at 4 °C for months to years, although the viability of SP particles will slowly diminish. Thus, for long-term storage to ensure the safety of your samples, we advise keeping a glycerol stock of the SP genome in addition to other methods.

Plaque assay

Materials:

- Overnight culture of the host *E. coli* strain.
- Purified phage solution.
- 4 mL aliquots of molten 2x YT top agar.
    - On the day of the plaque assay, melt the top agar using a microwave oven and aliquot 4 mL into sterile culture tubes. Keep the aliquots at 50 °C until use.
- LB plates.
    - Pre-warmed before use. Cold plates make it difficult to spread the top agar.

Procedure:

1. Make a 1:50 dilution of the overnight *E. coli* culture in 5 mL fresh LB broth + Amp.
2. Grow the diluted culture at 37 °C, 200 RPM, until OD_{600} 0.6-0.8. This typically takes 2-3 hr incubation.
3. Perform a serial dilution of the phage stock solution in ddH_2O until 10^{-6} dilution. We typically perform the dilution in 200 µL total volume per tube.
4. Separately transfer 30 µL serially diluted phage solution into fresh 1.5 mL tubes.
    - **Note:** These aliquots may be kept at room temperature for 30-60 min without adverse effects. Samples may be kept on ice for longer waiting periods.
5. Once OD_{600} of the *E. coli* culture reaches 0.6-0.8, add 270 µL culture to each of the tubes containing 30 µL diluted phage solutions. Incubate for 10 min to allow phage to infect host *E. coli* cells.
    - **Note:** Incubation times can be varied from 5-20 min without adversely affecting the result, *although extra precautions should be taken if the absolute quantity of plaques is important.*
    - **Note:** Alternatively, cultures of host cells grown to OD_{600} 0.6-1.0 can be stored at 4 °C for up to two weeks. Stored cells can be used in the plaque assay straight from the refrigerator after vortexing to resuspend the cells.
6. After incubation, transfer all 300 µL phage + *E. coli* mixture into a tube of 4 mL molten top agar. Mix well and immediately pour everything onto a LB plate. Allow the top agar to spread evenly across the plate. Agitation/tilting of the plate while the top agar is solidifying may cause clumping.
7. The top agar will solidify within minutes. Incubate the plate at 37 °C, overnight.

Instead of using a standard 100 mm petri dish, plaque assays can be poured using quarter well petri dishes and culture plates of various sizes to save space and reagents. Reagent volumes should be adjusted accordingly.

Isolating a single clone of phage particle from a solution of phage

1. Perform a plaque assay of the phage stock using *E. coli* 1059 as the host.
2. You should get plaques forming on the plates the next day. It is reasonable to assume that an individual plaque represents a uniform clone of phage particles.
3. Using a sterile pipette tip, transfer a single plaque from the plaque assay to 5 mL fresh LB + Amp. Grow the culture for 16 hr, 37 °C, 200 RPM.
4. The overnight culture can be used for subsequent analysis, such as plasmid miniprep, PCR, RCA (rolling circle amplification), or DNA sequencing. Picking plaques from a plate and

using directly for downstream steps is not recommended, since only low amounts of phage DNA may be present. We recommend outgrowth of plaques, which refers to allowing plaques to grow overnight to increase the initial amount of phage genome.

MP Check

Due to its strong mutagenic activity, MP stored as a glycerol stock in *E. coli* can still mutagenize DNA, including itself, even when the storage media is appropriately supplemented with 25 mM glucose. Therefore, we advise checking the mutagenic activity of MP each time before a PACE experiment to ensure that the plasmid is functioning properly. This procedure is called the "MP Check."

1. Prepare a pair of culture tubes containing *E. coli* S1030 that will be used for the PACE experiment in 5 mL LB with the appropriate amounts of Amp and Cm added.
2. Supplement the first tube with 25 mM glucose. This should suppress MP activity.
3. Supplement the second tube with 1 mM Arabinose. This should induce MP activity.
4. Incubate both culture tubes overnight and observe for bacterial growth the following day.

Note: The MP check can also be carried out by plating serially diluted cells onto 2xYT plates containing Amp+Cm that have been supplemented either with glucose or arabinose, as has been described[20].

After the overnight incubation, the culture tube supplemented with 25 mM glucose should have saturated bacterial growth, as expected. In contrast, there will be no growth in the tube supplemented with arabinose if the MP is functioning properly, since the mutagenic activity from a fully induced MP will quickly destroy the genetic integrity of the host cell, hampering bacterial growth. No growth in the MP Check confirms that the *E. coli* strain is ready for PACE. In contrast, observing growth in the arabinose-supplemented tube may mean that the MP has lost its functionality, or that the arabinose concentration was too low.

The MP Check should be done before setting up PACE evolution to avoid terminating the experiment if the check fails. Alternatively, the MP Check can be performed on the exact same *E. coli* cells that were used to prime the Chemostat flask; this can be done easily by setting up the MP Check as you also set up the Chemostat. Either way, performing this simple test on a routine basis will help maintain the integrity of a much longer experiment. MP checks can be conducted periodically on Chemostat cells to verify their integrity over the course of a long PACE.

PACE Checklist

A shopping list of the various parts listed in the selection below can be found in Figure S15, along with the corresponding part numbers to facilitate building your own PACE setup.

Pumps, shakers, & incubator

- A pair of peristaltic pumps, one each for the Chemostat and the Lagoon (Fig. S5A).
- A syringe pump for pumping arabinose (Fig. S5B).
- An air pump to aerate the Chemostat (optional).
- A pair of magnetic stirrers, one each for the Chemostat and the Lagoon.

- A tabletop incubator with an access hole on one of its sides.

Assemble the pumps as per the manufacturer's instructions. In many cases, the speed dials of the peristaltic pumps must be calibrated to obtain the desired flow rate (done by assembling the pump with their proper tubing and testing the flow with water). We advise you to become accustomed to assembling, disassembling, and operating the pumps before installing them in the incubator. Then, assemble these components in the incubator (Fig. S5C).

Tubes (Assemble tubes according to Fig. S6. Pictures of these tubes are shown in Figure S7.):

- Media to Chemostat.
- Chemostat to Lagoon.
- Chemostat to Waste.
- Lagoon to Waste.
- Air to Chemostat (if using).
- Air to Lagoon (if using).
- Air pump to Air filter (if using).
    - **Note:** it is possible to run PACE anaerobically with vents included in the set-up to relieve pressure buildup.
- Arabinose filter to Lagoon.
- Arabinose pump to Arabinose filter.

Chemostat & Lagoon flasks (Fig. S7):

- two 125 mL flasks with stirring rod inside.
- two silicone-tube caps.

Media Bottles:

- two 5 L bottles of DRM (Fig. S8A). You may use larger bottles if your autoclave has the capacity.
- two tubed-bottle caps (Fig. S8B).

Note: All tubes, flasks, caps, and media must be autoclaved before use. The open ends of these components should be covered with aluminum foil for autoclaving (Fig. S9).

Bacteria and phage:

- Colonies of *E. coli* S1030 freshly transformed with the appropriate AP and MP, with MP activity confirmed by MP Check. To prevent accumulation of mutations in the host genome, cells transformed with MP should not be stored for more than 2 days at 4 °C after the initial overnight incubation at 37 °C.
- Purified solution of the desired SP. A glycerol stock of the SP will work as well.

Disposables:

- 0.22 µm syringe filters.
- 60 mL disposable syringes.
- 1" sterile needles.
- 5" sterile needles.

- 1000x Ampicillin.
- 1000x Chloramphenicol.

Note: Here, we use disposable needles and a silicone cap as an easy solution to connect the tubes to the Chemostat and Lagoon. Alternately, re-usable, blunt-end needles can be used with a compatible tube cap.

PACE: Pump parameters & timeline

Chemostat & Lagoon parameters

Chemostat

- culture volume = 80 mL.
- flow speed of incoming media = 80 mL/hr.
- turnover rate = 1 culture volume/hr.
- flow rate may need adjustment for different host cell growth rates

Lagoon

- culture volume = 30 mL.
- flow speed of incoming media = 60 mL /hr.
- turnover rate = 2 culture volumes/hr.
- generation time = 2 infection cycles/hr.

Note: These parameters are for a 125 mL flask. Make adjustments when using containers of other sizes.

PACE timeline

- Prep-Day1: Setting up the Chemostat.
- Prep-Day2: Setting up the Lagoon and inducing mutagenesis with arabinose.
- Day1: First day after the start of arabinose induction. Sample the Lagoon to detect and analyze surviving SP.
- Day2, onwards: Sample the Lagoon every day for analysis. Experiment may end after Day3 but can be extended for a longer time.

Setting up PACE

Refer to Figures S10 and S11 for a detailed schematic diagram of the system.

Note: Aseptic techniques while setting up PACE.

Typically, brief exposure of sterile tubings and needles to air does not cause contamination in the PACE system. Consciously minimizing the handling time of sterile components, as well as using sprays of 2% bleach, prevents contamination.

Note: Items to check regularly during a PACE run

- All equipment functioning properly (pumps, tubes, stirring rods, air supply, etc.).
- No leaks inside and outside the incubator.

- Chemostat and Lagoon cultures look healthy. Both cultures should have a dense and well-established bacterial growth, but not be overgrown. Chemostat ODs should typically be between 0.1 to 0.8.
- Appropriate amount of media in the DRM bottle.
- Waste containers emptied before overflowing.

Prep-Day1: Installing the Chemostat

1. Let incubator warm to 37 °C.
2. Assemble the DRM bottle (Fig. S8C and S10).
    a. Add 2.5 mL 1000x Amp, 2.5 mL 1000x Cm, and 125 mL DRM supplement to the autoclaved 5 L DRM base solution. Mix well.
    b. Place a tubed-bottle cap on the 5 L DRM bottle and add a 0.2 µm syringe filter to its "filter" tube. The other end of the tube should be kept sealed with aluminum foil until it is connected to the Chemostat.
3. Assemble the Chemostat (Fig. S11B).
    a. Using a sterile loop, transfer a single colony of *E. coli* S1030 + AP + MP into a sterile 125 mL flask. Alternatively, an overnight culture of the bacteria can be used to prime the flask, as shown in Figure S11b. Immediately seal the flask with the silicone cap.
        i. Optional: Set up an MP Check as you assemble the Chemostat.
    b. Connect PACE tubings to the Chemostat using disposable needles (Fig. S11a). After attaching the needle to the appropriate end of each tube, insert the needle into the 125 mL tube through the silicone cap.
        i. The Chemostat end of (1) Media to Chemostat tube connects to a 1" needle.
        ii. The Chemostat end of (3) Chemostat to Waste tube connects to a 5" needle.
        iii. The Chemostat end of (5) Air filter to Chemostat tube connects to a 5" needle.
        iv. The Air filter end of (5) Air filter to Chemostat tube connects to a 0.2 µm syringe filter.
    c. Adjust the height of the 5" needles. The needle for the Chemostat to waste tube should be set at the 80 mL mark of the Chemostat. The needle for the air to Chemostat tube should be set below the needle for the Chemostat to Waste tube.
4. Place the 5 L DRM bottle and a waste container beside the 37 °C incubator beside the access hole (Fig. S11C).
5. Place the Chemostat on top of the Chemostat Stirrer inside the 37 °C incubator (Fig. S11C).
6. Connect the Chemostat to Air Pump (Fig. S11C).
    a. Connect the (8) Air Pump to filter tube to the (5) Air filter to Chemostat tube via the 0.2 µm syringe filter attached to the (5) Air filter to Chemostat tube.
7. Turn on the Chemostat Pump. Turn on the Chemostat Stirrer.
8. Connect the (1) Media to Chemostat tube to the Chemostat Pump (Fig. S11B). Make sure that the tube is flowing toward the Chemostat. Pass the Media end of this tube through the access hole and connect this end to the Chemostat tube of the 5 L DRM bottle.
9. Connect the (3) Chemostat to Waste tube to the Chemostat pump. Make sure that the tube is flowing toward the Waste Container. Pass the Waste end of this tube through the access hole and place the tube over the Waste Container. Seal the Waste Container with aluminum foil.

10. Turn the dial of the Chemostat Pump to start introducing DRM to the Chemostat. Once the flow has been established, allow DRM to fill the Chemostat to the 80 mL mark.
11. Check that the air pump and stirrer are working properly for the Chemostat.
12. Let the bacterial culture grow overnight in the Chemostat with the Chemostat Pump turned on. This will keep the culture healthier overnight, as glucose will not be depleted and the MP will not be induced.

Prep-Day2: Setting up the Lagoon and inoculating it with SP

1. Check that the bacterial culture is growing in the Chemostat. The culture should be at OD_{600} 0.1-0.8 to maintain the culture in log phase that will facilitate phage infection and induction of the error-prone DNA polymerase from the MP.
2. Assemble the Lagoon (Fig. 7).
    a. Transfer 1 mL purified phage solution into a 125 mL flask. Immediately seal the flask with the silicone cap.
    b. Connect PACE tubings to the Lagoon using disposable needles (Fig. 7A). After attaching the needle to the appropriate end of each tube, insert the needle into the 125 mL tube through the silicone cap.
        i. The Lagoon end of the Chemostat to Lagoon tube connects to a 1" needle.
        ii. The Chemostat end of the Chemostat to Lagoon tube connects to a 5" needle. Keep the needle placed in its package as this needle will be inserted into the Chemostat while installing the Lagoon in the incubator.
        iii. The Lagoon end of (4) Lagoon to Waste tube connects to a 5" needle.
        iv. The Lagoon end of (6) Air to Lagoon tube connects to a 1" needle.
        v. The Air filter end of (6) Air to Lagoon tube connects to a 0.2 µm syringe filter.
    c. Adjust the height of the 5" needles. The needle for the (4) Lagoon to Waste tube should be set at the 30 mL mark of the Lagoon.
3. Place a second Waste Container beside the 37 °C incubator, on the side of the access hole.
4. Place the Lagoon on top of the Lagoon Stirrer inside the 37 °C incubator (Fig. 7B).
5. Turn on the Lagoon Pump and the Lagoon Stirrer.
6. Install the (2) Chemostat to Lagoon tube in the Lagoon Pump. Make sure that the tube is flowing towards the Lagoon. Insert the 5" needle on the Chemostat end of this tube into the Chemostat. Position this needle below the needle for the Chemostat to Waste tube.
7. Install the (4) Lagoon to Waste tube in the Lagoon Pump. Ensure that the tube is flowing towards the Waste Container. Pass the Waste end of this tube through the access hole and place the tube over the second Waste Container. Seal the Waste Container with aluminum foil.
8. Turn the dial of Lagoon Pump to start introducing the culture in the Chemostat to the Lagoon. Once the flow has been established, allow the Lagoon to be filled with 30 mL bacterial culture and check that the (4) Lagoon to Waste tube is working properly.
9. Prepare the arabinose solution.
10. Assemble and install the arabinose syringe and pump.
    a. Attach the Filter end of the (7) Arabinose Filter to Lagoon tube to a 0.2 µm syringe filter. Connect the Lagoon end of this tube to the Lagoon using a 1" needle.

b. Attach the Arabinose Syringe end of the (9) Arabinose Syringe to Filter tube to a new 60 mL syringe. Load the syringe with the arabinose solution by dipping the filter end of this tube into the arabinose solution and pulling the plunger of the 60 mL syringe. The arabinose solution does not need to be kept sterile at this point.
c. Attach the Filter end of the (9) Arabinose Syringe to Filter tube to the 0.2 µm syringe filter that is already attached to the (7) Arabinose Filter to Lagoon tube.
d. Install the 60 mL plunger in the Arabinose Pump.
e. Start the arabinose pump to begin adding arabinose solution into the Lagoon. It may take time for arabinose to start dripping into the Lagoon.

11. Take note of the approximate time when you started arabinose induction.
12. Leave the system running overnight.
13. Start an overnight culture of *E. coli* S1059.

Note: It is not necessary to have the 0.2 µm syringe filter in between the syringe and the Lagoon, if the arabinose solution is filter sterilized first and then loaded in the arabinose pump aseptically. Refer to Figure S12 for how the PACE system should look like once fully assembled.

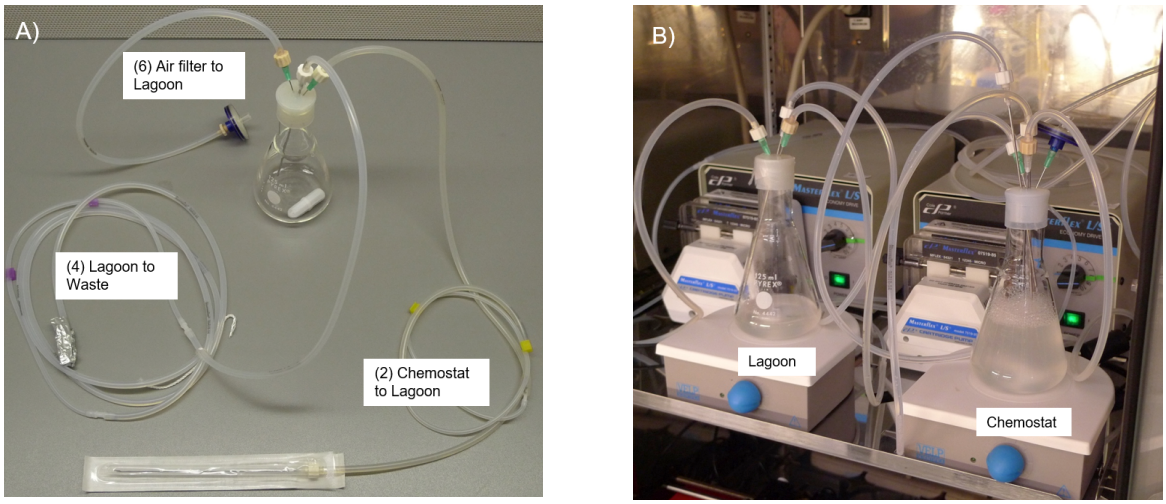


Figure 7. Setting up the Lagoon. **A**) Lagoon flask after assembly. The numbers correspond to Figure S6. **B**) Lagoon flask installed in the incubator.

Day1-Day3 Post Induction: Sampling SP surviving in the mutagenesis-induced Lagoon

1. Into a sterile tube, collect a 3-5 mL sample exiting the "Lagoon to Waste" line. This sample represents the current state of the Lagoon culture.

2. Alternatively, samples can be collected directly from the Lagoon using a syringe. This may be helpful if biofilm formation in the waste line is a concern.
3. Use the sampled Lagoon culture to prepare a purified phage solution.
4. Perform a plaque assay on the purified phage solution using *E. coli* S1059 as the host.
5. Keep the PACE running without any changes.
6. Every day after mutagenesis has been induced, take a sample of the Lagoon culture (as per Step 1 of Post induction) and:
    a. Store a portion of the Lagoon culture as a glycerol stock.
    b. Purify phage from the Lagoon culture to perform a plaque assay.
    c. Store a portion of the purified phage solution in 50% glycerol.
7. PACE can be extended beyond 3 days if desired. The culture within the Chemostat can be maintained for ~4 days before needing to be replaced to avoid having the cells adapt to Chemostat conditions, at which point they may begin to mutate.

Dismantling the PACE set-up

After the PACE experiments, all tubes and flasks must be disinfected with 2% bleach, washed with mild detergent, rinsed with water and dried. To begin the washing process, the pumps can be used to run bleach and water through the tubes and flasks before dismantling; this lowers the chance of phage contamination while dismantling the system. Disinfect the surfaces in the incubator with bleach, as appropriate. Washing the PACE components with MilliQ water will prevent corrosion and will allow for the re-using of the components several times before having to replace them due to wear and tear.

Expected results from PACE

Here, we present the seven-day evolution of SP-ME47 as an example. Details regarding SP-ME47 can be found under the Protocols section.

Experimental parameters of SP-ME47 PACE

- Protein of interest: ME47, a designed bHLH E-box-binding protein, expressed from SP (SP-ME47).
- Reporter system: PACE-B1H with E-box-controlled gIII reporter gene, located in the AP (AP/-7gcEbox).
- Selective pressure: ME47 must bind to its target E-box site in order for the SP genome to survive. Improved E-box binding by ME47 mutants leads to improved survival and propagation of their SPs.
- Expected direction of evolution: ME47 will evolve to bind the E-box sequence with *1)* higher affinity, and/or *2)* higher specificity.
- Induction of mutagenesis: Done using MP6, continually induced at ~7 mM arabinose.
- Duration of evolution: Seven days after starting the arabinose induction of MP6.
- Sample processing and analysis: Lagoon supernatant sampled and analyzed by plaque assay every day. Five plaques selected each day for Sanger sequencing. Refer to Figure 8 for sequencing results.

Evolution of SP-ME47 & analysis of sequencing data

The plaque assay analyzing the Lagoon supernatant one day after arabinose induction ("Day 1") produced plaques at an efficiency of 2.3 x 10^6 PFU/mL. The efficiency increased at Day 2 (7.0 x 10^7 PFU/mL) and Day 3 (6.8 x 10^8 PFU/mL) before stabilizing. Presence of plaques at Day 7 indicated successful phage survival over the duration of the experiment. Five plaques were selected for sequencing analysis for each of the seven days. Multiple sequence alignment (MSA) of the resulting sequences is shown in Figure 8.

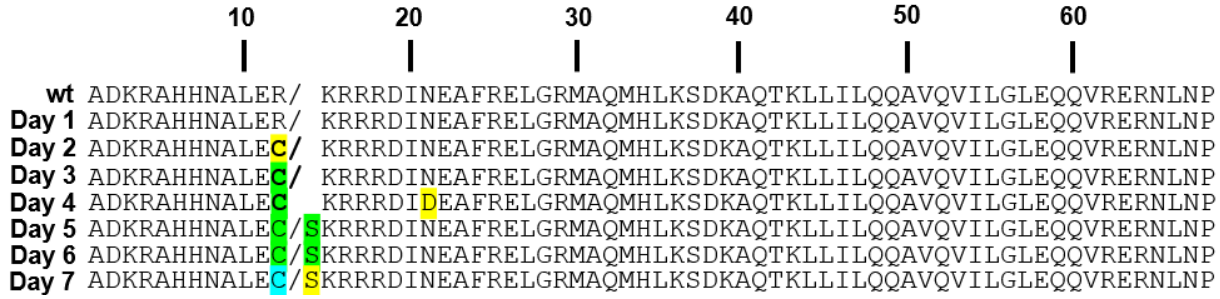


Figure 8. Multiple sequence alignment of ME47 CDS after 7 days of PACE. ME47 sequence corresponding to its basic region is shown. The sequence of original ME47 before mutagenesis is shown on top of the alignment, followed by representative sequences from Days 1-7 after arabinose induction. Mutations highlighted in yellow were present in less than half of the sampled population. Mutations shown in green were present in 50% of the population, while mutations in blue were present in >75% of the population. The R12C mutation was of interest, as it became the predominant variant of ME47 in the Lagoon by Day 6, whereas the N22D mutation did not persist in the Lagoon; we presumed that this was a chance mutation with no impact on phage propagation and did not explore it further.

When analyzing Figure 8, we must distinguish beneficial mutations from spontaneous ones that occur by chance. These beneficial mutations allow their SP to outcompete other SPs in the Lagoon, and therefore SP genomes carrying these mutations should increasingly dominate the Lagoon population over the seven days. In Figure 8, two mutations fit the criteria. Both mutations occur at the same location (residue 12) of the ME47 CDS, where the wild-type residue is R at this location, which was mutated to either C or S. For the first two days of PACE, the wild-type is the

most prevalent in the population (9 of 10 sequences are wild-type in Days 1 and 2). However, the frequency of mutations rises after Day 3, and these mutations take over the population by the last two days (9 of 10 sequences are mutants in Days 6 and 7). This strongly indicates that these ME47 mutations were beneficial for the survival of SP-ME47 in an environment where SP survival is correlated to ME47 activity. Needless to say, priorities were given to these mutants for further analysis.

Additional mutations were found throughout the ME47 sequence, such as those at position 22 in Figure 8. However, this mutation was only found twice across all 35 samples and failed to prevail in the population. This suggests that these mutations were non-beneficial, spontaneous changes that were sampled at the moment the mutation occurred before being purged from the population. Although these data are insufficient to confirm that the additional mutations were not beneficial, we judged these mutations to be of low priority (if any) for further analysis.

Confirming the effects of mutations outside the PACE system

The mutations uncovered by PACE must be evaluated in a separate assay to confirm their effect. Ideally, this test should be done without using material(s) produced from PACE to prevent accidental carryover of phage genome sequences that may have also mutated during evolution. It may be desirable to subclone the evolved genes into a standardized phage backbone, i.e., an SP backbone that has not been put through PACE, in order to separate the effects of gene-of-interest mutations from phage-backbone mutations.

The C-to-A and C-to-T mutations found in the ME47 CDS correspond to R12S and R12C mutations in the ME47 protein sequence, respectively. To further investigate the effect of these mutations, the R12S and R12C mutations were introduced into plasmid pB1H2w2/ME47 via site-

directed mutagenesis. The resulting mutants, pB1H2w2/ME47(R12S) and pB1H2w2/ME47(R12C), were tested in a B1H assay using an E-box regulated *hisB* gene as a reporter gene (see the Protocols section for description of the B1H assay). Results of the B1H assay are shown in Figure 9A. The B1H assay clearly shows that both ME47/R12S and ME47/R12C more strongly upregulate reporter gene activation, as compared to original ME47. There were no observable differences between the transcriptional activation activities of ME47/R12S and ME47/R12C. Furthermore, both ME47 mutants were shown to retain their E-box-specific binding activities towards the E-box sequence, as they do not activate the *hisB* reporter controlled by nonspecific DNA (Fig. 9B). These mutations were introduced directly into the pB1Hw2w vectors via site-directed mutagenesis, and therefore they are the sole reason responsible for the increased performance of ME47. Taken together, these results exemplify a successful usage of PACE to obtain ME47 mutants with a much-improved E-box binding activity.

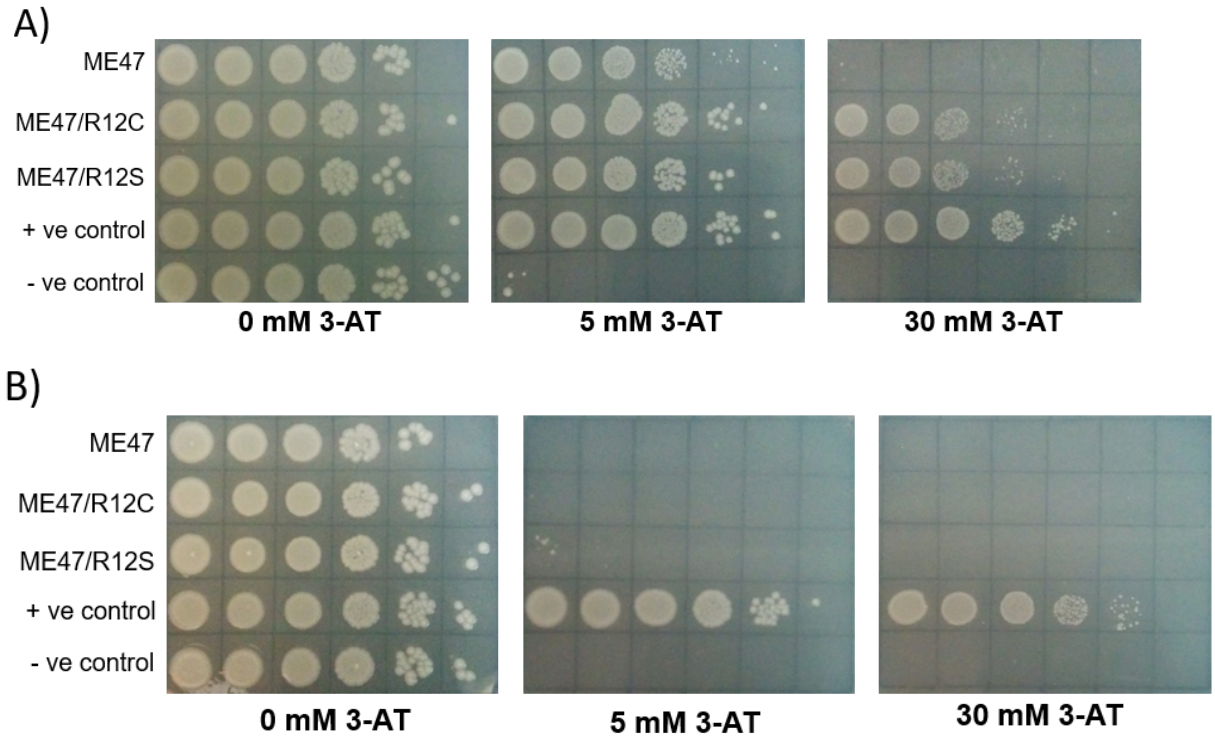


Figure 9. The ME47 mutants generated by PACE outperform original ME47 in the B1H assay. **A**) B1H assay of ME47 and its mutants tested against the *hisB* reporter gene controlled by the -7 GC Max E-box. Both ME47/R12C and ME47/R12S show strong growth until 30 mM 3-AT, in contrast to ME47. **B**) B1H assay of ME47 and its mutants tested against the *hisB* reporter gene controlled by -7 GC NS DNA. Both ME47/R12C and ME47/R12S show minimal growth at 5 mM 3-AT, confirming their specificity towards the E-box.

Additional considerations for developing PACE

Finetuning the PACE B1H selection system

In our PACE-B1H selection system, there are numerous ways to alter the stringency of the system in order to find conditions that are more likely to give rise to mutations in the protein of interest. These include decreasing the amount of pIII that is generated if the system is autoactivating, producing pIII outside of the context of the PACE selection system, and increasing the rate or amount of generated pIII that allows the phage to propagate to avoid being washed out of the Lagoon. As was mentioned earlier, the copy number of the double-stranded viral genome within the host can reach up to ~100 copies, which is important to consider when fine-tuning the amount of pIII present in the circuit[16]. The general strategy is to alter components of the PACE system to change the amount of pIII that is produced as a result of the activity of the protein of interest. Below are several parameters that can be tackled toward altering the PACE-B1H selection system; however, this is not an exhaustive list of changes that can be explored.

<u>Parameters in the B1H selection system that can be tuned to alter stringency</u>

1. Spacer length between cognate DNA and promoter to affect transcription from the AP.
2. Linker length between the omega subunit of RNA polymerase (RNAP) and DNA-binding domain on the SP.
3. Promoter binding capabilities of the omega subunit of RNAP fused to protein of interest.
4. Promoter sequence such that the rate of transcription is changed.
5. Length/composition of the ribosome binding site (RBS) to modify the amount of translation from the resulting mRNA transcript.
6. Codons of the CDS for the protein of interest to adjust gIII expression.
7. Altering substrate availability where suitable.

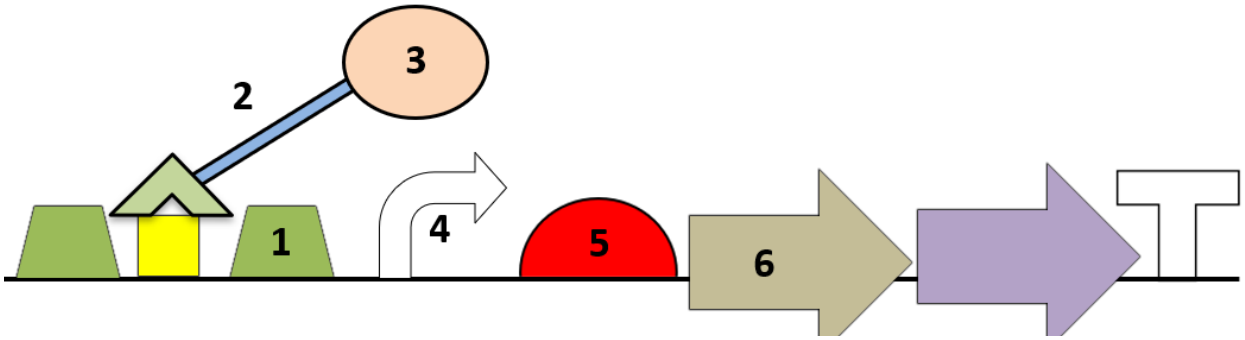


Figure 10. Overview of the B1H-PACE circuit. Schematic depiction of how the accessory plasmid (AP) is organized, and how the AP and protein of interest-omega subunit fusion from the SP interact with one another to affect gIII expression in our B1H-PACE with ME47. The numbers correspond to the list outlining parameters that can be altered to fine-tune gIII expression. In brief, the E-box motif (yellow) is flanked by sequences (blue trapezoid) that can be modified to alter the distance between cognate DNA and promoter. The length of the linker (blue rectangle) between ME47 and the omega subunit of RNA polymerase can be altered to optimally position the omega subunit on the promoter. The promoter sequence can also be altered to give differing amounts of transcription from the RNA polymerase[26, 27]. The RBS sequence (red semi-circle) can be altered to change how easily the ribosome binds to it to affect translation[36]. Altering the first codon of gIII (gray arrow) can also decrease the amount of translation that occurs[36]. A LuxAB CDS (purple arrow) can be placed downstream of the gIII CDS to allow for an indirect means to measure gIII expression.

Strategy 1 involves altering the distance between the cognate DNA that is bound by ME47 (alternatively, could be another DNA binding element) and the leading edge of the weak *lac* promoter[26, 27], while strategy 2 alters the distance between the omega subunit and DNA-binding domain. By altering either of these distances (linker lengths), we can ultimately fine-tune the ease with which the omega subunit of RNAP interacts with the promoter to affect transcription, resulting in an increase/decrease of pIII production[19, 22, 37, 38]. This could also be theoretically achieved by mutating the omega subunit, so that it interacts more strongly/weakly with the promoter; however, this is a significantly more difficult way to tackle the problem and is not recommended[39]. We can also alter the translation of gIII by changing the Shine-Dalgarno sequence (denoted above as RBS or ribosome binding site) to which the ribosome binds[36]. The canonical RBS in *E.coli* is AGGAGG; examples of tactics we can try to alter stringency include

adjusting the amount of translation by changing the consensus RBS sequence, changing the distance between the consensus sequence and starting Met residue, and/or altering the second codon after the initiating Met[36, 40-42]. The following program (https://salislab.net/software/)[43] can be used to predict the amount of translation from a given RBS construct before testing the constructs in the PACE system. The codon usage for the CDS of the gIII protein can be changed by introducing codons that are used more/less frequently; thus, we can observe a corresponding change in the amount of pIII production[44, 45]. Lastly, if the PACE circuit requires evolution to recognize a new substrate, lowering the concentration of the substrate within the Lagoon would increase the stringency of the selection[21]. It may be that higher levels of substrate are needed initially to allow for the accumulation of mutations in a lower stringency environment before transitioning to the higher stringency setting[23].

Possibly, fine-tuning of the PACE selection circuit to obtain an appropriate level of phage propagation may involve altering more than one parameter and this is the most challenging aspect of PACE. A good resource for a similar analysis of possible parameters that can be fine-tuned for a B2H-PACE selection system was outlined in Badran *et al*[19].

Evolution of protein-protein interactions in the context of PACE

PACE can also be used to evolve protein-protein interactions to improve upon existing protein-protein interactions or to create novel interactions. Two strategies have been outlined thus far, the split T7 RNAP strategy[24, 46] and the bacterial two-hybrid (B2H) strategy[19]. The T7 RNAP can be split into two domains that spontaneously reassemble to reform the functional RNAP enzyme[47]. By fusing the two proteins of interest whose binding we want to evolve to the domains of the split RNAP and cloning each fusion into either the AP or SP[24], we can imagine that

proteins that evolve to recognize and bind one another better will reform the active T7 RNAP resulting in more pIII production thereby leading to more phage propagation (Fig. 11). A recent improvement to the system includes fusions to either end of the RNAP domains, thereby improving the ease and adaptability of the system[24], including developing the systems as a versatile biosensor platform.

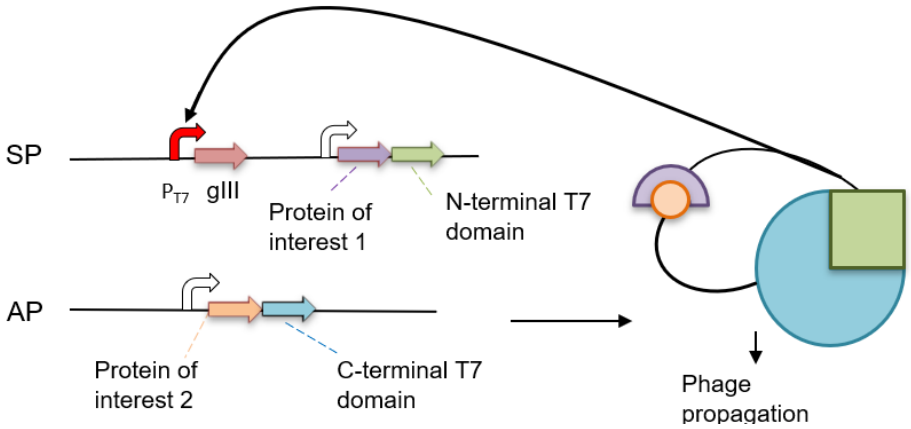


Figure 11. Outline of split T7 RNAP PACE. Adapted from Pu *et al*[24]. By fusing one of the proteins of interest to the N-terminal domain of the T7 RNAP and cloning the fusion into the SP, and then fusing the other protein of interest to the C-terminal domain of the T7 RNAP and cloning that fusion into the AP, you can set up a PACE selection to develop mutations that improve the protein-protein interaction by tying gIII expression to the protein-protein interaction. The stronger the protein-protein interaction, the easier it will be to reconstitute the intact T7 RNAP that can interact with P_{T7} (red) to mediate expression of gIII (brown) needed for phage propagation.

Additionally, a B2H-like system can be used to evolve protein-protein interactions as described to evolve the BT toxin Cry1Ac to bind to a different receptor to overcome BT toxin resistance[19]. Cry1Ac was fused to the omega subunit of RNAP and cloned into the SP, while fragments of the receptor that Cry1Ac was intended to bind were fused to the 434 phage cI repressor DNA-binding domain (Fig. S14).

Additional protein activities compatible with PACE

Above we have outlined techniques for linking gIII expression to DNA binding activity and to protein-protein interactions; however, there exist other protein activities that can be used to generate a selection circuit in PACE. These techniques include, but are not limited to, using enzymatic activities such as recombinase activity to flip the gIII promoter into the proper orientation that allows it to be regulated from its upstream promoter, and proteolysis by utilizing a proteolysis-activated RNAP to drive gIII expression[10, 48]. PACE can even be used to alter the substrate specificity of proteases to create custom proteases that can be used for biotechnological/therapeutic purposes[49]. Additionally, other DNA-binding proteins such, as Cas9 and TALENs, have been fused to the omega subunit of RNAP to generate the AD that drives gIII expression[22, 50]. PACE can also be used to evolve drug resistance in proteins and be used as a means to identify the vulnerabilities of drug candidates to the evolution of clinically relevant drug resistance[48].

The stepping-stone strategy for evolving new functions in PACE: Going from A→ B→ C

One of the initial applications of PACE was to evolve T7 RNAP to recognize the T3 promoter instead of the T7 promoter[10, 11]. The initial strategy for changing the promoter that T7 RNAP recognizes was to initially evolve the RNAP against a hybrid "stepping-stone" promoter that contained elements of both the T3 and T7 promoters that controlled expression of gIII from the AP (Fig. 12). This was done such that the wild-type RNAP would have some limited activity against the hybrid promoter, ensuring some minimal amount of gIII expression (leading to phage propagation), while the RNAP accumulated mutations to improve binding to the T3 elements of the hybrid promoter. After allowing the T7 RNAP to accumulate mutations to better recognize the

stepping-stone promoter, the partially evolved RNAP was challenged against the T3 promoter to evolve the RNAP to recognize the full T3 promoter.

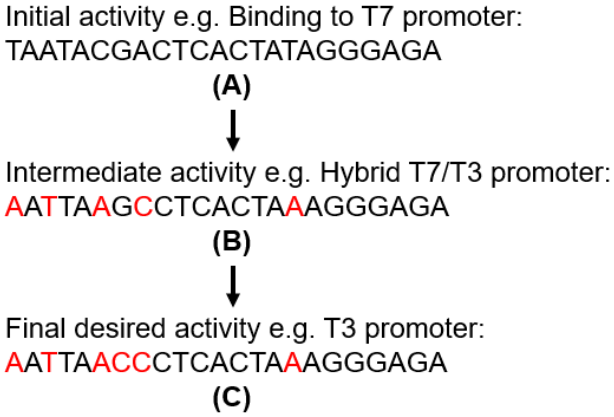


Figure 12. Stepping-stone strategy. Adapted from Leconte *et al*[11]. Promoter sequences used to change the sequence from which T7 RNAP enables transcription, with sites that were changed from the initial T7 promoter shown in red. T7 RNAP was evolved to recognize the T3 promoter within 96 hours by using the stepping-stone strategy to accumulate beneficial mutations in the RNAP.

If the goal is to evolve a new or modified feature in the protein of interest, this gradual stepping-stone strategy, where we initially evolve the protein of interest against conditions that are semi-permissible like the T7/T3 hybrid promoter, can be used before fully challenging the protein against the endpoint conditions of the evolution. This stepping-stone technique has also been used to evolve a new Bt toxin to overcome insect resistance[19], proteases with altered substrate specificities [48, 49], as well as variants of the T7 RNAP that support fusions made to its C-terminus[46], highlighting the potential effectiveness of this technique. The biggest constraint to using a stepping-stone approach in PACE is the design of the stepping-stone AP, such that it allows some level of phage propagation while still presenting enough of an evolutionary challenge for the protein to select for beneficial mutations. This necessitates that the protein of interest be well characterized in order to be able to design an appropriate stepping-stone(s) for the selection process.

Using negative selection in PACE to select against the protein's intrinsic properties

While the initial PACE experiments showed successful evolution of T7 RNAP toward recognition of the T3 promoter to effect expression of gIII, the evolved T7 RNAP still retained activity on the original T7 promoter[35], meaning that substrate recognition merely broadened and was not truly altered as had happened in other directed evolution experiments[51]. This can be one of the shortcomings of the stepping-stone strategies for evolution, in that the evolved protein will most likely retain some of its native function, which is typically undesirable.

To address this, a form of negative selection was developed where the protein of interest was forced to evolve to no longer carry out its original function while developing the new desired properties. To this effect, AP_{neg} was designed[35], where a dominant negative mutant of gIII ($gIII^-$) was cloned into the AP instead of the wild-type gIII and placed under the control of a theophylline-activated riboswitch (Fig. S13). The $gIII^-$ is N-C83 where the N1/N2 domains of the pIII are intact, but the C domain has an internal deletion of 70 residues[14, 17]. The truncated C domain can mediate attachment to a phage particle but cannot catalyze detachment of phage from the host cell membrane; the M13 phage is then trapped when it infects the host, resulting in it being washed out with the host from the Lagoon[35, 46]. By modifying AP_{neg} to produce $pIII^-$ when the protein of interest carries out its initial function, we can establish a powerful negative selective pressure: the phage host contains either the original AP or AP_{neg}; therefore, the ongoing mutagenesis within the PACE system can evolve the protein of interest to no longer carry out its intrinsic function while acquiring the newly evolving function[21, 35, 46].

Allowing for genetic drift in PACE by utilizing the drift plasmid

Alternating between periods of low and high stringency within a given PACE selection was initially done by physically changing the host that was introduced into the Lagoon in a set-up that necessitated having two chemostats, with each containing either a high stringency host or a low stringency host to allow for drift. This issue was remedied by the creation of the "drift" plasmid[35], which allows for the inducible expression of the gIII protein that enables the researcher to decide when and by what degree to reduce selective pressure in the PACE system. The drift plasmid has the gIII protein under control of a $P_{psp\text{-}tet}$[52] so that infection by M13 phage (which would be continuous during PACE) and the presence of tetracycline/anhydrotetracycline (which can be provided) are needed for expression of the gIII protein.

By transforming S1030 cells that have the desired AP with the drift plasmid and providing tetracycline to the PACE system, we can remove selective pressure in the system and allow for a period of neutral drift to occur, where mutagenesis is ongoing in the absence of selection. The net effect is similar to that of the dual chemostat set-up described above; we can alternate between periods of no selective pressure where a library of random mutants of the protein of interest can accumulate in the Lagoon, then switch to a period of high stringency to select for the most fit members of the library. After a period of high stringency, we can induce neutral drift again to replenish the phage titer in the Lagoon and expand the library before repeating the selection process; this can be done for as long as it takes to acquire the desired activity. Care needs to be taken when inducing neutral drift to avoid accumulating cheater phage that will impede the PACE experiment.

How convergent evolution & different evolutionary trajectories may impact PACE

Several studies have examined evolutionary outcomes of parallel populations from a common ancestor; while phenotypic outcomes might be the same across different populations, the genetic changes that give rise to those phenotypes can vary[5, 6]. Very early on, it was shown that this is also a feature of PACE, where the evolutionary trajectories that are accessible to the protein of interest are heavily influenced by the conditions under which the PACE experiment takes place[11, 12, 23]. By simply altering the mutagenesis rate/stringency of the T7 RNAP selection, the Liu group was able to alter the nature of the mutations that arose in the T7 RNAP, as well as the abundance of the mutations within the phage population[11].

Other PACE studies have shown that varying substrate availability and concentration, for example, can affect the nature of mutations that arise in PACE. Dickinson *et al.* evolved the T7 RNAP via two distinct pathways before evolving the two separate populations toward a common final target. They found that "stochastic occurrences and prior selection history" greatly impacted the convergent evolution of the two distinct populations[12]. Epistasis has also been shown to significantly impact the suite of mutations that are tolerated within a protein[6, 12]. For example, in a study of cephalosporin antibiotic resistance mutations in β-lactamase, several active site mutations were tolerated only in the presence of a pre-existing stabilizing mutation[53], and this was also seen in PACE as well, where a destabilizing mutation in an evolved Bt toxin had to be lost (via reversion mutations) in order to obtain a stable protein[19]. No two PACE experiments will be the same; depending on the selection scheme used and conditions within the Lagoon as well as chance, it is reasonable to expect to see different mutations arise in the protein of interest. This makes techniques, like PANCE, that can parallelize these selections very attractive to maximize the portion of the evolutionary landscape that can be sampled.

Acknowledgements
We thank Dr. David Liu and his lab members, especially Shannon Miller and Aditya Raguram, for their assistance in helping us develop our PACE systems. JAS is grateful for funding from CHRP, NSERC, and the University's VP Research.

References

1. Tsien RY. The Green Fluorescent Protein. Ann Rev Biochem. 1998;67:509-44.
2. Remington SJ. Green fluorescent protein: A perspective. Prot Sci. 2011;20:1509-19.
3. Marshall SA, Lazar GA, Chirino AJ, Desjarlais JR. Rational design and engineering of therapeutic proteins. Drug Discov Today. 2003;8:212-21.
4. Selzer T, Albeck S, Schreiber G. Rational design of faster associating and tighter binding protein complexes. Nat Struct Mol Biol. 2000;7:537-41.
5. Packer MS, Liu DR. Methods for the directed evolution of proteins. Nat Rev Genetics. 2015;16:379-94.
6. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol. 2009;10:866-76.
7. Arnold FH, & Georgiou, G. Directed Enzyme Evolution. Arnold FH, Georgiou G, editors: Humana Press; 2003.
8. Serebriiskii IG, Golemis EA. Two-hybrid system and false positives. Approaches to detection and elimination. Meth Mol Biol. 2001;177:123-34.
9. Vidalain P-O, Boxem M, Ge H, Li M, Vidal M. Increasing specificity in high-throughput yeast two-hybrid experiments. Methods. 2004;32:363-70.
10. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. Nature. 2011;472:499-505.
11. Leconte AM, Dickinson BC, Yang DD, Chen IA, Allen B, Liu DR. A Population-Based Experimental Model for Protein Evolution: Effects of Mutation Rate and Selection Stringency on Evolutionary Outcomes. Biochemistry. 2013;52:1490-99.
12. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc Natl Acad Sci USA. 2013;110:9007-12.
13. Badran AH, Liu DR. In Vivo Continuous Directed Evolution. Curr Opin Chem Biol. 2014;24:1-10.
14. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr Iss Mol Biol. 2011;13:51-76.
15. Bennett NJ, Rakonjac J. Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. J Mol Biol. 2006;356:266-73.
16. Smeal SW, Schmitt MA, Rodrigues Pereira R, Prasa A, Fisk JD. Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. Virology. 2017;500:259-74.
17. Rakonjac J, Feng JN, Model P. Filamentous phage are released from the bacterial membrane by a two step mechanism involving a short C-terminal fragment of pIII. J Mol Biol. 1999;289:1253-65.
18. Rakonjac J, Model P. Roles of pIII in filamentous phage assembly. J Mol Biol. 1998;282:25-41.
19. Badran AH, Guzov VM, Huai Q, Kemp MM, Ishwanath P, Kain W, et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature. 2016;533:58-63.
20. Badran AH, Liu DR. Development of Potent In Vivo Mutagenesis Plasmids with Broad Mutational Spectra. Nat Comm. 2015;6:8425-34.
21. Bryson DI, Fan C, Guo LT, Miller C, Söll D, Liu DR. Continuous directed evolution of aminoacyl-tRNA synthetases. Nat Chem Biol. 2017;13:1253-60.

22. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature. 2018;556:57-63.
23. Roth TB, Woolston BM, Stephanopoulos G, Liu DR. Phage-assisted evolution of Bacillus methanolicus methanol dehydrogenase 2. ACS Synth Biol. 2019;8:796-806.
24. Pu J, Zinkus-Boltz J, Dickinson BC. Evolution of a split RNA polymerase as a versatile biosensor platform. Nat Chem Biol. 2017;13:432-38.
25. Xu J, Chen G, De Jong AT, Shahravan SH, Shin JA. Max-E47, a designed minimalist protein that targets the E-box DNA site in vivo and in vitro. J Am Chem Soc. 2009;131:7839-48.
26. Meng X, Brodsky MH, Wolfe SA. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nat Biotech. 2005;23:988-94.
27. Meng X, Wolfe SA. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. Nat Protocol. 2006;1:30-45.
28. Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH, Wolfe SA. A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nuc Acids Res. 2008;36:2547-60.
29. Ouwerkerk PBF, Meijer AH. Yeast One-Hybrid Screening for DNA-Protein Interactions. Current Protocols in Molecular Biology: John Wiley; 2001. p. 12..1-..22.
30. Dove SL, Hochschild A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. Genes Dev. 1998;12:745-54.
31. Struhl K, Cameron JR, Davis RW. Functional genetic expression of eukaryotic DNA in Escherichia coli. Proc Natl Acad Sci USA. 1976;73:1471-75.
32. Brennan MB, Struhl K. Mechanisms of increasing expression of a yeast gene in Escherichia coli. J Mol Biol. 1980;136:333-38.
33. Swanson HI, Yang J-H. Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site. Nucl Acid Res. 1999;27:3205-12.
34. Thuronyi BW, Koblan LW, Levy JM, Yeh W-H, Zheng C, Newby GA, et al. Continuous evolution of base editors with expanded target compatibility and improved activity. Nat Biotechnol. 2019;37.
35. Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR. Negative Selection And Stringency Modulation Enable Phage-Assisted Continuous Evolution (PACE) of Enzymes With Altered Specificity. Nat Chem Biol. 2014;10:216-22.
36. Chen H, Bjerknes M, Kumar R, Jay E. Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. Nucl Acid Res. 1994;22:4953-57.
37. Cheetham GM, Jeruzalmi D, Steitz TA. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. Nature. 1999;399:80-3.
38. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013;65:1357-69.
39. Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, et al. Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci USA. 2001;98:892-97.
40. Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, et al. Translation initiation in Escherichia coli: sequences withing the ribosome-binding site. Mol Microbiol. 1992;6:1219-29.
41. Salis HM. The ribosome binding site calculator. Methods Enzymol. 2011;498:19-42.

42. Stanssens P, Remaut E, Fiers W. Alterations upstream from the Shine-Dalgarno region and their effect on bacterial gene expression. Gene. 1985;36:211-33.
43. Espah Borujeni A, Salis HM. Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. J Am Chem Soc. 2016;138:7016-23.
44. Novoa EM, de Pouplana LR. Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet. 2012;28:574-81.
45. Hannig G, Makrides SC. Strategies for optimizing heterologous protein expression in Escherichia coli. Trends Biotechnol. 1998;16:54-60.
46. Pu J, Disare M, Dickinson BC. Evolution of C-terminal modification tolerance in full-length and split T7 RNA Polymerase biosensors. ChemBioChem. 2019;20:1547-53.
47. Shis DL, Bennett MR. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. Proc Natl Acad Sci USA. 2013;110:5028-33.
48. Dickinson BC, Packer MS, Badran AH, Liu DR. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat Comm. 2014;5:5352. DOI: 10.1038/ncomms6352.
49. Packer MS, Rees HA, Liu DR. Phage-assisted continuous evolution of proteases with altered substrate specificity. Nat Comm. 2017;8:956. DOI: 10.1038/s41467-017-01055-9.
50. Hubbard BP, Badran AH, Zuris JA, Guilinger JP, Davis KM, Chen L, et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat Methods. 2015;12:939-42.
51. Tracewell CA, Arnold FH. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol. 2009;13:3-9.
52. Brissette JL, Weiner L, Ripmaster TL, Model P. Characterization and sequence of the Escherichia coli stress-induced psp operon. J Mol Biol. 1991;220:35-48.
53. Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol. 2002;320:85-95.

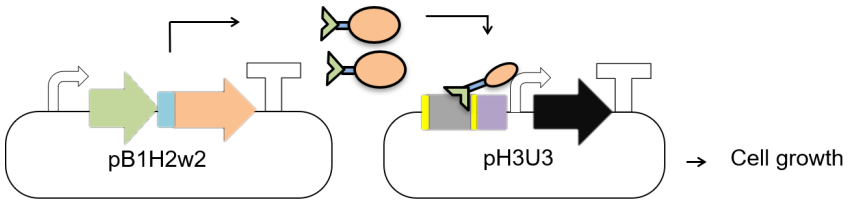


Figure S1. Initial B1H set-up for ME47. ME47 (peach) is fused to the omega subunit of RNA polymerase (light green) via a 12 amino-acid linker (light blue), this fusion protein is under the control of the *lac* promoter on pB1H2w2. When expressed after the addition of IPTG, the fusion protein will interact with the AT E-box motif (light grey for the core motif, yellow for the flanking sequence) that has a 13-nucleotide spacer (purple) between it and the 5' edge of the -35 weak *lac* promotor that regulates expression of the *hisB* reporter gene (black) that will allow US0 cells to grow in histidine deficient media.

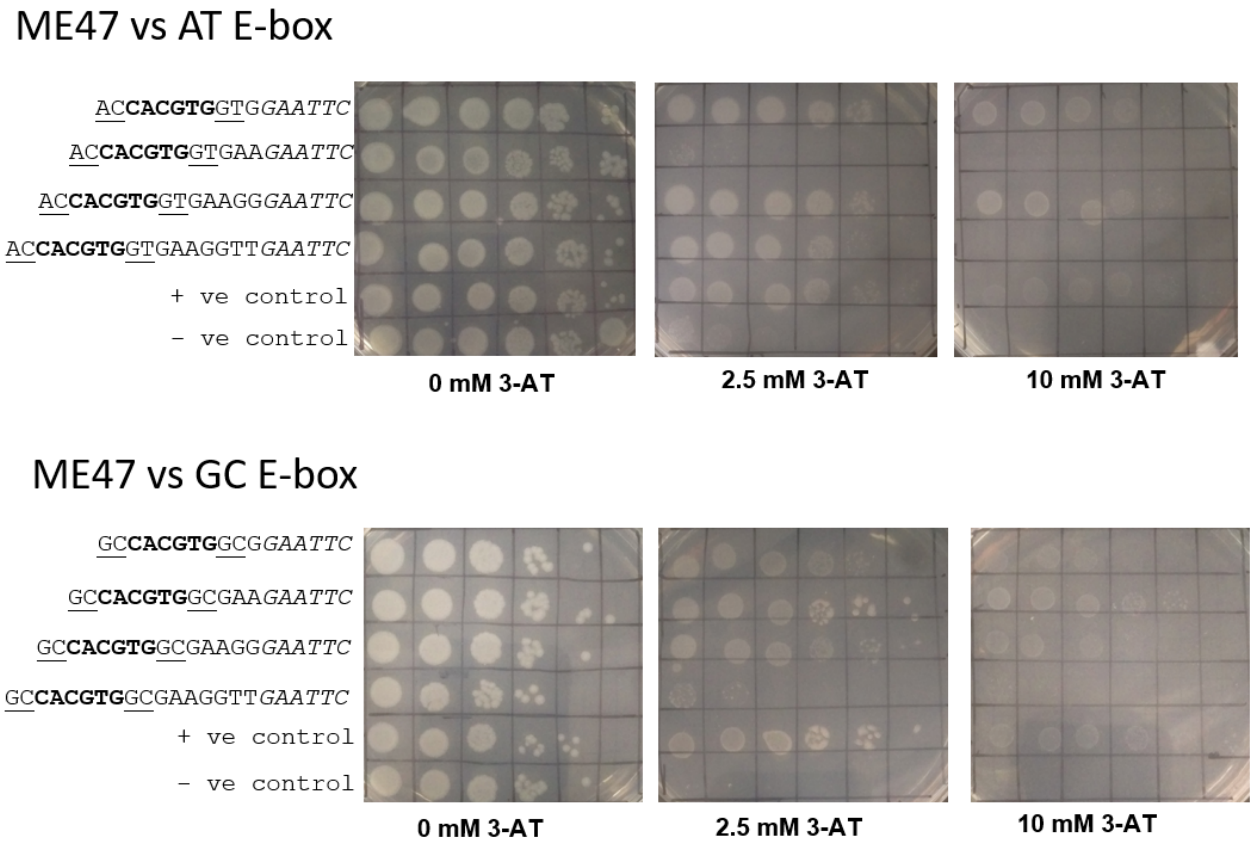


Figure S2. Representative B1H assays for the GC and AT libraries. The library spanned 7 to 25 nucleotide spacers, increasing in increments of 2 nucleotides (shown here are 7,9,11,13 nucleotide spacers for both AT and GC E-boxes). The ideal E-box reporter would give a signal similar to the positive control even at the higher concentrations of 3-AT, unlike the -13 AT E-box construct. Serial dilutions were carried out from left to right going from 10^{-1} to 10^{-6}. The positive control consisted of US0 cells transformed with Zif268 and its cognate DNA in lieu of the E-box), while the negative control consisted of US0 cells transformed with Zif268 and empty pH3U3[1]. Omega-ME47 produced strong signals from the -7 and -11 AT E-box constructs and -7, -9, and -11 GC E-box constructs, making these promising candidates for the PACE-B1H.

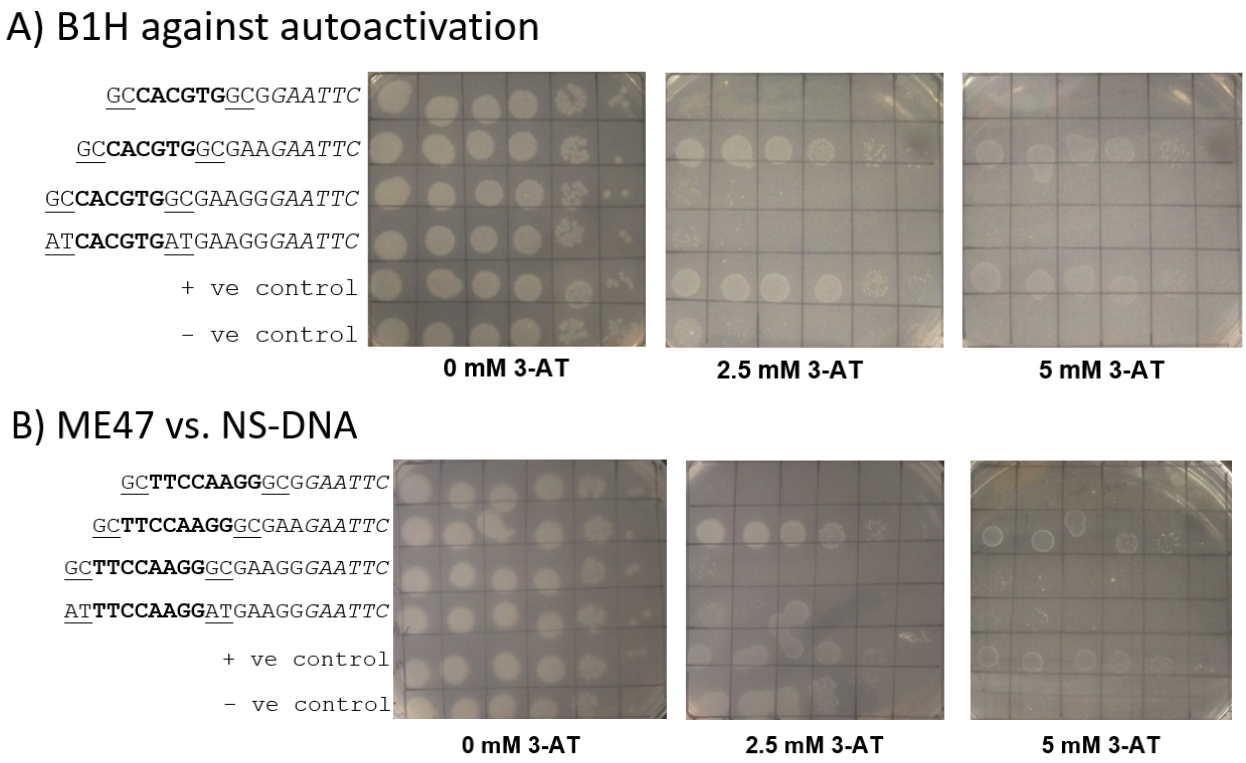


Fig. S3. B1H assays against autoactivation. A) The autoactivation B1H consisted of US0 cells transformed with the various library members on the pH3U3 plasmid without omega-ME47 subunit fusion on the pB1H2w2 plasmid (shown here are library members with -7 GC, -9 GC, -11 GC, and -11 AT nucleotide spacers for both conditions). In the absence of the omega-ME47 subunit fusion protein, no signal should be seen, as such, the -9 GC E-box would be considered an autoactivating E-box while the -7 GC E-box is not autoactivating. **B) B1H assays against nonspecific activity.** The nonspecific B1H consisted of US0 cells transformed with the omega-ME47 subunit fusion on the pB1H2w2 plasmid and the same spacer library members used in S3A on the pH3U3 vector, whose E-box was replaced with nonspecific DNA (5' TTCCAAGG in lieu of 5'CACGTG). The controls were the same as outlined earlier. The -7 GC, -11 AT and -11 GC constructs did not produce a signal in the nonspecific B1H, making them promising candidates for use in the PACE selection circuit.

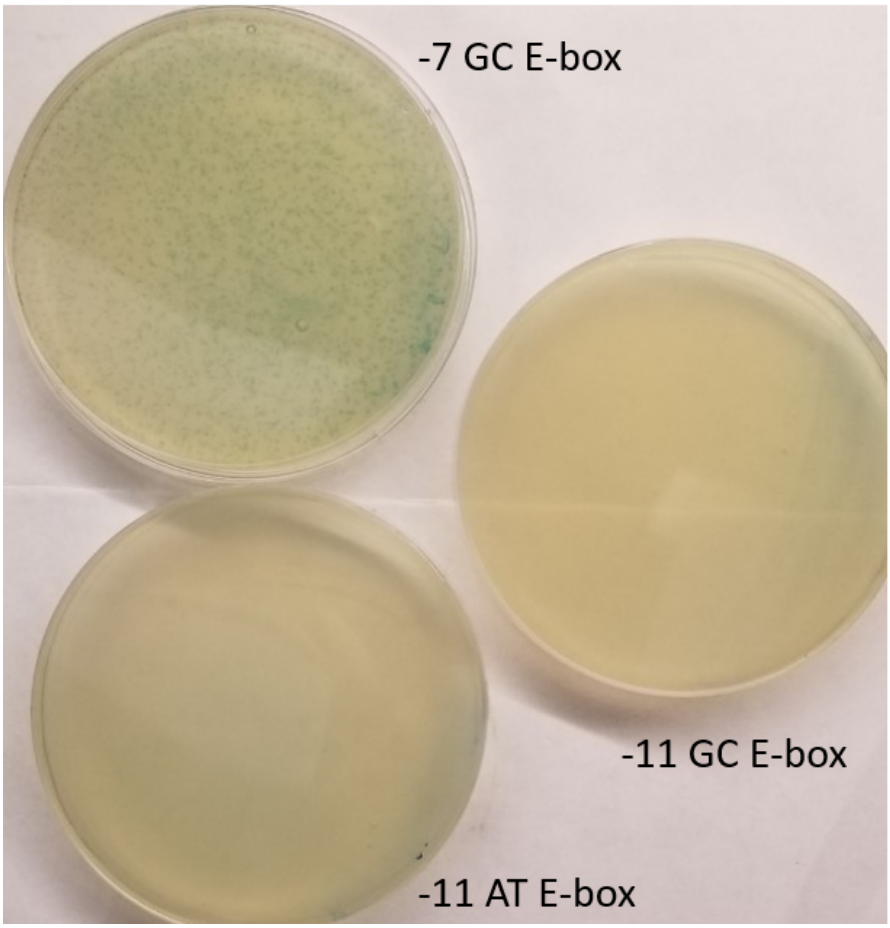


Figure S4. Plaque assay results for the AP library. 2060 cells harboring the various AP constructs were infected with M13 phage containing SP-ME47 to see which AP permits phage propagation. The 2xYT top agar used for the plaque assay was supplemented with Bluo-Gal to allow for the formation of blue plaques to facilitate visualization of the plaques. Dark blue points are plaques were the 2060 cells with a valid AP construct were successfully infected by the SP-ME47-containing phage, indicating that the AP construct is viable for PACE. Shown here are plaque assays done against SP-ME47 that had been diluted 100-fold to allow for visualization of individual plaques. Only the -7 GC E-box allowed for plaque formation, marking it as the AP to be used in the PACE selection circuit.

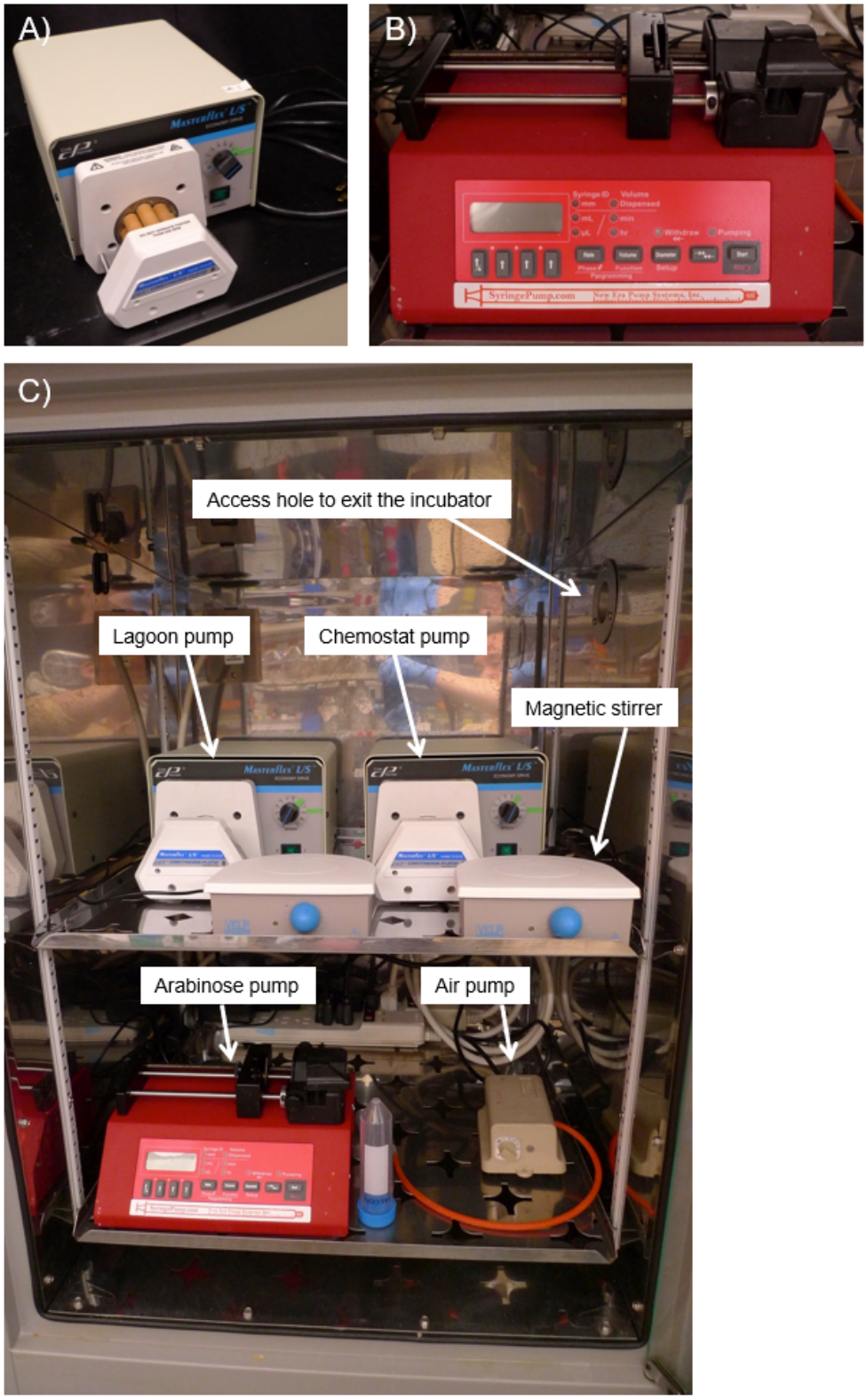
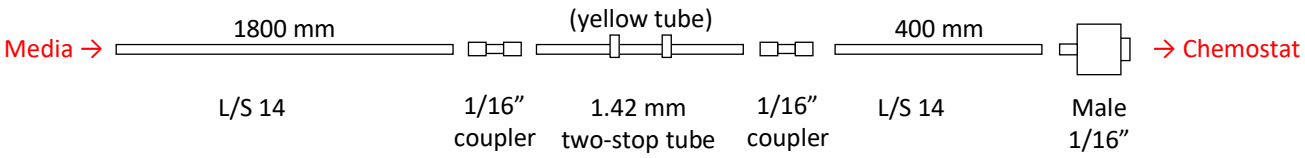
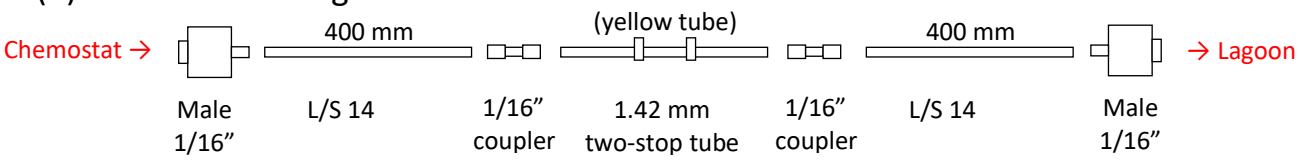


Figure S5. PACE pumps and incubator. **A**) The peristaltic pump used for the Chemostat/Lagoon. **B**) Syringe pump for the arabinose solution. **C**) The basic components of PACE installed in the tabletop incubator. A 50 mL Falcon tube was included for size comparison.

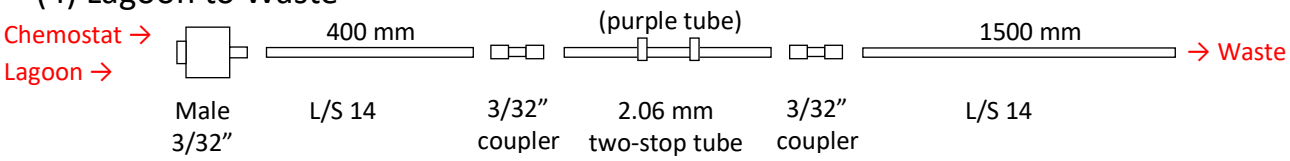
(1) Media to Chemostat



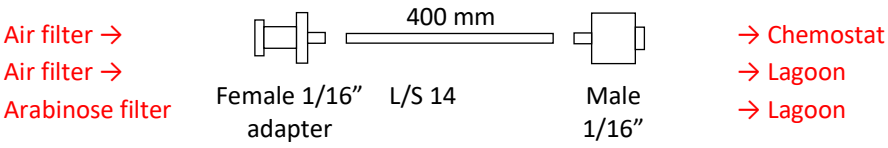
(2) Chemostat to Lagoon



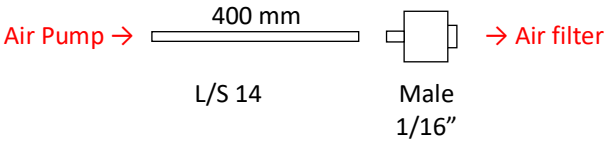
(3) Chemostat to Waste
(4) Lagoon to Waste



(5) Air Filter to Chemostat
(6) Air to Lagoon
(7) Arabinose Filter to Lagoon



(8) Air Pump to Filter



(9) Arabinose Syringe to Filter

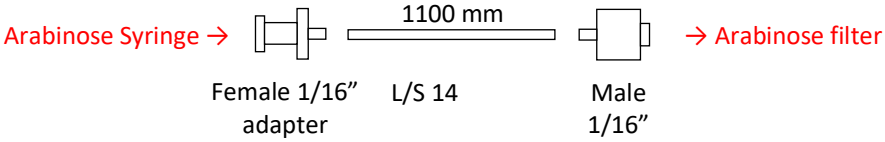


Figure S6. PACE tubing connections (above, previous page). Diagrams are not to scale. The L/S 14 tube needs to be cut for the desired length. Do *not* cut the two-stop tubes. The destinations in/out of the tubes are indicated in red.

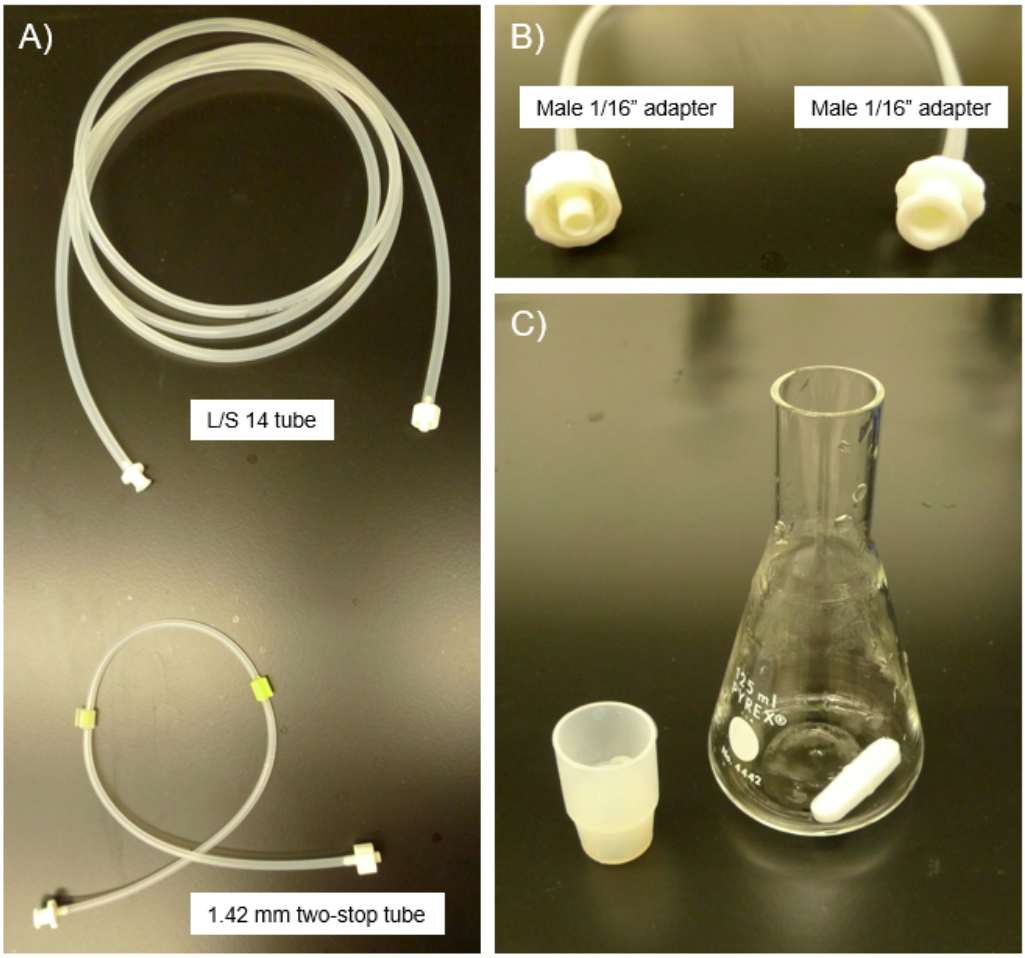


Figure S7. Closeup of PACE tubes and flasks. **A**) L/S 14 Tube and 1.42 mm two-stop tube, both connected to a pair of Male/Female 1/16 " adapters. **B**) Close-up of the Male/Female 1/16" adapters. **C**) Flask used as the Chemostat/Lagoon, shown with the silicone cap and a magnetic stirrer.

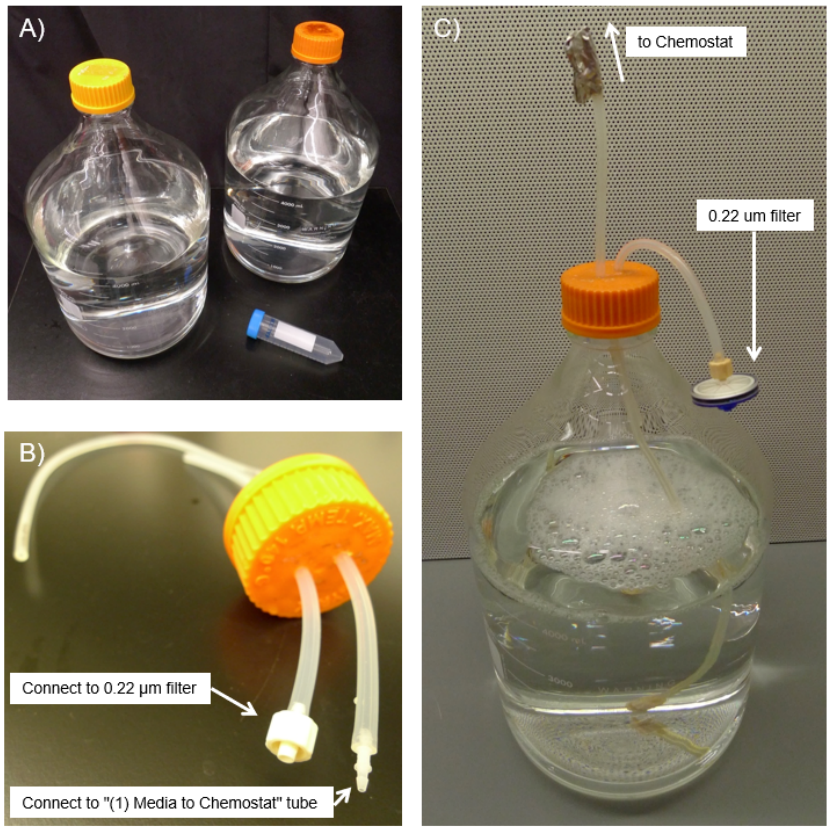


Figure S8. Closeup of PACE media bottle. **A)** A pair of 5 L media bottle used in our PACE setup. A 50 mL Falcon tube was included for size comparison. **B)** Close up of the Media bottle cap. The media bottle cap was made in-house by drilling a hole in a Pyrex bottle cap and then fitting two L/S-14 tubes of appropriate length. The number correspond to Figure S6. **C)** PACE media bottle after assembly.

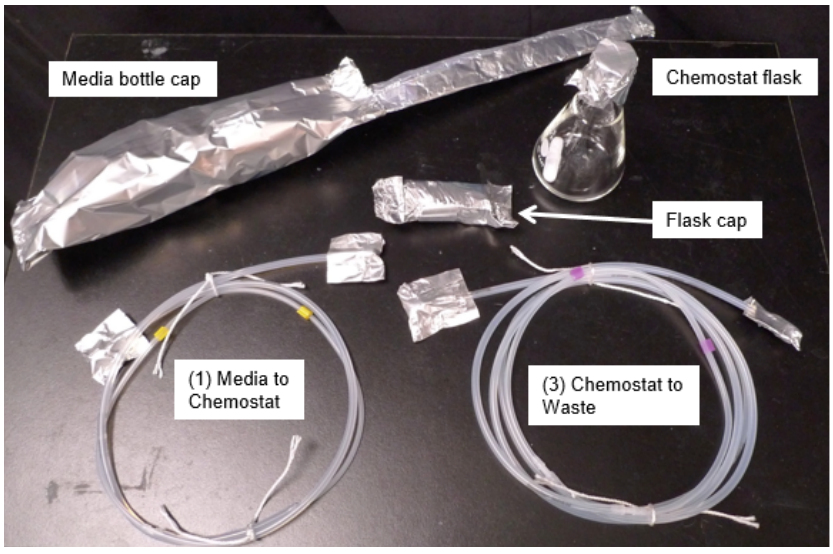


Figure S9. Components for PACE are packaged in aluminum foil before autoclaving. The numbers correspond to Figure S6.

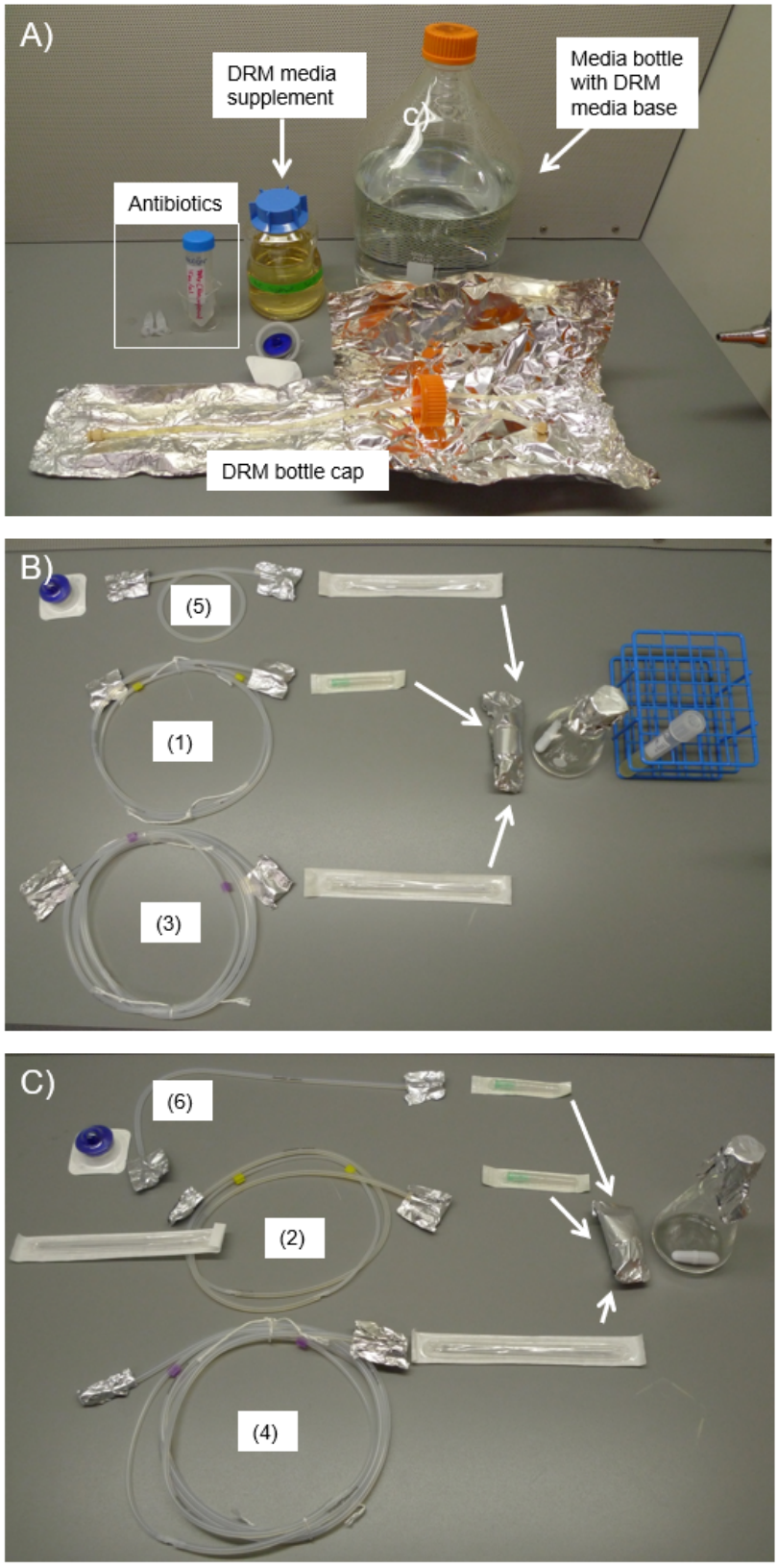


Figure S10. All components required to assemble the DRM bottle **A)**, the Chemostat **B)**, and the Lagoon **C**). The numbers correspond to Figure S6.

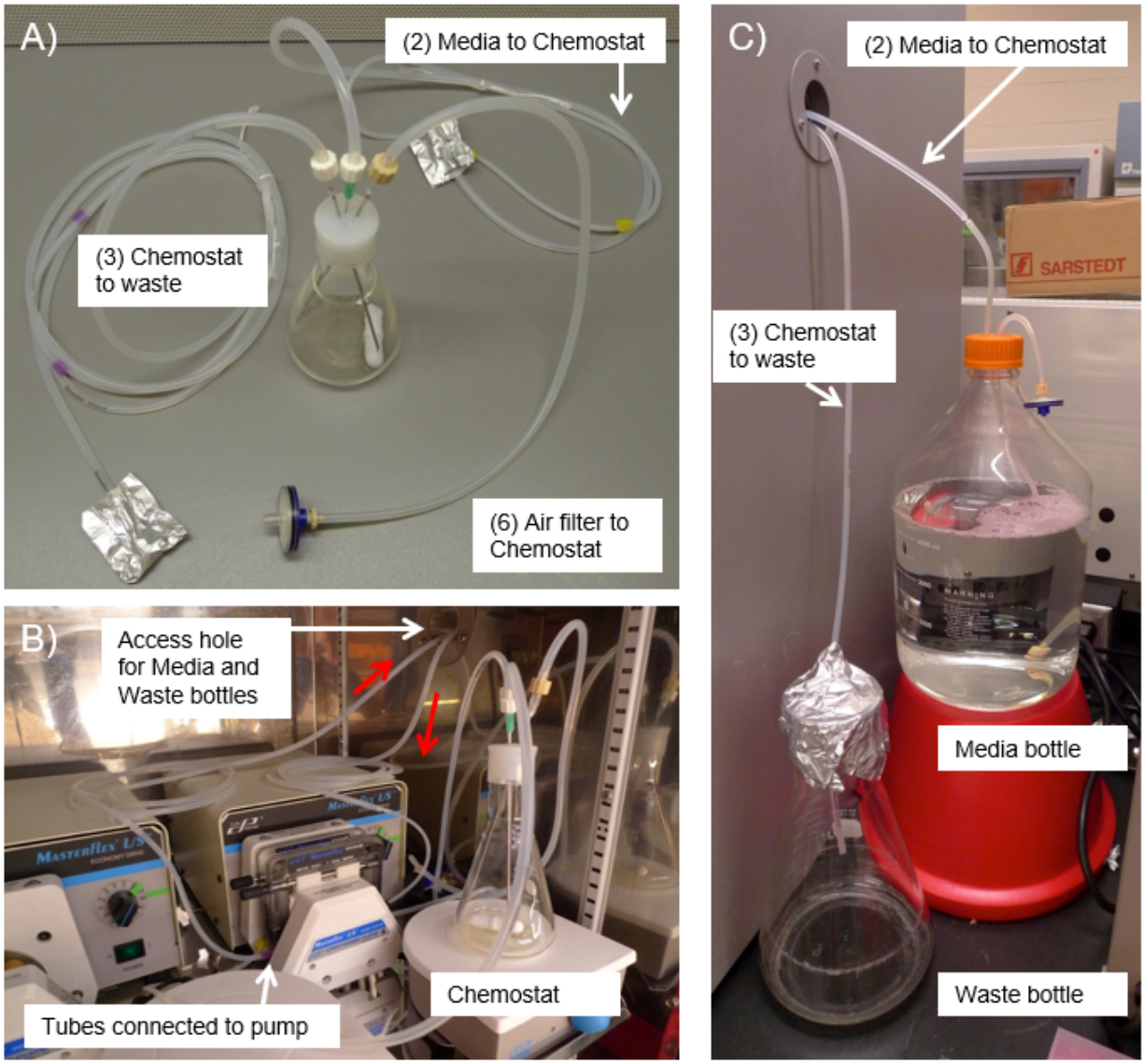


Figure S11. Setting up the Chemostat. **A**) Chemostat flask after assembly. The numbers correspond to Figure S6. **B**) Chemostat flask installed in the incubator. The "(2) Media to Chemostat" and "(3) Chemostat to Waste" tubes are connected to the right pump using the pump clamps. These tubes exit the incubator via the access hole to reach the Media bottle and Waste bottle. **C**) Outside the incubator where "(2) Media to Chemostat" and "(3) Chemostat to Waste" tubes are connected to the Media bottle and the Waste bottle, respectively.

Figure S12. Representative view of the fully assembled PACE system. This is what the PACE system looks like when fully set up.

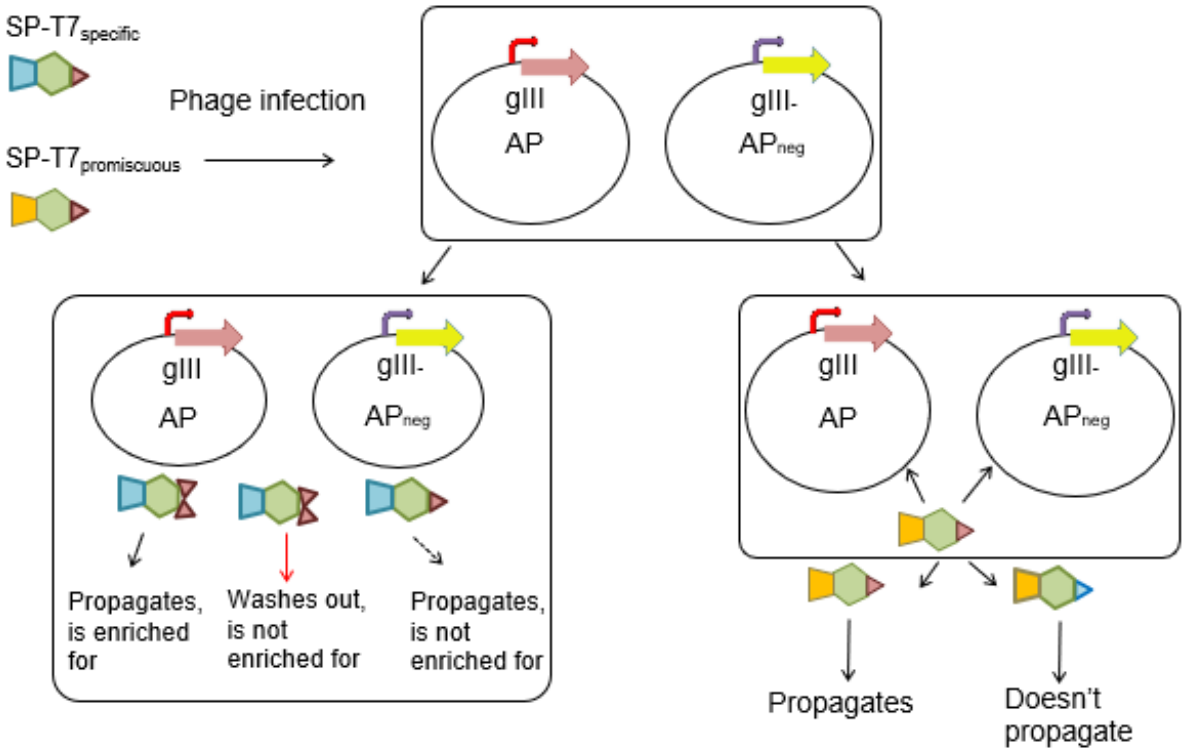


Figure S13. Negative selection. Adapted from Carlson *et al*[2]. S1030 host cells were transformed with one of two APs; AP_{neg} has $pIII^-$ production (blue) under the control of the T3 promoter and the theophylline riboswitch (purple); the standard gIII (brown) is under the control of the T7 promoter (red). The Lagoon was seeded with a $1{:}10^6$ ratio of specific T7 RNAP: promiscuous T7 RNAP binds both T7/T3 promoters. Upon the addition of theophylline, enrichment of the specific T7 RNAP occurred, validating this as a valid negative selection strategy.

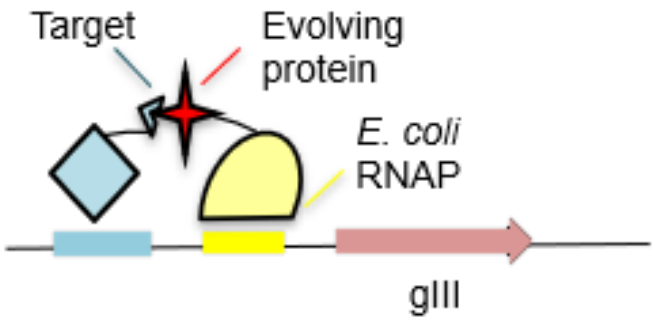


Figure S14. An alternative selection circuit for evolving protein-protein interactions. Adapted from Badran *et al*[3]. 434 phage *cI* repressor binds its cognate DNA and presents the receptor fragments to be bound by the omega-Cry1Ac subunit fusion. If the Cry1Ac fusion recognizes and binds the receptor fragments, the RNAP will be positioned in such a way as to be able to affect the transcription of the downstream gIII to allow for phage propagation. This can be adapted so that a DNA binding protein is fused to the target protein (blue) that binds its cognate DNA waiting to interact with the evolving protein fused to the *E.coli* RNAP (yellow) to drive gIII expression allowing for phage propagation.

Figure S15. List of materials used in PACE. Below is a "shopping list" for PACE set-up.

Media pumps and related accessories

Main pump
Masterflex L/S economy variable-speed drive, 7 to 200 rpm, 115 VAC.
HV-07554-80

Arabinose Pump
New Era NE-1000 Programmable Single Syringe Pump

Pump Head
Masterflex L/S 4-channel, 8-roller cartridge pump head.
SC-07519-20

Pump Head Cartridges
Masterflex L/S small cartridges for 07519-20 and -25 pump heads.
SC-07519-85

MasterFlex tubes

Microbore two-stop tube sets, silicone (platinum-cured); 0.89 mm ID. Pk of 6.
EW-06421-26

Microbore two-stop tube sets, silicone (platinum-cured); 1.42 mm ID. Pk of 6.
EW-06421-34

Microbore two-stop tube sets, silicone (platinum-cured); 2.06 mm ID. Pk of 6.
EW-06421-42

L/S 14 Silicone Tube (platinum-cured)
T-96410-14

Male Luer x 1/16" barb adapter
RK-45518-00

Male Luer x 3/32" barb adapter
RK-45518-02

Female Luer x 1/16" barb adapter
RK-45508-00

3/32" barb x 3/32" barb
RK-45508-02

Needles

BD Blunt Fill, Luer, 18-gauge, 100/pk 1.5 inch
BD305180305180 VWR (media into chemostat, cells into Lagoon).

Air-Tite Vet Premium Hypodermic Needle 16-gauge, 4"
N164 (chemostat to waste, Lagoon to waste)

Air-Tite Vet Premium Hypodermic Needle 22-gauge, 1"
N221 (inducer needle)

Chemostat, Lagoon and other vessels

125 mL Erlenmeyer flasks
RK-34502-58

15 mL Drosophila flasks (for smaller volumes)
B7990-5

5 L media bottles
PYREX brand 1395 media bottles, 5L
T-34514-28

Miscellaneous

Magnetic stirrer
Low-Profile magnetic stirrer
T-86579-10

Magnetic stirring rod of appropriate length (3 cm is enough)
58948-950 VWR, 28 mm long, 7.9 mm thick.

Air pump
Use any commercially available aquarium pump

Autoclavable silicone caps
Folding Skirt Stoppers, silicone
T-62995-96

References

1. Christy B, Nathans D. DNA binding site of the growth factor-inducible protein Zif268. Proc Natl Acad Sci USA. 1989;86:8737-41.
2. Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR. Negative Selection And Stringency Modulation Enable Phage-Assisted Continuous Evolution (PACE) of Enzymes With Altered Specificity. Nat Chem Biol. 2014;10:216-22.
3. Badran AH, Guzov VM, Huai Q, Kemp MM, Ishwanath P, Kain W, et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature. 2016;533:58-63.