Phage-Assisted Continuous Evolution (PACE): A Guide Focused on Evolving Protein–DNA Interactions

Serban C. Popa, Ichiro Inamoto, Benjamin W. Thuronyi, and Jumi A. Shin*

ABSTRACT: The uptake of directed evolution methods is increasing, as these powerful systems can be utilized to develop new biomolecules with altered/novel activities, for example, proteins with new catalytic functions or substrate specificities and nucleic acids that recognize an intended target. Especially useful are systems that incorporate continuous evolution, where the protein under selective pressure undergoes continuous mutagenesis with little-to-no input from the researcher once the system is started. However, continuous evolution methods can be challenging to implement and a daunting investment of 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) toward the evolution of proteins that bind to a specific DNA target. We discuss factors to consider before adopting PACE for a given evolution scheme with focus on the PACE bacterial one-hybrid selection system and what optimization of a PACE selection circuit may look like using the evolution of the DNA-binding protein ME47 as a case study. We outline different types of selection circuits and techniques that may be added onto a basic PACE setup. With this information, researchers will be better equipped to determine whether PACE is a valid strategy to adopt for their research program 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 directed 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 clustered regularly interspaced short palindromic repeats (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 ref 1), 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.

Given such utility of the protein scaffold, it is logical for researchers to develop methods that enable us to exploit protein structure for our own research goals. PACE, which is phage-assisted continuous evolution,2 allows us to harness the utility of proteins by providing a means for evolution of proteins under selective pressure that is controlled by the researcher. Just as proteins possess a wide variety in structure and function, PACE and other continuous evolution systems can be modified in various ways toward solving diverse research problems. Here, we focus on the setup and validation of a basic PACE system for the evolution of a protein–DNA interaction in the PACE bacterial one-hybrid (B1H)3 selection circuit.

Protein Engineering and 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. 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: for example, 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. Directed evolution has been extensively applied to enzymes to create “green” biocatalysts with improved properties such as thermostability, robustness, stereoselectivity, novel chemistry and reactions, and improved or altered substrate specificities. Mutations are typically incorporated into the protein’s DNA-coding sequence (CDS) through methods such as the error-prone polymerase chain reaction (PCR), saturation mutagenesis, DNA shuffling, and more recently base-editing technologies like clustered regularly interspaced short palindromic repeats (CRISPR) to generate a DNA library for the protein of interest (recently reviewed in ref 9).

By selecting proteins with improved attributes while introducing random mutations into their sequences, directed evolution can circumvent the limitations of rational design. A solid understanding of the protein’s activity and mode of action is still 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. 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 performed where only library members with a level of 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. 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.

Directed evolution allows proteins to sample larger swaths of the evolutionary landscape when compared to rational design, such that proteins can gain new functions by “simple adaptive walks involving small numbers of mutations.” 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. However, the use of directed evolution brings new challenges to the table, one of which is the design of a selective assay. 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. 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 performed where only library members with a level of 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. 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.

Directed evolution allows proteins to sample larger swaths of the evolutionary landscape when compared to rational design, such that proteins can gain new functions by “simple adaptive walks involving small numbers of mutations.” 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. 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.

Seminal work done by Lenski showed that it is possible to reproduce adaptive evolution in organisms in a lab setting, admittedly on very long time scales and with significant human intervention. To remedy the amount of human intervention required in directed evolution, various continuous evolution systems have been designed in order to accelerate the rates of mutagenesis coupled to systems where human intervention is minimal. Historically, this entailed using mutator hosts like Escherichia coli XL1-Red that feature damaged DNA repair systems to evolve proteins, but these strains are liable to mutating the entire host genome leading to host instability. More modern strategies tie replication of the protein of interest to the activity of an orthogonal DNA polymerase or take advantage of viral–host interactions to create orthogonal circuits. PACE is one such system that accelerates the evolutionary process by employing highly error-prone DNA polymerases in tandem with the rapid replicative life cycle 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 theoretically used toward evolution of any genetically encoded molecule, including nucleic-acid aptamers.

The M13 Bacteriophage in Continuous Evolution. PACE is a true evolution system, in which evolving genes are subjected to continuous, seamless cycles of mutagenesis and selection. Liu and co-workers 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.

To accomplish directed evolution without constant human intervention, PACE utilizes the continuous infection of E. coli host cells by a modified version of the M13 bacteriophage. The mature M13 bacteriophage particle features a rod-shaped protein shell carrying a circular single-stranded phage DNA (Figure 1). The protein shell contains five different phage coat proteins. The majority of the coat is built from more than 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 the M13 phage.

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 (Figure 2A). 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 the production of mature M13 particles. Thirty minutes after the phage infection, dozens of double-stranded, transcriptionally active phage molecules are present within the host cell. Phage are made via a rolling-circle replication from the initial viral single-stranded DNA-coding sequence (CDS)}
Thus, the SP expresses the protein bacteriophage whose gIII gene is replaced by that for 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). Upon entry into the host, the single-stranded viral genome is converted to a double-stranded genome by host DNA polymerases; expression of the M13 genome can begin by utilizing the host machinery.

The PACE gIII selection system is the equivalent of the "selective assay" discussed above that establishes the direction of evolution, and the SP is the vector that expresses the protein of interest. The competition among 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 setup 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 °C. The first pump continuously delivers fresh media into the Chemostat, while transporting excess media and cells out of the Chemostat to the waste container (Figure 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 (Figure 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 the SP population—note this is an important factor when considering gene dosage effects for PACE (discussed below). 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 the 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.

Phage protein pIII, which is encoded by phage gene gIII, is essential for phage maturation and infectivity. The infectivity of M13 phage scales with increasing levels of pIII over a range of 2 orders of magnitude. 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, Figures 2 and 3). 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, gIII 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 (Figure 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.
to propagate in the Lagoon for hundreds of generations without interruption.

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. 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; for example, the flow rate of a 30 mL Lagoon may be 80 mL/h. At this rate, the culture in the Lagoon is refreshed in less than 30 min, allowing E. coli cells to linger in the Lagoon for less than half an hour. Since the fastest doubling time for E. coli is ~20 min, an average uninfected cell in the Lagoon has barely enough time to divide; moreover, host cells infected with M13 phage will experience a significant increase in their doubling times. This limits the ability of the host E. coli to produce progeny that persist in the Lagoon, which prevents mutations from accumulating in their genomes: such mutations 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 selection stringency. As a result of these combined 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 the 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. Minimalizing 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 any recombinant phage in the Lagoon.

During PACE, the mutagenesis rate during DNA replication is greatly increased using the Mutagenesis Plasmid (MP) that increases the rate of protein evolution. 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. Currently, the most effective variant of the MP is MP6 (available from Addgene), which increases the mutagenesis rate to more than 300 000-fold over the basal level of mutation in E. coli. This high mutagenesis rate allows for the rapid accumulation of mutations that persist in the protein after arabinose induction of MP. In theory, the PACE evolution experiments yield mutations from the very first generation, with enrichment of fitness-improving mutations developing within 1–3 d of continuous propagation. The PACE duration depends on the nature of the desired changes and the evolutionary trajectories the protein takes to achieve the desired function. PACE experiments usually run for at least several days, and multiplex evolution experiment trajectories have exceeded 20 d.

The Supporting Information provides a detailed discussion about protocols, experimental considerations, PACE setup, troubleshooting, variations of PACE, and reagents including cells, plasmids, hardware, etc.

Workflow for Building the PACE System. This protocol aims to be a guide for establishing a basic, economical PACE system on a lab bench, with particular emphasis on establishing a selection circuit to evolve protein–DNA interactions. The principles outlined here, however, can assist in the design of selection circuits to evolve different protein properties, such as altered substrate specificities or protein–protein interactions as previously described. 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.

A generalized 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 to evolve the protein–DNA interaction of choice and (2) installing the Chemostat-Lagoon pump system for PACE evolution. We roughly estimate that it takes one to two 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.

The two major steps of PACE development are described in separate sections. We start by discussing development of a typical PACE selection circuit followed by the description of the expected results from PACE-B1H evolution on protein ME47 as an example. 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 use the development of the PACE-B1H gIII selection circuit on ME47 as an example to highlight the key concepts in Protocol 1, below. In the Supporting Information, we then provide detailed descriptions of how to set up a basic PACE Chemostat-Lagoon pump system; this protocol includes photographs, timelines, stringency considerations, tips and troubleshooting (Protocol 2: PACE setup, reagents, and protocols). This protocol focuses on the basic PACE setup; more complex PACE procedures are briefly discussed in the Supporting Information (Other permutations of PACE outline
Fig 5. Typical workflow for the development of a PACE selection circuit. Before embarking on 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 the effort invested to perform the refinement of various systems before obtaining a successful PACE circuit. Designing and constructing the selection could take from 0.5 to 1.5 years depending on how much troubleshooting is required. Evolving the protein of interest in PACE once the system is set up could take up to a year to perform the actual experiment, interpret the results, and examine fitness-improving mutations.

additional uses, including split RNA polymerase from the T7 bacteriophage (T7 RNAP), stepping-stone strategy, negative selection, and the drift plasmid.

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. 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. This makes PANCE slower than PACE at developing mutations, but PANCE has the advantage of not requiring a specialized setup. PANCE is useful for the preliminary evolution of a protein of interest to obtain variants with high enough activity to support PACE, which requires a 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.

### RESULTS AND DISCUSSION

**Protocol 1: Developing a PACE-B1H Selection System.** Establishing a proper PACE selection circuit is key to success. Here, we discuss guidelines for the construction of a PACE selection circuit by using the development of our PACE-B1H as an example. In general, 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). The desired protein activity leads to pIII production, which allows for the production of next-generation phage particles (Figure 2B). When the selection circuit is being developed, 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 autoactivated by endogenous molecules, that is, that the system has low background, and (3) verifying that the gIII cassette is only activated through specific interaction with the protein of interest, that is, 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 the 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-<i>E. coli</i> infrastructure. The phage-<i>E. coli</i> 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, for example, 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; for example, the Liu lab utilized LuxA/B as the reporter, and the gIII stop codon overlaps with LuxAB by one base (TA'TG, where the last A of the stop codon of gIII is the first A of the start codon for the Lux reporter). Given the overlapping nature of this construct, the expression of LuxAB depends on gIII transcription and translation such that luciferase activity can be used as an indirect means to monitor any 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 66 amino acids that specifically targets the E-box motif (5'-CACGTG). 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-B1H, SPs carrying variants of ME47 with fitness-improving 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 system was developed by the Wolfe lab to assay the affinity and specificity of a given protein–DNA interaction in E. coli. At the molecular level, the B1H functions similarly to related systems in yeast (Figure 6A; reviewed in ref 28). 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). The system uses HIS3 as the reporter gene, which is located on a second vector, pH3U3. 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 (∆hisBΔpyrF), whose native hisB gene (hisB is the bacterial analogue of the yeast HIS3 gene and will be used to refer to the HIS3 reporter gene) was deleted, to grow in histidine-deficient media. 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. 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.

To construct the B1H reporter system in E. coli, we cloned the sequence expressing the ME47 BHHL 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 (Figure S1). Similarly, we cloned a variant of the E-box sequence, the AT-E-box, into 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.

Figure 6. Schematic diagram illustrating the (A) B1H system and the (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.

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 GCCAAGTGC, 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, 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, from 7 to 25 base pairs (bp) away from the −35 weak promoter (Figure S2).

We tested ME47 against the various GC- and AT-E-box reporter constructs (Figure S2). As before, we saw no growth when ME47 was paired against the −13 AT-E-box. In contrast, other reporters, such as the −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 nonautoactivating 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: for example, 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 3′). 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 6 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 the 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-B1H 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.\(^2,3\)

We then transformed the four AP/E-box vectors separately into E. coli 1030.\(^4\) There are many strains of host cells that can be used for PACE; however, we made use of two, in particular: 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 pIII regardless of the activity from the desired protein from the SP and is a permissive host. The S1030 cells lack the constitutive AP of S1059 and need 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).\(^20\)

In theory, SP-ME47 should be able to activate gIII on the AP/E-boxes via the omega-ME47 construct (Figure 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 plaqueing is a stringent test for propagation, but it is not quantitative. However, overnight propagation tests will give a qualitative measure of phage fitness on the developed selection circuit.\(^12\) 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 subinfecting amount of SP-ME47 for 10 min at 37 °C (typically 10^5 PFU/mL is adequate). It is important that a subinfecting amount of SP is 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 postinfection 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 greater than 10^4-fold propagation for wild-type phage and greater than or equal to 10^5-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 numbers given here are an idealized target for what you should be aiming for when developing your own selection circuit; however, these numbers may vary.

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 selective pressure on the protein to improve. We advise not doing PACE with a circuit that gives more than 10^5-fold overnight propagation on a given circuit, though it may possibly still give useful results depending on the selection that is being performed.

Expected Results from PACE-B1H. Here, we present the seven-day evolution of SP-ME47 as an example. Details regarding SP-ME47 can be found under Protocol 1 above and the Supporting Information.

Experimental Parameters of SP-ME47 PACE-B1H.

- 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: 7 d after the arabinose induction of MP6 is started.
- Sample processing and analysis: Lagoon supernatant sampled and analyzed by plaque assay every day. Five plaques selected each day for Sanger sequencing.

Evolution of SP-ME47 and Analysis of Sequencing Data. The plaque assay analyzing the Lagoon supernatant 1 d after arabinose induction (“Day 1”) produced plaques at an efficiency of 2.3 × 10^6 PFU/mL. The efficiency increased at Day 2 (7.0 × 10^7 PFU/mL) and Day 3 (6.8 × 10^8 PFU/mL) before stabilizing. The 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. Alignment and discussion of the resulting sequences is shown in Figure S5.

We looked for mutations that persisted and eventually dominated in the Lagoon over the seven days of PACE. We distinguish these persistent mutations from spontaneous changes that occur by chance; these persistent mutations allow their SP to outcompete other SPs in the Lagoon, and therefore SP genomes carrying these mutations should increasingly dominate the Lagoon population. We uncovered two mutations, incidentally both at Arg12 of the ME47 CDS, where the wild-type R was mutated to either C (ME47/R12C) or S (ME47/R12S). These two mutants dominated the population by Days 6 and 7 (Figure S5). Therefore, ME47/R12S and ME47/R12C were prioritized for further analysis. Although other mutations were observed in the ME47 CDS, these did not prevail in the population. We judged that these changes were spontaneous and were likely captured at the
Excision of the omega subunit and such that the omega subunit would no longer be subject to selective pressure. 33,34 

Recent advances in PACE have introduced a split translational unit that can then mediate gIII expression (please refer to ref 33 for a thorough description of this setup, as it is beyond the scope of this paper). 

Figure 7. PACE-B1H-derived ME47 mutants 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. The 0 mM 3-AT controls show that spotting for all samples is consistent over 10-fold dilutions from left to right covering 6 orders of magnitude. Both ME47/R12C and ME47/R12S show strong growth until 30 mM 3-AT, in contrast to ME47, where cell growth is completely inhibited. (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 toward the E-box.

Confirming the Effects of Mutations Outside the PACE-B1H System. In general, 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 (e.g., to avoid carryover of an evolved omega subunit of the RNAP to which our transcription factors are fused). It may be desirable to subclone the evolved genes into a standardized phage backbone, that is, 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. Mutations in the omega subunit of RNAP in similar 2-hybrid or 1-hybrid PACE have been documented, 20 reinforcing the necessity of confirming that mutations that arise from PACE-B1H do indeed improve the activity of the protein of interest and not that of the omega subunit. Recent advances in PACE have introduced a split intron strategy, where one has the ability to encode an omega subunit on a complementary plasmid that is akin to the AP, such that the omega subunit would no longer be subject to selective pressure. 33,34 

Excision of the omega subunit and subsequent addition to the protein of interest can be mediated by the addition of small molecules to the functional translational unit that can then mediate gIII expression (please refer to ref 33 for a thorough description of this setup, as it is beyond the scope of this paper).

The C-to-A and C-to-T mutations found in the ME47 CDS correspond to ME47/R12S and ME47/R12C, respectively. To investigate the effect of these mutations, the R12S and R12C mutations were introduced into the wild-type plasmid pB1Hw2/ME47 via site-directed mutagenesis. The resulting mutants, pB1Hw2/ME47/R12S and pB1Hw2/ME47/R12C, were tested in a B1H assay using an E-box regulated hisB gene as a reporter gene (see the Protocols section for a description of the B1H assay). 

Supporting Information discusses a number of parameters that can be tackled toward altering the PACE-B1H selection system; this is not an exhaustive list, however, of changes that can be explored.

The general strategy is to alter the 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. Figure S6 in the Supporting Information discusses a number of parameters that can be altered to change the characteristics of the PACE-B1H selection system. 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. 19 

The C-to-A and C-to-T mutations found in the ME47 CDS correspond to ME47/R12S and ME47/R12C, respectively. To investigate the effect of these mutations, the R12S and R12C mutations were introduced into the wild-type plasmid pB1Hw2/ME47 via site-directed mutagenesis. The resulting mutants, pB1Hw2/ME47/R12S and pB1Hw2/ME47/R12C, were tested in a B1H assay using an E-box regulated hisB gene as a reporter gene (see the Protocols section for a description of the B1H assay).
How Convergent Evolution and 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.\(^3,5\) 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.\(^15,16,26\) For example, 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.\(^15\)

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.\(^16\) Epistasis has also been shown to significantly impact the suite of mutations that are tolerated within a protein.\(^5,16\) 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,\(^3,5\) 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.\(^20\) 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 that different mutations will arise in the protein of interest. This makes techniques, like PANCE,\(^26\) that can parallelize these selections very attractive to use to maximize the portion of the evolutionary landscape that can be sampled.

### ASSOCIATED CONTENT

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c03508.

Discussion of reagents, methods, conditions, hardware setup, troubleshooting, B1H. Plaque assays. Sequencing data. Alterations of stringency. PACE protocols (detailed setup, reagents). Photographs of setup. Diagrams of tubing connections for setup. Photographs of components. Detailed diagrams and ordering of putting together the PACE system. List of materials. Discussion of other variations of PACE (PDF)

### AUTHOR INFORMATION

**Corresponding Author**

Jumi A. Shin — Department of Chemistry, University of Toronto, Mississauga, ON L5L 1C6, Canada; orcid.org/0000-0001-9939-2072; Email: jumi.shin@utoronto.ca

**Authors**

Serban C. Popa — Department of Chemistry, University of Toronto, Mississauga, ON L5L 1C6, Canada

Ichiro Inamoto — Department of Chemistry, University of Toronto, Mississauga, ON L5L 1C6, Canada

### ACKNOWLEDGMENTS

We thank Dr. D. Liu and his lab members, especially B. Hubbard, S. Miller, and A. Raguram, for their assistance in helping us develop our PACE systems and providing plasmids. J.A.S. is grateful for funding from CHRP, NSERC, and the University’s VP Research.

### REFERENCES

(1) Tsien, R. Y. The Green Fluorescent Protein. Annu. Rev. Biochem. 1998, 67, 509–44.
(2) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. Nature 2011, 472, 499–505.
(3) Meng, X.; Brodsky, M. H.; Wolfe, S. A. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nat. Biotechnol. 2005, 23, 988–994.
(4) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. Nat. Rev. Genet. 2015, 16, 379–94.
(5) Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 2009, 10, 866–76.
(6) Liao, H.; McKenzie, T.; Hageman, R. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. Proc. Natl. Acad. Sci. U. S. A. 1986, 83, 576–80.
(7) Salazar, O.; Cirino, P. C.; Arnold, F. H. Thermostabilization of a cytochrome P450 peroxxygenase. ChemBioChem 2003, 4, 891–93.
(8) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. Science 2013, 339, 307–10.
(9) Simon, A. J.; d’Ooidt, S.; Ellington, A. D. Synthetic evolution. Nat. Biotechnol. 2019, 37, 730–743.
(10) Arnold, F. H.; Georgiou, G. Directed Enzyme Evolution; Humana Press, 2003; Vol. 230.
(11) Serebriiskii, I. G.; Golemis, E. A. Two-hybrid system and false positives. Approaches to detection and elimination. Methods Mol. Biol. 2001, 177, 123–134.
(12) Cooper, T. F.; Rozen, D. E.; Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 2003, 100, 1072–77.
(13) Rasala, T. S.; Pajunen, M. I.; Savilahti, H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. Anal. Biochem. 2009, 388, 71–80.
(14) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K.; Javanpour, A. A.; Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. Cell 2018, 175, 1946–57.
(15) Leconte, A. M.; Dickinson, B. C.; Yang, D. D.; Chen, I. A.; Allen, B.; Liu, D. R. A Population-Based Experimental Model for Protein Evolution: Effects of Mutation Rate and Selection Stringency on Evolutionary Outcomes. Biochemistry 2013, 52, 1490–99.
(16) Dickinson, B. C.; Leconte, A. M.; Allen, B.; Esvelt, K. M.; Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc. Natl. Acad. Sci. U. S. A. 2013, 110, 9007–12.
(17) Badran, A. H.; Liu, D. R. In Vivo Continuous Directed Evolution. Curr. Opin. Chem. Biol. 2015, 24, 1–10.
(18) Rakonjac, J.; Bennett, N. J.; Spagnuolo, J.; Gagic, D.; Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr. Iss. Mol. Biol. 2011, 13, 51–76.
(19) Smeal, S. W.; Schmitt, M. A.; Rodrigues Pereira, R.; Prasad, A.; Fisk, J. D. Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. Virology 2017, 500, 259–274.
(20) Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A. M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; Wang, P.; Malvar, T.; Liu, D. R. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature 2016, 533, 58–63.
(21) Badran, A. H.; Liu, D. R. Development of Potent In Vivo Mutagenesis Plasmids with Broad Mutational Spectra. Nat. Commun. 2015, 6, 8425–34.
(22) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Soll, D.; Liu, D. R. Continuous directed evolution of aminocyl-tRNA synthetases. Nat. Chem. Biol. 2017, 13, 1253–60.
(23) Packer, M. S.; Rees, H. A.; Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. Nat. Commun. 2017, 8, 956–66.
(24) Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. 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, A. T.; Shahrvan, S. H.; Shin, J. A. Max-E47, a designed minimalist protein that targets the E-box DNA site in vivo and in vitro. J. Am. Chem. Soc. 2009, 131, 7839–7848.
(26) Roth, T. B.; Woolston, B. M.; Stephanopoulos, G.; Liu, D. R. Phage-assisted evolution of Bacillus methanolicus methanol dehydrogenase 2. ACS Synth. Biol. 2019, 8, 796–806.
(27) Inamoto, I.; Sheoran, I.; Popa, S. C.; Hussain, M.; Shin, J. A. Combining rational design and continuous evolution on minimalist proteins that target DNA. bioRxiv 2020.
(28) Ouwerkerk, P. B. F.; Meijer, A. H. Yeast One-Hybrid Screening for DNA-Protein Interactions. In Current Protocols in Molecular Biology; John Wiley, 2001; pp 12.12.1–12.22.
(29) Dove, S. L.; 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.
(30) Swanson, H. I.; Yang, J.-H. Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site. Nucleic Acids Res. 1999, 27, 3205–3212.
(31) Thuronyi, B. W.; Koblan, L. W.; Levy, J. M.; Yeh, W.-H.; Zheng, C.; Newby, G. A.; Wilson, C.; Bhaumik, M.; Shubina-Oleinik, O.; Holt, J. R.; Liu, D. R. Continuous evolution of base editors with expanded target compatibility and improved activity. Nat. Biotechnol. 2019, 37, 1070.
(32) Carlson, J. C.; Badran, A. H.; Guggiana-Nilo, D. A.; Liu, D. R. Negative Selection And Stringency Modulation Enable Phage-Assisted Continuous Evolution (PACE) of Enzymes With Altered Specificity. Nat. Chem. Biol. 2014, 10, 216–22.
(33) Miller, S. M.; Wang, T.; Randolph, P. B.; Arbab, M.; Shen, M. W.; Huang, T. P.; Matuszek, Z.; Newby, G. A.; Rees, H. A.; Liu, D. R. Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nat. Biotechnol. 2020, 38, 471–81.
(34) Yuen, C. M.; Rodda, S. J.; Vokes, S. A.; McMahon, A. P.; Liu, D. R. Control of transcription factor activity and osteoblast differentiation in mammalian cells using an evolved small-molecule-dependent intein. J. Am. Chem. Soc. 2006, 128, 8939–46.
(35) Wang, X.; Minasov, G.; Shoichet, B. K. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J. Mol. Biol. 2002, 320, 85–95.