Continuous Directed Evolution of Enzymes with Novel Substrate Specificity

Continuous Directed Evolution of Enzymes with Novel Substrate Specificity

A dissertation presented

by

Jacob Charles Carlson

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Chemical Biology

Harvard University
Cambridge, Massachusetts

May 2013

Dissertation Advisor: Professor David R. Liu Jacob Charles Carlson

Continuous Directed Evolution of Enzymes with Novel Substrate Specificity

Abstract

Methodological advances in directed evolution have already made it possible to discover useful biomolecules within months to years. A further acceleration of this process might make it possible to address outstanding challenges, or needs for which the current timescale is a fundamental barrier. To realize these goals would require transformative methodological advances in directed evolution.

In Chapter One, current methods in directed evolution are briefly reviewed. In Chapter Two, a general platform for continuous directed evolution is presented. The method is used to evolve T7 RNA polymerase enzymes with novel promoter activity on the days timescale. In Chapter Three, a method is developed for tuning selection stringency during continuous evolution, thus obviating the requirement for a minimal starting library activity. In Chapter Four, a method is developed for simultaneous positive and negative selection, thus allowing explicit selection for substrate specific enzymes. In Chapter Five, the advances in stringency modulation and negative selection are combined to evolve highly substrate specific enzymes starting from an inactive starting library. In a continuous evolutionary arc of less than three days, we discover T7 RNA polymerase enzymes with a degree of specificity for the T3 promoter exceeding that of the wild type enzyme for its native substrate.

Table of Contents

Acknowledgements

Many parts of this work were done as a collaboration, and it was especially enjoyable to work with Ahmed Badran and Drago Guggiana Nilo. David Liu provided a resource-rich environment that was very helpful for this research.

The lab was made especially enjoyable for me by the good company, as both friends and colleagues, of Vikram Pattanayak, Brent Dorr, John Guilinger, Grace Chen, Ed Curtis, Yevgeny Brudno, Walter Kowtoniuk, Courtney Yuen, and Brian Tse.

I don't think that I would have made it to the end without the care and support of my committee and graduate program. Andrew Murray, Christopher T. Walsh, Dan Kahne, Jason Millberg, and Garth McCavana helped me navigate a very precarious situation, and I am extremely grateful for that. Claire Shindler helped me stay (semi)sane through it. In short, they saved me. I especially want to thank Dan again.

I'm really fortunate to have some wonderful friends and family. Josh has been such a great friend and comrade. Steph has been so supportive and patient with me. Al has been a caring partner and a best friend, and she has given me a family-away-from-home. Dad and Aud are just the best family. The older I get, the more I realize how alike we are, and I feel lucky that I get to know them and be like them. I love these people very much.

Chapter One

Introduction

1.1 Introduction

Directed evolution mimics nature's cycle of diversification, selection, and amplification to discover biomolecules with novel properties; an evolving molecule is "directed" toward a new goal for which selection pressure may not exist in nature[1]. In the diversification stage, a parent gene undergoes mutation to create a "library" of candidate genes, of which some small fraction potentially encode molecules with improved properties. This library is then interrogated, by screening or selection, to identify or enrich variants with the desired property. Finally, the output of this stage is amplified, recovered for analysis, and resubjected to diversification. This laboratory evolution scheme can be used to dramatically improve protein activity toward a variety of user-defined criteria. Two recent examples – far-red fluorescent proteins[2], and site selective transpeptidases[3] – illustrate some key principles of the modern art (Figure 1.1).

Lin et. al. describe an effort to evolve far-red fluorescent proteins with excitation and emission spectra compatible with intravital imaging[2]. The starting point for this evolution was mKate, a red fluoresecent protein with an available X-ray crystal structure[4]. The authors relied heavily on this structural information to design a very small starting library (theoretical complexity of 4e2) with saturation at two residues proximal to the chromophore. This library was easily interrogated in a simple plate-based screen to yield a single mutant with improved brightness and photostability. From this output, a subsequent library design was guided by previous experience with other red-shifted fluorescent proteins, and a modest library (theoretical complexity 1.6e5) with saturation at four positions was screened. An additional two rounds of random mutagenesis and screening yielded mNeptune, a superior red-shifted fluorescent protein that is four mutations away from the parent. Here, the approach made use of prior information

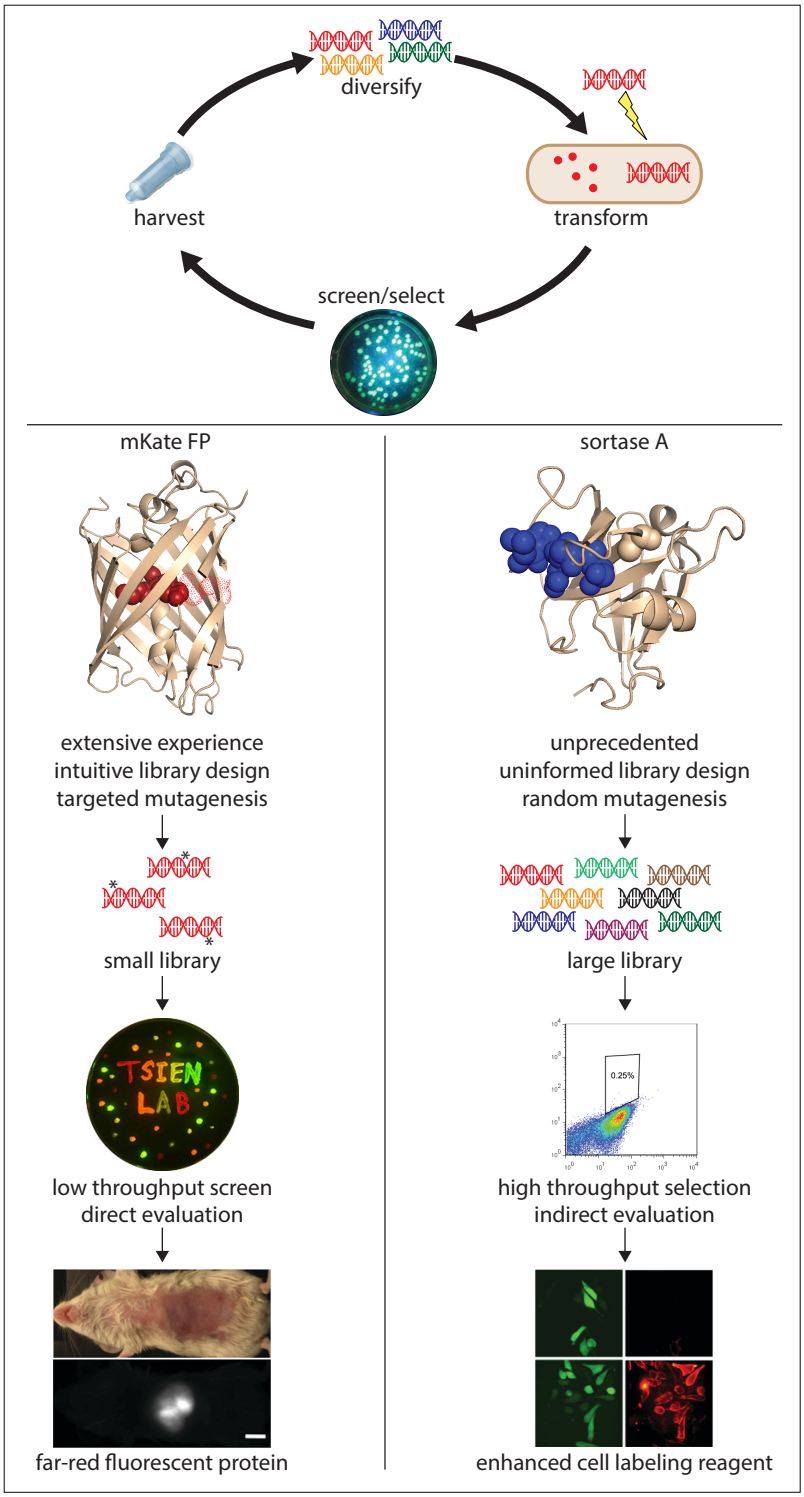


Figure 1.1. Top. The canonical directed evolution cycle. Bottom left. Scheme for discovery of far-red shifted fluorescent proteins (adapted from Lin et. al.[2], plate image from Tsien Lab). Bottom right. Scheme for discovery of more active transpeptidases (adapted from Liu and coworkers[3]).

and experience to guide the design of sequential small, high quality libraries that could be easily interrogated with a simple screen. This success was achieved with remarkable economy.

A more complex challenge was addressed in evolving sortase A variants with improved transpeptidation activity[3]. Here, the crystal structure of sortase A was also known[5], but a wealth of experience and informed intuition was not available to guide the library design for this new evolution target. Thus, the authors chose a large, randomly mutagenized library (complexity 7.8e7) that could not be interrogated with a simple low through-put screen like that used in the previous example. Instead, the authors crafted a selection in which displayed enzymes with high activity would self-label their host cells, allowing them to pass a FACS-based selection. Since the library recovered after sorting was an enriched *mixture*, rather than discretely chosen individual clones, several rounds of sorting were required to bring the population to a high abundance of active clones that could be randomly chosen and analyzed. After four rounds of sorting the output mixture showed some convergence, thus providing key information on possible determinants of improved activity; this information was absent during the initial library design. With this understanding, a second library was generated by in vitro homologous recombination of the post-selection mixture, with a bias toward incorporation of the convergent mutations found in the first rounds. Sorting of this library yielded dramatically improved clones that combined these early mutations. Here, prospective guidance on initial library design was lacking, thus requiring a selection strategy with sufficient power to resolve active variants from a large, randomized library. A subsequent, informed library of high quality was then screened to quickly arrive at active variants.

It is obvious, but worth stating, that the potential success of such efforts is directly related to the probability that unknown hypothetical constructs with optimal or improved activity will be physically created and interrogated. Several measures can be taken to favor success. First, the quality of the library can be improved by using informed design decisions to increase the fraction of library members likely to have the desired activity[6]. Second, the capacity of the screen or selection can be increased so that a larger library can be interrogated in a single round. Finally, the efficiency of the overall protocol can be improved so that more iterations of mutation and selection can be performed in a reasonable timeframe, thus increasing the total number of variants explored. This last improvement allows for the interrogation of advanced libraries that are launched from the improvements gained in earlier rounds; this search strategy can be very efficient if fit variants can be accessed by a reasonably smooth uphill path (a critical, and often untestable assumption)[7]. For outstanding challenges in protein evolution, it is possible that large libraries, with many rounds of selection, will be required.

Methodological advances have already made it possible to discover useful biomolecules - thermostable cellulases[8], site-selective DNA binding domains[9], novel biosynthetic enzymes[10], to name a few - within months to years. Though there is obvious benefit to improving the efficiency of this (or any) discovery process, a drastic acceleration might also make it possible to address demands for which the current timescale is a more fundamental barrier; customized research reagents and patient-tailored therapeutics are two hypothetical applications with time-sensitive prompts. To satisfy these demands would require transformative methodological advances in directed evolution.

This thesis describes the development of a general platform for greatly accelerated in vivo directed evolution (Chapter 2), the addition of drift (Chapter 3) and negative selection (Chapter 4) capabilities to this platform, and the use of these advances (Chapter 5) to invert the specificity of an RNA polymerase in a single evolutionary trajectory of less than three days.

1.2 References

1. Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153-173 (2008).

2. Lin, M. Z. *et al.* Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chem. Biol.* **16**, 1169-1179 (2009).

3. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11399-11404 (2011).

4. Pletnev, S. *et al.* A crystallographic study of bright far-red fluorescent protein mKate reveals pH-induced cis-trans isomerization of the chromophore. *J. Biol. Chem.* **283**, 28980-28987 (2008).

5. Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. & Narayana, S. V. Crystal structures of Staphylococcus aureus sortase A and its substrate complex. *J. Biol. Chem.* **279**, 31383-31389 (2004).

6. Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 447-452 (2005).

7. Voigt, C. A., Kauffman, S. & Wang, Z. G. Rational evolutionary design: the theory of in vitro protein evolution. *Adv. Prot. Chem.* **55**, 79-160 (2000).

8. Heinzelman, P. *et al.* A family of thermostable fungal cellulases created by structure-guided recombination. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5610-5615 (2009).

9. Maeder, M. L., Thibodeau-Beganny, S., Sander, J. D., Voytas, D. F. & Joung, J. K. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat. Protoc.* **4**, 1471-1501 (2009).

10. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440**, 1078-1082 (2006).

Chapter Two

A System for the Continuous Directed Evolution of Biomolecules

Jacob Charles Carlson, Kevin M. Esvelt, David R. Liu

The continuous evolution system was co-developed by Kevin Esvelt and Jacob Carlson. Kevin Esvelt performed the T7 RNA polymerase evolution experiments described here, and designed the initial PACE apparatus.

This chapter is adapted from published work in *Nature*, **472**, 499-503 (2011)

2.1 Introduction

The canonical directed evolution protocol iterates mutagenesis, selection, and amplification until the researcher meets the desired endpoint or a practical time constraint. A typical experimental cycle requires error prone PCR of a starting gene, insertion of the gene library into an expression vector context, transformation of that library into an expression host, treatment of the host to effect selection, amplification of those survivors, recovery of the DNA, and re-entry to diversification. Notably, many of these steps are simple but time consuming DNA manipulations (Figure 2.1).

A typical virus uses self and host machinery to rapidly execute DNA manipulations; the virus injects its encoding information into a new host cell, is amplified, and then recovers its genetic information back into the environment for transfer to a new host. Note that many of the tedious steps in the directed evolution cycle have analogs in the viral life cycle (Figure 2.1).

Husimi and coworkers recognized the potential of a bacteriophage and host cell system to perform continuous evolution experiments[1]. They designed an apparatus to automate the selection of bacteriophages through a constantly renewed population of host cells. Briefly, a population of host cells, maintained in a "turbidostat', was flowed at a constant rate into a "cellstat" that contained a population of M13 bacteriophage. Phage infected newly added host cells, replicated, and new progeny were then secreted back into the medium where they infected more fresh host cells. Husimi demonstrated that more fit M13 phage variants could experience rapid enrichment in this setup, even with small differences in replication rates[2]. They even proposed the use of this system for applied molecular evolution experiments, but this goal went unrealized.

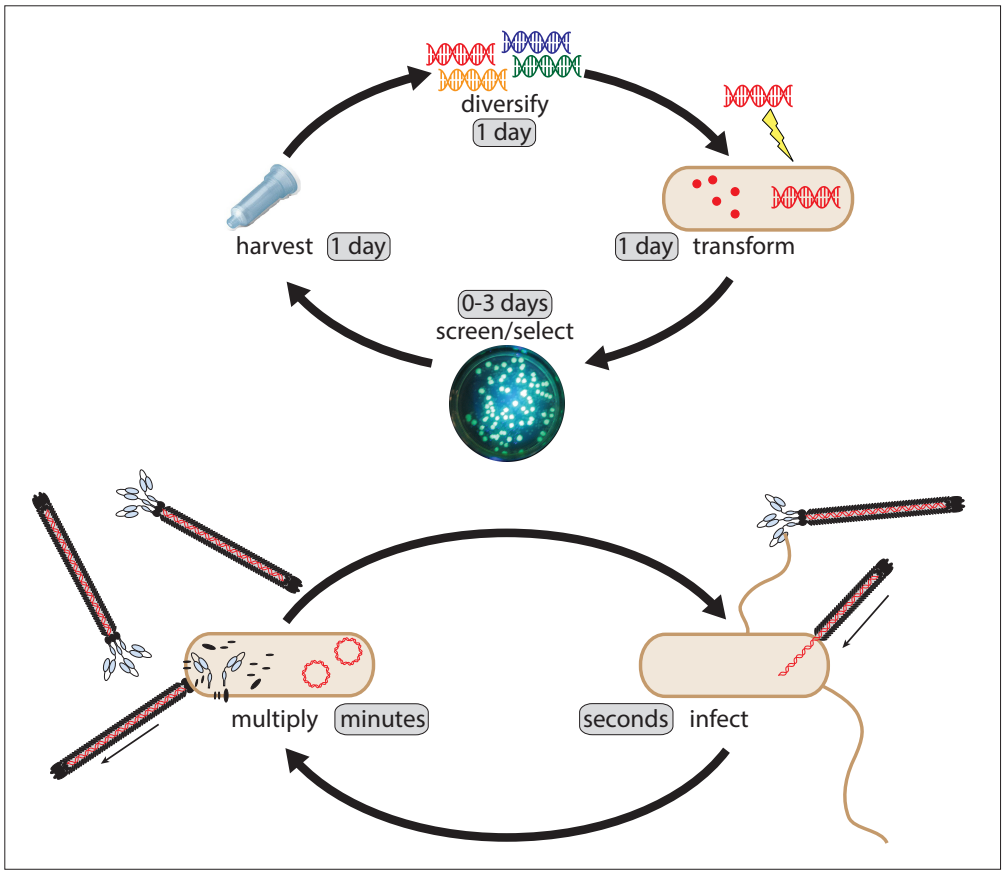


Figure 2.1. Top. The cycle and timescale of conventional directed evolution. Bottom, the cycle and timescale of bacteriophage replication

2.2 Results and Discussion

2.2.1 Principle of a continuous evolution platform

The efficiency of continuous bacteriophage selection can be harnessed for directed evolution by modifying key aspects of the viral life cycle. First, enhanced mutagenesis during viral vector replication is required to constantly renew a diverse library. Second, the autonomy of the native phage life cycle must be replaced with selection-dependent replication; the selection creates a link between the researcher-defined goal of the evolving, virus-borne gene and the replication rate of the encoding virus.

2.2.2 The life cycle of the filamentous bacteriophage

Our strategy for viral evolution utilizes the E. coli male-specific filamentous bacteriophage M13[3]. This phage does not kill its host cell upon release of progeny, it has no biphasic lytic/lysogenic lifestyle, and there are no firm limits on the minimal or maximal genome size that can be propagated[4]. The M13 phage particle is a single-stranded DNA genome encased in a polymeric coat composed of thousands of monomers of the major coat protein pVIII, thus forming a filament (Figure 2.2). At one end of the particle are an estimated five molecules each of the head cap proteins pVII and pIX, while the other end contains five molecules each of the tail cap proteins pVI and pIII[5]. To initiate an infection, pIII binds first to the distal tip of *E. coli* conjugative F-pili[6], which, through their normal dynamics of extension and retraction[7], draw the bound phage to the periphery of the host cell where pIII again interacts with TolA[8], the secondary receptor. The particle is then uncoated as the genome is injected into the cytoplasm of the host cell, and its elegant developmental program begins.

The ssDNA phage DNA is engaged by host polymerases that synthesize the complementary strand[9], thus creating a dsDNA that is the template for further replication, and a substrate for phage transcript synthesis. The genome is transcribed in two cascades[10], one comprised of gII-X-V-VII-IX-VIII and the other of III-VI-I-XI-IV. The locations of promoters, and rates of translation[11] of resulting transcripts, are carefully balanced to afford the correct dynamics and stoichiometry of proteins required to assemble and compose the phage particle. pII, which is likely the first phage protein synthesized, nicks the dsDNA template within the replication origin[12], thus allowing host polymerases to synthesize a new positive strand by rolling circle replication. This positive strand can serve as either the template for a new round of

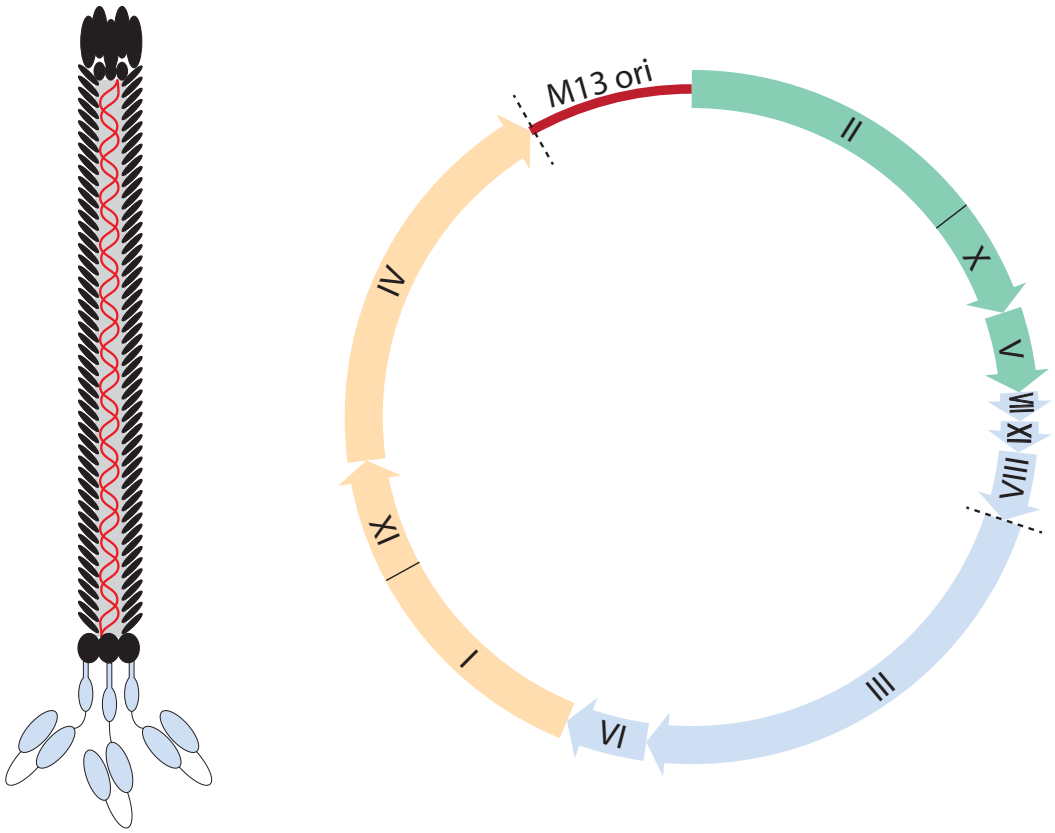


Figure 2.2. Left. Structure of the filamentous phage particle. Right. Genome organization of M13.

complementary strand synthesis, or can be escorted to the cell membrane to be packaged into a progeny phage particle. This decision is regulated by a simple developmental circuit composed of pII, pX, and pV[13]. Early in the replication cycle, pII is abundant and promotes the re-entry of newly synthesized positive strands into the DNA replication cycle by excluding strand access by pV. The concentration of dsDNA (the transcription template) increases and all phage proteins become more abundant. At this stage, pV begins to downregulate expression of pII by binding site specifically to a structure in the translation initiation region of the gII 5' UTR, reducing de novo pII synthesis[14]. As a result, pV outcompetes pII for access to new positive strands, and newly synthesized strands are coated in pV for escort to the membrane. Here, phage particle synthesis is coincident with export; the ssDNA genome is adorned with its protein coat as it is

extruded. This process is catalyzed by the morphogenetic proteins pI, pXI, and pIV[15, 16], which together form an assembly site and extrusion pore that connects to the extracellular medium via a gated channel. Head cap proteins pVII and pIX are added, and the length of the ssDNA sheds a coat of pV in exchange for the major coat protein pVIII[5]. As the extruded particle reaches unit length, pVI and pIII are added to the tail cap of the particle, and pIII catalyzes excision of the particle from the membrane[17], thus releasing the progeny.

2.2.3 Selection-dependent phage replication

This phage life cycle can be made selection-dependent by removing an essential gene from the phage genome, thus decoupling its expression from the native program. The evolving gene can be encoded on the phage vector in place of this essential gene, which is instead borne on a host vector under selection-dependent expression control. Many robust schemes exist that link potential target activities (DNA binding, protein binding, various enzyme activities, etc.) to gene expression[18], thus establishing the crucial connection between phage replication and the activity of interest.

Each phage protein is essential to the replicative program, so in principle it is possible to use any of these genes to form the basis for a selection. However, the phage can be deconstructed in multiple ways, which lead to different requirements for the choice of selection gene (Figure 2.3). The selection phage system is conceptually the simplest. The evolving gene replaces any essential gene on the native phage genome, and this gene is encoded on an accessory plasmid (AP) to be carried by the host cells. Thus, host cells, prior to infection, contain only a single phage gene that is then activated by the evolving gene carried by the phage;

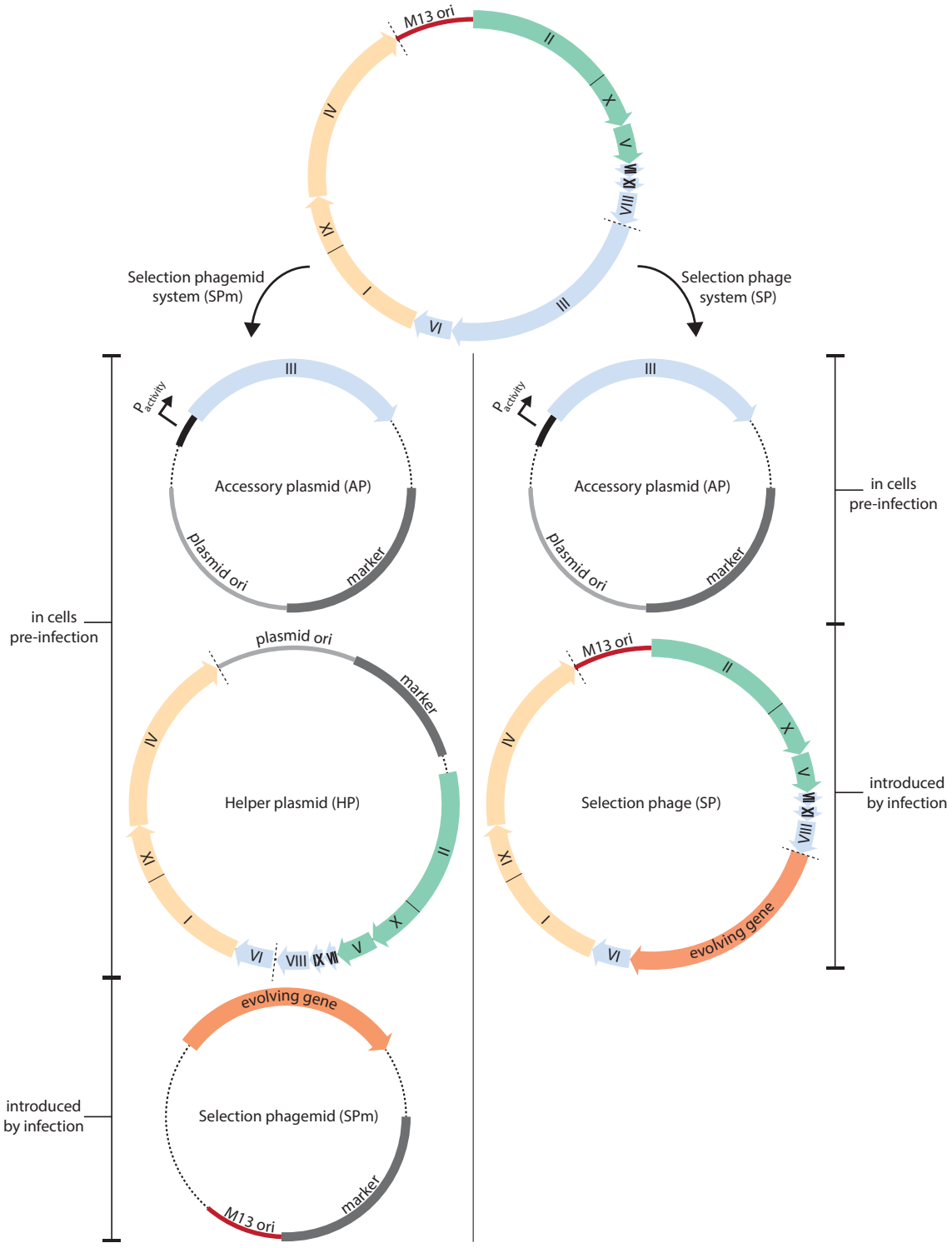


Figure 2.3. Left. A selection phagemid based system deconstructs the M13 genome into three parts. Right. A selection phage based system deconstructs the genome into two parts.

this selection phage also carries in the remaining essential genes to build the progeny particles. In the selection phagemid system, an additional dissection is introduced by constitutively expressing the remaining phage genes from a helper plasmid (HP) carried by the host cells. In this format, only the evolving gene is passaged on the phagemid vector, and all phage genes are present in the host cell prior to infection (with one under selection dependent control on the AP). Here, it is again possible to choose any essential gene to form the basis of the selection, but there is an additional requirement: if any of the phage proteins render host cells resistant to infection, they *must* be unexpressed prior to infection. pIII is well known to render cells resistant to infection[19], so gIII would have to be the basis for the selection in the phagemid format. If placed under activity-dependent expression control, this gIII would remain off prior to infection and could then be stimulated in an activity dependent way by the evolving gene product carried on the phagemid.

The phagemid format was the first one tested. A "helper plasmid" (HP) was constructed in which all phage genes except gIII were encoded on a p15a-origin plasmid (does *not* include the phage replication origin). gIII was then encoded on an "accessory plasmid" (AP) in which gIII transcript synthesis would be controlled by an activity-dependent promoter. Finally, the evolving gene was encoded on a "selection phagemid" (SPm) that also contained the phage replication origin. When combined in a single cell, these plasmids together express all the phage proteins required to assemble the phagemid, which uniquely contains the phage origin and packaging signal, into a phage particle (assuming that the SPm-encoded gene can activate expression from the AP). Indeed, in mock experiments, this format was very capable of producing infectious phagemids with the expected dependence on gIII expression. However,

continuous experiments failed to yield robust propagation. Indeed, even a positive control phagemid, which directly encoded gIII, could not be robustly propagated, suggesting either that phagemid production in continuous culture did not correspond to that observed in discrete experiments, or that host cells were not capable of receiving phagemid via infection.

Using a new assay for host-cell competence, the helper plasmid was identified as a source of infection-resistance in host cells (Figure 2.4). This observation conflicted with previous lab experience, and a published report of a helper plasmid used in phage display[20]. pIII is widely known to render cells resistant to infection by causing membrane destabilization and F-pilus retraction[19], but was not encoded on the helper plasmid, suggesting the influence of another phage protein. pV was identified as a second source of infection resistance, an observation that is also supported by literature precedent[21]. The mechanism is different from that caused by pIII (retraction of F-pili through TolA-mediated pleiotropic effects). In the native phage life cycle, pV coats newly synthesized positive ssDNA strands and escorts them to the membrane for export. Here, when pV is present prior to an infection, the newly infecting strand is subject to a cell fate decision: either host polymerases reach the ssDNA and initiate complementary strand synthesis, or pV reaches the strand first and escorts it back out for export (though the export would be "incomplete" since pIII would not be present). The tuning of pII expression from the HP was used to partially rescue this defect[13] and allowed propagation of phagemids, but it was clear that a selection phage-based system would still be more robust.

This finding motivated the transition to the phage-based system; pIII would remain the basis for the selection and all other phage proteins (including pV) would be introduced by infection with the evolving phage.

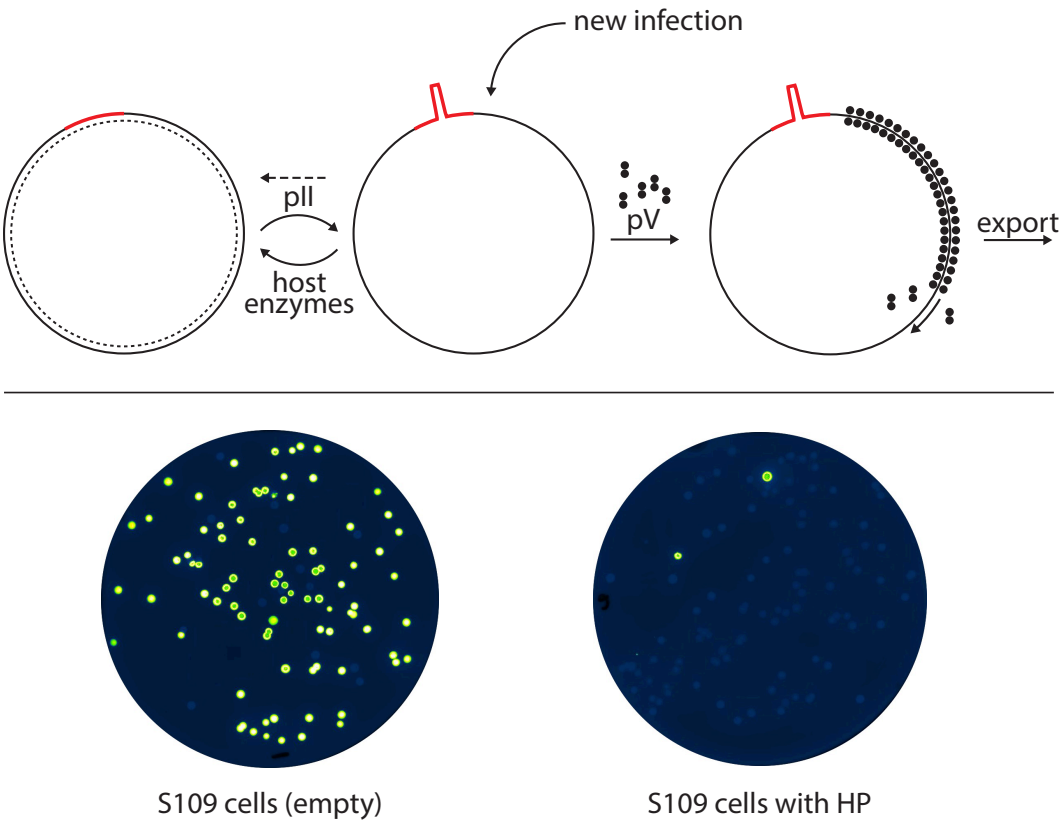


Figure 2.4. Top. Interplay between pV and pII in controlling phage DNA replication. Bottom. Recipient cells carrying HP are less infectable than empty cells. Cells were incubated with phagemids encoding YFP and then plated. The fraction of yellow colonies (which received the phagemid by infection) reports on the infectability of the original recipient culure.

2.2.4 Implementation of PACE system

In phage-assisted continuous evolution (PACE), *E. coli* host cells continuously flow through a fixed-volume vessel (the "lagoon") containing a replicating population of phage DNA vectors ("selection phage", SP) encoding the gene(s) of interest. The average residence time of host cells in the lagoon is less than the time required for *E. coli* replication. As a result, mutations accumulate only in the evolving SP population, the only DNA that can replicate faster than the rate of lagoon dilution. The mutation of host cells in the lagoon should therefore have minimal

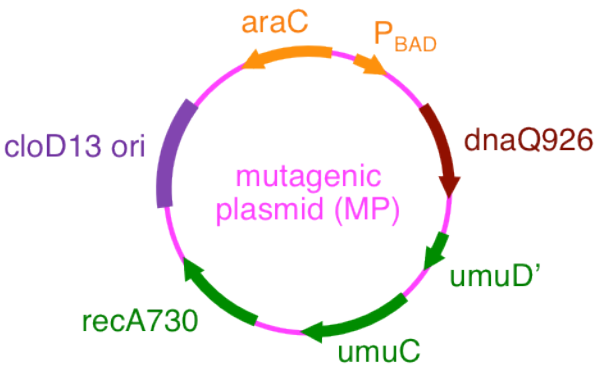
impact on the outcome of the selection over many rounds of phage replication, and mutagenesis conditions are not limited to those that preserve E. coli viability.

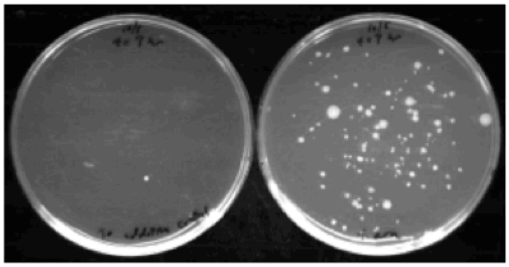
Due to the speed of the phage life cycle (progeny phage production begins ~10 minutes post-infection)[22], PACE can mediate many generations of selective phage replication in a single day. We observed activity-dependent phage vectors that tolerate lagoon flow rates up to 3.2 volumes per hour, corresponding to ~115 population doublings and an average of ~38 phage generations per 24 hours. More conservative flow rates of 2.0-2.5 volumes per hour allow 24-30 generations per day and reduce the risk of complete phage loss (washout) during selections. Multiple lagoons can evolve genes in parallel, with each 100 mL lagoon containing ~5 x 10^{10} host cells selectively replicating active phage variants. Importantly, PACE requires no intervention during evolution and obviates the need to create DNA libraries, transform cells, extract genes, or perform DNA cloning steps during each round.

2.2.5 Enhanced mutagenesis

We constructed an arabinose-inducible mutagenesis plasmid (MP) that elevates the error rate during DNA replication in the lagoon by suppressing proofreading and enhancing error-prone lesion bypass[23, 24]. Full induction increased the observed mutagenesis rate by ~100-fold (Figure 2.5), inducing all possible transitions and transversions. This enhanced mutation rate is sufficient to sample all possible single and double mutants of a given sequence each generation, in principle enabling single-mutation fitness valleys to be traversed during PACE.

The basal mutation rate of replicating filamentous phage (5.3×10^{-7} substitutions/bp) is sufficient to generate all possible single but not double mutants of a given library member in a





observed mutations (*n* = 72)

G/C to A/T: 43%
A/T to G/C: 12%
G/C to T/A: 15%
A/T to C/G: 8%
G/C to C/G: 7%
A/T to T/A: 10%

mutant colonies after 3 hours

	– arabinose	+ arabinose
	1	114
	2	204
	9	861
	1	60
total:	13	1239

Figure 2.5. A mutagenesis plasmid (MP) increases the mutation rate in a lagoon by up to ~100-fold. Phage bearing lacI were propagated with and without arabinose and used to infect a reporter strain that expresses spectR under lacI control. Colonies indicate infection with phage carrying mutated lacI.

100 mL lagoon. For a target gene 1,000 base-pairs in length, a basal mutation rate of $5x10^{-7}$ yields $2.5x10^{7}$ base substitutions spread over $5x10^{10}$ copies of the gene in a 100 mL lagoon after one phage generation, easily covering all single point mutants but not all double mutants. Arabinose induction of the MP can increase the mutation rate to ~$5x10^{-5}$, yielding ~$2.5x10^{9}$ mutations spread over $5x10^{10}$ copies of the gene after one generation. The vast majority are single mutants which together comprise a target area of ~$2.5x10^{12}$ base pairs, suggesting that some $1.2x10^{8}$ are double mutants, sufficient to cover all $9x10^{6}$ possible double mutants.

2.2.6 Evolution of T3 promoter recognition by T7 RNA polymerase

Bacteriophage T7 RNA polymerase (T7 RNAP) is widely used to transcribe RNA in vitro and in cells. T7 RNAP is highly specific for its promoter sequence (TAATACGACTCACTATA),

and exhibits virtually no detectable activity on the consensus promoter of the related bacteriophage T3 (AATTAACCCTCACTAAA, differences underlined)[25, 26]. Despite decades of study and several attempts to engineer the specificity of T7 RNAP towards other promoters[27] including that of T3, a mutant T7 RNAP capable of recognizing the T3 promoter has not been previously reported.

To remove potential interference from evolutionary improvements to the phage vector rather than to T7 RNAP, we propagated an SP expressing wild-type T7 RNAP for three days on host cells containing an AP with the wild-type T7 promoter driving gIII expression. A single plaque presumed to represent vector-optimized SP contained a single mutation (P314T) in T7 RNAP. We confirmed that the activity of the P314T mutant does not significantly differ from that of wild-type T7 RNAP.

This starting SP failed to propagate on host cells containing the T3 promoter AP. We therefore propagated the phage on cells containing a hybrid T7/T3 promoter AP with the T7 promoter base at the important -11 position[25] but all other positions changed to their T3 counterparts. Two initially identical lagoons were evolved in parallel on the hybrid promoter AP for 60 hours, then on the complete T3 promoter AP for 48 hours, and finally on a high-stringency, very low-copy T3 promoter AP for 84 hours(Figure 2.6).

In both lagoons phage persisted after 8 days of PACE, surviving a net dilution of 10^{167} fold, the equivalent of 555 phage population doublings and ~200 rounds of evolution by the average phage. We isolated, sequenced, and characterized phage vectors from each lagoon after 48, 108, and 192 hours, observing up to eight, ten, and 11 non-silent mutations in single T7 RNAP genes at each time point.

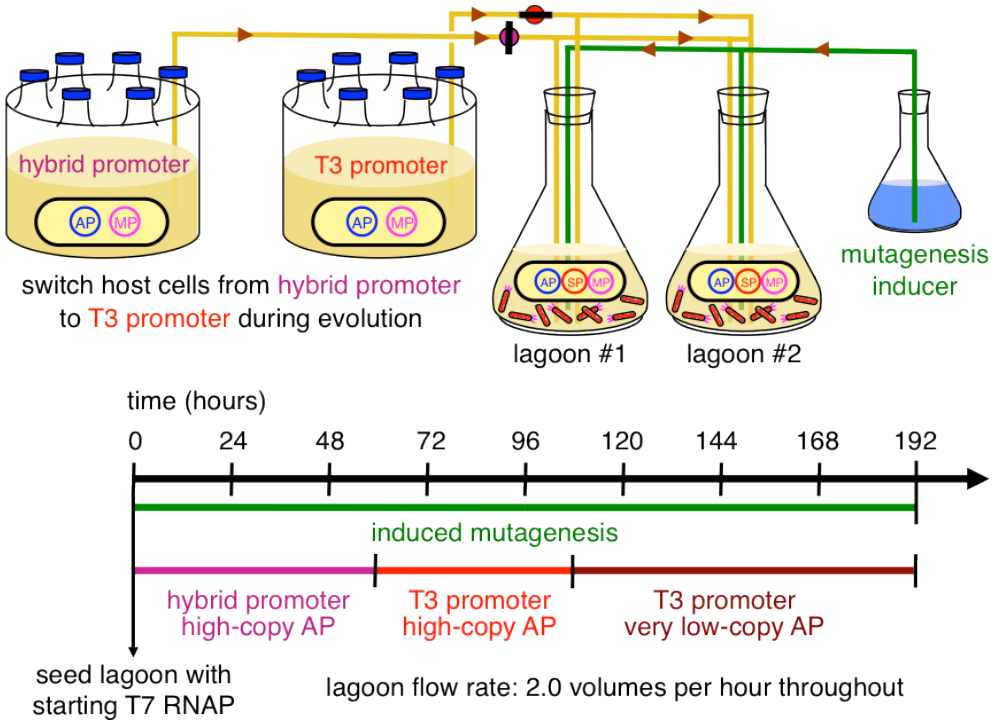


Figure 2.6. Schedule for evolution of T7 RNAP mutants that can recognize the T3 promoter.

Protein-encoding regions (without upstream promoter sequences) of evolved mutant T7 RNAP genes were subcloned into assay plasmids that quantitatively link transcriptional activity to beta-galactosidase expression in cells[28]. We defined the activity of wild-type T7 RNAP on the T7 promoter to be 100%. The starting T7 RNAP exhibited undetectable (< 3%) levels of activity on the T3 promoter in these cell-based assays. The assayed mutants exhibited > 200% activity after 108 hours of PACE, and > 600% activity following high-stringency PACE at 192 hours, improvements of more than 200-fold (Figure 2.7). These results collectively establish the ability of PACE to very rapidly evolve large changes in enzyme activity and specificity with minimal intervention by the researcher.

Several evolved T7 RNAP mutants were also purified and assayed in vitro using radioactive nucleotide incorporation assays. Purified T7 RNAP mutants exhibited activity levels

on the T3 promoter in vitro exceeding that of wild-type T7 RNAP on the T7 promoter, representing improvements of up to 89-fold compared with the starting enzyme. These results indicate that PACE resulted in large improvements in substrate binding or catalytic rate. Evolved

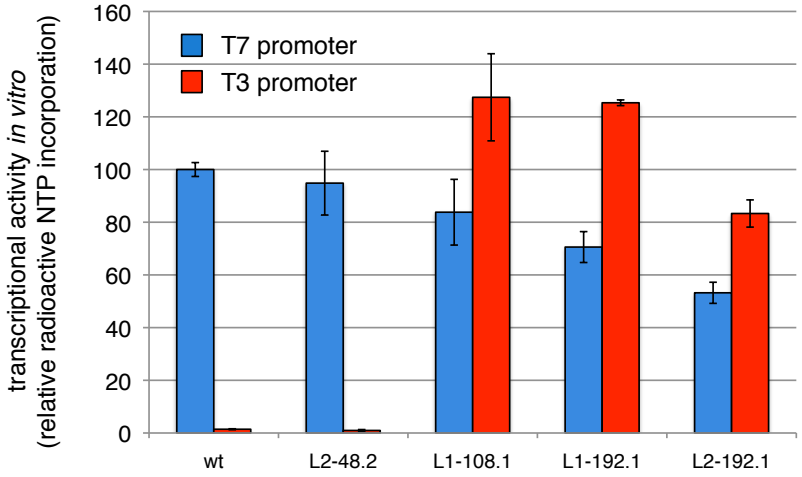


Figure 2.7. In vitro activity of evolved RNAP clones

activity improvements were higher in cells than in vitro, suggesting that these enzymes also evolved improvements in features such as expression level, polymerase folding, or stability that are specific to the context of the cytoplasm.

2.3 Materials and Methods

2.3.1 General Materials and Methods

All DNA cloning was performed with Mach1 cells or NEB Turbo cells. Early discrete infection assays and PACE experiments were performed with PirPlus DH10βF'DOT cells. Plaque assays and PACE experiments with T7 RNAP were performed with using *E. coli* S109 cells derived from DH10B by replacement of the proBA locus with the *pir116* allele, as previously described[29]. This modification enables R6K based phagemids to replicate in recipient cells,

which is needed for some infectability assays. Similarly, the *lacI* cassette was deleted from the F plasmid and from the chromosome to enable mutagenesis assays. The modified F plasmid derives from *E. coli* ER2738 (NEB) The complete genotype of the resulting strain is F´proA+B+ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupG rpsLΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ^-.

Discrete (non-continuous) infectivity assays were performed by co-transforming phage bearing antibiotic resistance genes with the appropriate accessory plasmid into competent cells to generate phage donors. Colonies were picked and grown overnight in 2xYT media with both antibiotics. 2 µL donor cells were mixed with 198 µL F^+ recipient cells in mid-exponential phase containing an antibiotic resistance gene not found in the donor. Mixtures were incubated at 37 °C for 1.5 hours and 20 µL of serial dilutions were spread on plates containing the donor and the recipient antibiotics. Infection was quantified by the number of resulting colonies after incubation at 37 °C overnight. For plaque assays, phage DNA was transformed into electrocompetent cells containing the appropriate accessory plasmid and recovered for 1 hr at 37 °C, or simply isolated from a lagoon. Serial dilutions were mixed with 300 µL F^+ recipient cells grown to exponential phase in 14 mL Falcon culture tubes and incubated at 37°C for 15 minutes. 3 mL top agar (7 g/L from LB broth base) at 50 °C was added to each tube, briefly vortexed, and poured onto minimal agar plates incubated at 37 °C for 8 hours or overnight to generate plaques.

2.3.2 Assay for infection-competence of recipient cells

Recipient cultures were grown to mid-log phase, and 100 μL of culture was mixed by pipetting with 2 μL of pJC126b phagemid prep (encodes R6K origin, f1 origin, pLac-YFP) totaling ~1e9 cfu. Reactions were incubated for 2 minutes and then stopped by transfer of 1 μL to 1 mL of fresh 2xYT media. Cells were pelleted, resuspended in 1 mL fresh 2xYT, and an aliquot was diluted 10-fold into fresh media. 20 uL was plated to 2xYT-agar plates with 1 mM IPTG and grown overnight. Colonies were scanned on a Typhoon laser imager (488 nm laser, 520/40 filter).

2.3.3 Turbidostat and lagoon setup and operation

Turbidostats were constructed from BioProbe flasks on magnetic stir plates. Each flask was equipped with a TruCell2 cell density meter held in a GL32 probe holder with compression fitting. GL45 and GL32 septa pierced with needles transferred media to and from the turbidostat via an 8-channel peristaltic pump with Tygon tubing. A needle set at the desired turbidostat volume level pumped excess cells to the waste container. A 0.2 µm filter attached to a 14-gauge needle piercing the septum vented the turbidostat vessel. A two-way valve controlled media flow to the turbidostat, connected such that a closed valve state returned the media to source. The valve opened and closed in response to TruCell2 4-20 mA output processed by a digital panel meter programmed with the desired set point. Panel meters were unlocked according to the instruction manual and programmed to the following settings: Input dc_A, Setup 30_ 10, Config 00000, Filtr 11009, dec.pt ddd.dd, lo in 00.400, lo rd 004.00, hi in 02.000, hi rd 020.00, Alset 00036, deu1h 000.01, deu2h 000.01. The digital panel meter, adapter, and valve were connected

with a solderless breadboard according to the diagram shown in Supplementary Fig. 12. Lagoons consisted of 100 mL Pyrex bottles with GL45 septa pierced with needles for fluid delivery, a 0.2 μm filter-terminated vent line, and a magnetic stir bar. Excess lagoon volume was continually pumped to waste via a waste needle set at the desired lagoon volume.

Each 20 L media carboy received 140 g anhydrous potassium phosphate dibasic, 40 g potassium phosphate monobasic, 20 g ammonium sulphate, and 20 mL Tween 80 in 20 L H_2O. Carboys were loosely capped with Polyvent filling/venting closures with an autoclavable 0.2 μm filter fastened to the venting port. Media was autoclaved until visibly boiling (typically 120 min at 30 psi, 121 °C) and allowed to cool overnight. Media supplement was prepared from 90 g glucose, 10 g sodium citrate, 0.25 g anhydrous magnesium sulphate, 10 g casamino acids, 0.15 g tetracycline-HCl, 0.6 g carbenicillin, 0.6 g spectinomycin, and 0.5 g (L)-leucine, dissolved in 500 mL H_2O, and filtered with a Nalgene 500 mL filtration unit. 500 mL media supplement was added to each carboy under conditions that minimize the risk of media contamination (in the case of the reported experiments, immediately following 1 hour of germicidal UV irradiation).

The autoclavable components of a turbidostat apparatus include the BioProbe flask, TruCell2 probe, needles, vent filter, and tubing. All such components were autoclaved fully assembled except for tubing, which was connected while hot. Upon equilibrating to ambient temperature, the peristaltic pump responsible for media addition and waste removal was started with the valve opened until the desired volume was reached. The TruCell2 probe was connected to its transmitter and zeroed. Turbidostats were seeded with 100 μL of an overnight culture of host cells. Turbidostats and lagoon cultures were grown at 37 °C.

Serial dilution plating was used to generate a calibration curve to determine TruCell2 output and panel meter setting corresponding to the desired cell density. For these experiments, both panel meter alarms were programmed to open the valve at 6.80 mA. Cells were pumped from the turbidostat to the lagoons via peristaltic pumps with silicone (platinum) two-stop tubing. Calibration curves relating pump speed in rpm to volumetric flow rate were determined experimentally with a timer and graduated cylinder for each tubing size.

2.3.4 PACE experiments

Turbidostats and lagoons were assembled as described above. Upon the turbidostat reaching the desired set point of 6.80 mA (corresponding to 5 x 10^8 cells/mL in our hands), lagoons were connected to turbidostats, waste needles were set at the desired volume, and lagoon pumps were set to a flow rate corresponding to the desired dilution rate. Each lagoon was seeded with 100 μL of an overnight culture producing selection phage. To induced elevated mutagenesis, 10% filter-sterilized (L)-arabinose was delivered by a separate peristaltic pump to each lagoon requiring enhanced mutagenesis to a final concentration of 1%. Lagoon aliquots were taken by sampling lagoon waste lines at the luer lock just after the peristaltic pump. Individual clones were isolated by plaque assay or amplified by PCR, assembled into a T7 RNAP activity assay plasmid, and transformed into cells containing a *lacZ* reporter plasmid. Active clones were picked by blue/white screening.

2.3.5 Selection phage optimization.

T7 RNAP was subcloned into VCSM13 helper phage encoding kanamycin resistance, generating HP-T7RNAP A. To ensure that improvements to the phage genome did not interfere with the evolution of T7 RNAP, HP-T7RNAP A was propagated in a lagoon fed by S109 host cells containing AP-T7 A and DP-QUR and supplemented with arabinose for 72 hours at 2.0 volumes/hour. Individual plaques were isolated and their T7 RNAP genes sequenced. One plaque contained only a single point mutation in T7 RNAP, P314T, and was chosen as the preoptimized selection phage for T7 RNAP evolution. Sequencing of the rest of the selection phage revealed numerous changes relative to the parental VCSM13. Notably, the entire p15a-Kan^R cassette inserted into the intergenic (IG) region to create VCSM13 had been perfectly deleted to reconstitute the wild-type M13 IG region. Other changes included N79S, F286S, and I360T mutations in gIV, a K249R mutation in gII, three silent mutations back to the corresponding M13 base, two other silent mutations, and the deletion of one thymine residue in the terminator before gIII, possibly increasing the expression of T7 RNAP. These regional patterns of variation parallel those observed by Husimi in more extensive filamentous phage evolution experiments. This evolved phage, designated SP-T7RNAP P314T, was used as the starting selection phage for all subsequent PACE experiments.

2.3.6 Mutagenesis assays

The *lacI* gene was cloned into VCSM13 between the p15a origin and kan^r to generate VCSM13-lacI. A turbidostat was grown to a set point equivalent to $5x10^8$ cells/mL with S109 cells containing the mutagenesis plasmid (MP). Lagoons were seeded with 10 μL VCSM13-lacI

and run at 2.5 volumes/hour for 3 hours. One lagoon was supplemented with 10% arabinose to a final concentration of 1%, while the other was not. Aliquots were removed after 3 hours and each was used to infect a 100-fold greater volume of recipient cell culture of S109 cells containing a reporter plasmid conferring carbenicillin and spectinomycin resistance with a *lacI* binding site (*lacO*) capable of repressing spectinomycin resistance. Mixtures were incubated for 1.5 hr at 37°C, spread on 2xYT plates containing spectinomycin, kanamycin, and carbenicillin, and incubated at 37°C overnight. Colonies were counted for induced and uninduced lagoons to estimate the fold increase in mutagenesis. 72 colonies were sequenced to determine the frequencies of all transitions and transversions within mutated *lacI* genes. All sequenced colonies contained at least one mutation capable of inactivating repressor function.

2.3.7 Cell-based T7 RNAP activity assays

Overnight cultures of S109 cells grown in 2xYT containing reporter plasmid and expression vector were diluted 4-fold in fresh 2xYT media. 20 µL of the diluted culture was mixed with 80 µL Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM beta-mercaptoethanol, pH 7.0) in Falcon Microtest 96-well OptiLux assay plates and the absorbance at 595 nm was measured using a Spectra M5 plate reader. 25 µL of 1 mg/mL methylumbelliferyl-beta-(D)-galactopyranoside was added to each well and the time recorded. Plates were incubated at 30 °C and fluorescence was measured at 360/460nm on a Spectra M5 plate reader. Plates were measured at multiple time points to avoid saturation of the spectrophotometer or consumption of the substrate, depending on the activity level of the T7 RNAP enzyme being assayed. MUG fluorescence units were calculated as previously described.

The activity level of wild-type T7 RNAP on the T7 promoter was defined as 100%; activities > 3% were considered significantly above the background level of this assay.

2.3.8 T7 RNAP protein purification

T7 RNAP variants were cloned into pT7-911Q, a His-tagged T7 RNAP expression vector. Overnight cultures grown at 30°C were diluted 1:500 in LB broth containing 50 µg/mL carbenicillin and 2% glucose. Upon reaching OD_{600} = ~0.5, cultures were centrifuged at 4000 *g* for 5 minutes and resuspended in LB broth with 0.4 mM IPTG and 50 µg/mL carbenicillin. Cultures were grown for 4 hours at 30°C, spun at 8000 *g* for 6 minutes, and the pellet was frozen overnight. Binding buffer consisted of 50 mM Tris, 300 mM NaCl, 5% glycerol, 5 mM beta-mercaptoethanol, and 10 mM imidazole at pH 8.0. Wash buffer consisted of 50 mM Tris, 800 mM NaCl, 5% glycerol, 5 mM beta-mercaptoethanol, and 20 mM imidazole at pH 8.0. Elution buffer was equivalent to wash buffer with 500 mM imidazole. Pellets from 25 mL culture were resuspended in 1 mL wash buffer and cells were lysed by sonication while kept on ice using a Misonix CL4 sonicator at maximal microtip power for 45 seconds in 1-second bursts. Cell debris was spun down at 20,000 *g* for 15 minutes at 4°C. Ni-NTA spin columns were equilibrated with 500 µL binding buffer and spun at 800 *g* for 2 minutes. Lysate supernatant was loaded onto each column and spun at 300 *g* for 5 minutes. Columns were washed twice with 500 µL wash buffer, spinning at 800 *g* for each, then eluted twice with 250 µL elution buffer. Proteins were dialyzed into 20 mM Tris, 100 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mM DTT, pH 8.0, and concentrated using Amicon Ultra-0.5 30K concentration columns.

2.3.9 *In vitro* T7 RNAP activity assays

Purified T7 RNAP variant concentrations were determined by Bradford assay and then by Coomassie stain on a 4-12% NuPage gel. Templates were prepared by PCR amplification of 150 bp fragments of the reporter plasmids used for *in vivo* assays including the promoter and the start of the lacZ gene. Templates were purified by MinElute spin column. Transcription reactions were performed in 1x RNA polymerase buffer consisting of 40 mM Tris-HCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 2 mM spermidine pH 7.9 with 1 mM rNTPs, 1 ng template DNA, purified polymerase variant, and 2 mCi [α-^{32}P]ATP. Reactions were incubated at 37 °C for 20 minutes, mixed with an equivalent volume of loading dye consisting of 7 M urea, 178 mM Tris-Cl, 178 mM H_3BO_3, 4 mM EDTA and 0.002% bromophenol blue, then electrophoresed on Criterion 5% or 10% TBE-urea denaturing gels. RNAs were transcribed from double-stranded templates of sequence

5’-TAATACGACTCACTATA**G**GGAGAGCCACCACCACCACCACCACCA-3’,

5’- TAATACGACTCACTATA**C**CCCCCGCCACCACCACCACCACCACCA -3’, and 5’-TAATACGACTCACTATA**A**AAAAAGCCACCACCACCACCACCACCA -3’ (the +1 base is underlined in bold). Transcripts were electrophoresed on Criterion 15% TBE-urea gels, exposed to phosphor screens, and imaged on a Typhoon Trio phosphorimager. Bands corresponding to transcription products were quantified with ImageJ software.

2.4 References

1. Husimi, Y. Selection and evolution of bacteriophages in cellstat. *Adv. Biophys.* **25**, 1-43 (1989).

2. Husimi, Y., Nishigaki, K., Kinoshita, Y. & Tanaka, T. Cellstat-a continuous culture system of a bacteriophage for the study of the mutation rate and the selection process of the DNA level. *Rev. Sci. Instrum.* **53**, 517-522 (1982).

3. Salivar, W.O., Tzagoloff, H. & Pratt, D. Some physical-chemical and biological properties of the rod-shaped coliphage M13. *Virology* **24**, 359-371 (1964).

4. Lopez, J. & Webster, R. E. Morphogenesis of filamentous bacteriophage f1: orientation of extrusion and production of polyphage. *Virology* **127**, 177-193 (1983).

5. Endemann, H. & Model, P. Location of filamentous phage minor coat proteins in phage and in infected cells. *J. Mol. Biol.* **250**, 496-506 (1995).

6. Jacobson, A. Role of F pili in the penetration of bacteriophage fl. *J. Virol.* **10**, 835-843 (1972).

7. Clarke, M., Maddera, L., Harris, R. L. & Silverman, P. M. F-pili dynamics by live-cell imaging. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17978-17981 (2008).

8. Riechmann, L. & Holliger, P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell* **90**, 351-360 (1997).

9 Geider, K. & Kornberg, A. Conversion of the M13 viral single strand to the double-stranded replicative forms by purified proteins. *J. Biol. Chem.* **249**, 3999-4005 (1974).

10. Edens, L., Konings, R. N. & Schoenmakers, J. G. A cascade mechanism of transcription in bacteriophage M13 DNA. *Virology* **86**, 354-367 (1978).

11. Blumer, K. J., Ivey, M. R. & Steege, D. A. Translational control of phage f1 gene expression by differential activities of the gene V, VII, IX and VIII initiation sites. *J. Mol. Biol.* **197**, 439-451 (1987).

12. Dotto, G. P., Horiuchi, K. & Zinder, N. D. Initiation and termination of phage f1 plus-strand synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7122-7126 (1982).

13. Fulford, W. & Model, P. Bacteriophage f1 DNA replication genes. II. The roles of gene V protein and gene II protein in complementary strand synthesis. *J. Mol. Biol.* **203**, 39-48 (1988).

14. Model, P., McGill, C., Mazur, B. & Fulford, W. D. The replication of bacteriophage f1: gene V protein regulates the synthesis of gene II protein. *Cell* **29**, 329-335 (1982).

15. Haigh, N. G. & Webster, R. E. The pI and pXI assembly proteins serve separate and essential roles in filamentous phage assembly. *J. Mol. Biol.* **293**, 1017-1027 (1999).

16. Linderoth, N. A., Model, P. & Russel, M. Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. *J. Bacteriol.* **178**, 1962-1970 (1996).

17. Rakonjac, J. & Model, P. Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282**, 25-41 (1998).

18. Vidal, M. & Legrain, P. Yeast forward and reverse n-hybrid systems. *Nucleic Acids Res.* **27**, 919-929 (1999).

19. Boeke, J. D., Model, P. & Zinder, N. D. Effects of bacteriophage f1 gene III protein on the host cell membrane. *Mol. Gen. Genet.* **186**, 185-192 (1982).

20. Chasteen, L., Ayriss, J., Pavlik, P. & Bradbury, A. R. M. Eliminating helper phage from phage display. *Nucleic Acids Res.* **34**, e145 (2006).

21. Terwilliger, T. C., Fulford, W. D. & Zabin, H. B. A genetic selection for temperature-sensitive variants of the gene V protein of bacteriophage f1. *Nucleic Acids Res.* **16**, 9027-9039 (1988).

22. Calendar, R. The Bacteriophages (Oxford Univ. Press, 2006)

23. Fijalkowska, I. J. & Schaaper, R. M. Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2856-2861 (1996).

24. Opperman, T., Murli, S., Smith, B. T. & Walker, G. C. A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9218-9223 (1999).

25. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228**, 506-515 (1992).

26. Ikeda, R. A., Chang, L. L. & Warshamana, G. S. Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. *Biochemistry* **32**, 9115-9124 (1993).

27. Raskin, C. A., Diaz, G. A. & McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3147-3151 (1993).

28. Vidal-Aroca, F. *et al.* One-step high-throughput assay for quantitative detection of beta-galactosidase activity in intact gram-negative bacteria, yeast, and mammalian cells. *BioTechniques* **40**, 433-434, 436, 438 passim (2006).

29. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640-6645 (2000).

Chapter Three

A Method for Stringency Modulation in Continuous Evolution

Jacob Charles Carlson, Drago Guggiana Nilo, David R. Liu

Drago Guggiana Nilo built the real-time luminescence monitor. Jacob Carlson designed and performed all other experiments.

3.1 Introduction

We previously used PACE to evolve T7 RNA polymerase mutants that initiate transcription at the T3 promoter, which differs from the T7 promoter at 6 of 17 base positions. We first challenged the wild-type enzyme to evolve activity directly on the T3 promoter, but this attempt failed, suggesting that the stringency of this initial selection was too high. We then designed a hybrid T7/T3 promoter to serve as a lower-stringency evolutionary stepping stone, and from the library evolved on this hybrid promoter the transition to recognition of the full T3 promoter was realized. Fortunately, our hybrid promoter design was guided by extensive literature knowledge of the sequence determinants of T7 and T3 promoter recognition[1]. However, we recognize that for many potential applications of PACE, the choice of suitable intermediate substrates may not be intuitive or informed by literature precedent. Therefore, such a strategy lacks immediate generality and may limit the extent to which a protein activity can be practically altered. To address this limitation, we designed a generic modulator of stringency that can be engaged and tuned with the simple addition of a small molecule, and that operates irrespective of the activity being evolved. Using this advance, we demonstrate that T3-active polymerases can be rapidly evolved from the T7-specific wild-type enzyme without the use of intermediate substrates.

Technical advances in the use of a chemostat grown recipient cell culture, and a real-time monitor of luminescence, were used in this and subsequent chapters and are described here.

3.2 Results and Discussion

3.2.1 Development of chemostat culture and real-time monitor

Previous continuous evolution experiments made use of a turbidostat to maintain a regular culture of recipient cells. This setup is expensive and prone to technical difficulty, so a chemostat culture system was implemented. Cells grown in chemostat culture are at high density, which exacerbates problems with cell clumping and clogging of tubes. Disruption of biofilm genes (as described in Materials and Methods) resolved issues with clumping.

The ability of chemostat grown cells to support propagation was found to be growth rate dependent. At chemostat dilution rates of ~1.6 vol/hr, propagation of a positive control SP-T7 on cells bearing a pT7-gIII cassette was robust (Figure 3.1).

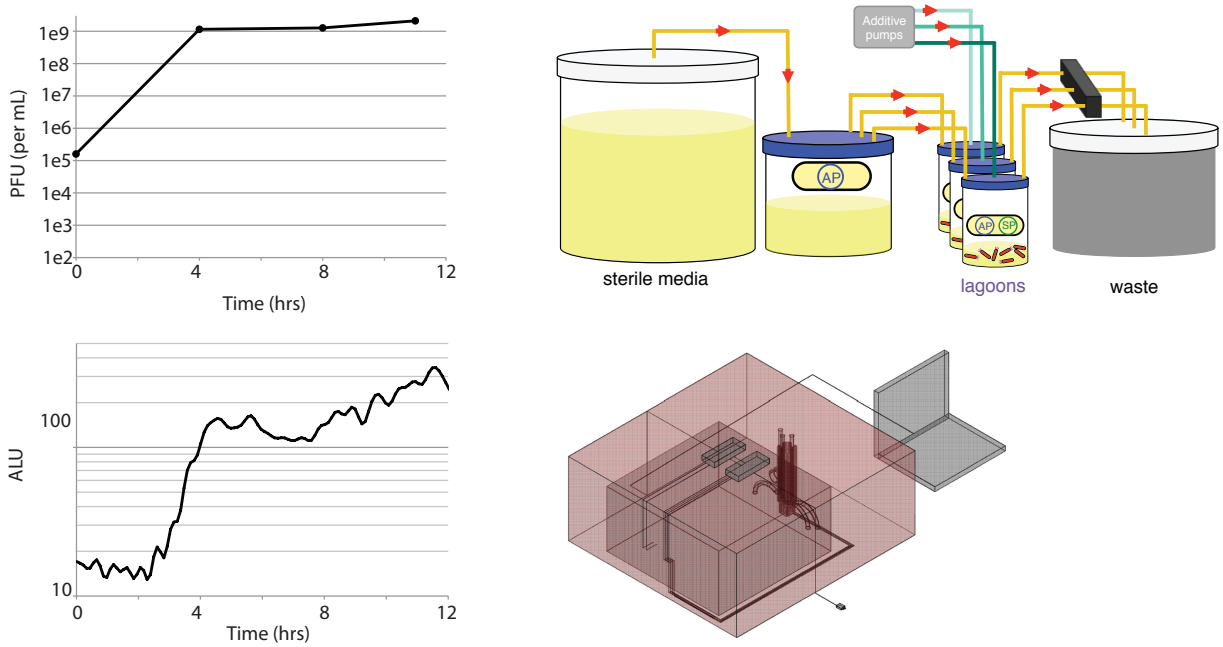


Figure 3.1. Demonstration of chemostat and real-time monitor with propagation of an SP-T7 phage. Upper left, Tracking of phage titer over time using chemostat grown cells. Upper right, Setup of chemostat. Lower left, Real-time luminescence monitoring of lagoon. Lower right, Setup of real-time monitor.

A real-time monitor of phage propagation was implemented by constructing a device for the in-line measurement of lagoon cell luminesence deriving from a coupled gIII-luciferase reporter. The details of construction are described in Materials and Methods. Luminescence signal was found to correlate strongly with the titer of a positive control phage (Figure 3.1)

3.2.2 Development of Small Molecule Controlled Drift Capability

To create a system for tunable selection stringency, we built upon existing strategies for small-molecule controlled gene expression. The TetR repressor of Tn10 represses promoters by binding to cognate operators at positions that occlude RNA polymerase or transcription factor binding[2]. This repression is relieved by addition of tetracycline and its more potent analogs (e.g. anhydrotetracycline, ATc)[3]. We constructed an AP with gIII downstream of a modified pTetA promoter (containing an additional TetR operator) and demonstrated ATc-dependent gene expression and phage production (Figure 3.2), but found that this AP did not support robust phage propagation in a PACE format when ATc was added to a lagoon (Figure 3.2). Phage were amplified very slowly, implying that the overall replication rate was near the limit of the dilution rate.

We hypothesized that the problem was related to the infectability of the cells. It is known that low levels of pIII-expression will result in retraction of F-pili (the primary receptor for phage infection), thus rendering cells resistant to infection[4]. We speculated that our recipient cells, which begin expressing pIII immediately upon entering an ATc induced lagoon, were becoming phage-resistant prior to an initial infection and hindering

propagation. In discrete tests, recipient cells carrying induced pTet-gIII do not support activity independent plaque formation with SP-T7 or wild type phage, and are not competent in an explicit infection assay (Figure 3.3). Though these discrete tests cannot

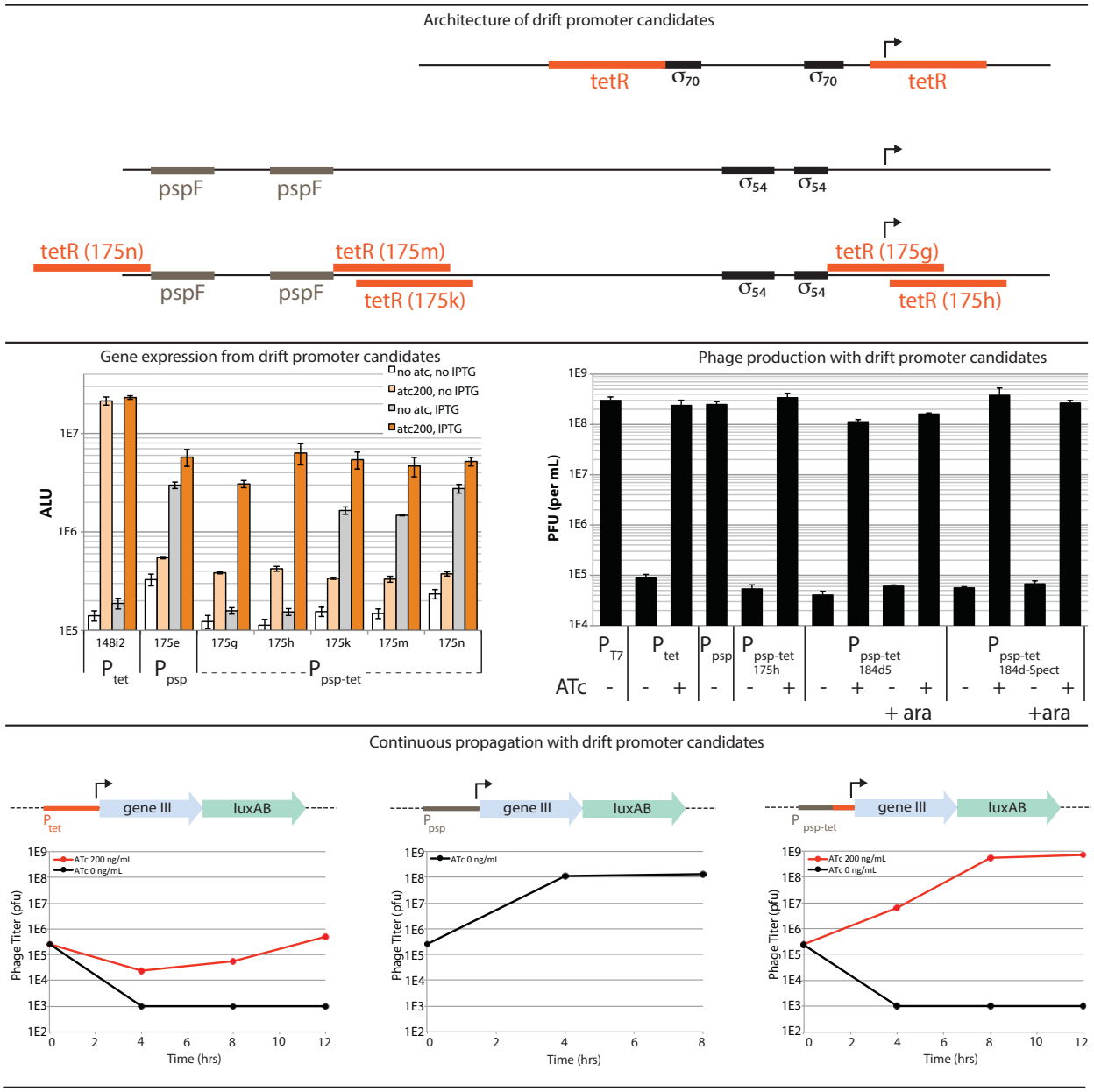


Figure 3.2. Top. Architecture of pTet, psp, and psp-tet candidates. Middle left. Luciferase gene expression measurements of candidate drift cassettes in the presence and absence of ATc and IPTG (which induces pIV expression from a separate pLac-pIV cassette). Middle right. Phage production from discrete cultures in the presence and absence of ATc. Bottom. Continuous propagation of phage with indicated drift cassettes.

capture the more dynamic induction that would occur with recipient cells entering a lagoon, it is likely that the same phenomenon is obstructive to continuous propagation. Robust drift might be achieved by including an explicit requirement for phage infection prior to induction.

To create a system in which activity-independent gIII expression requires both ATc *and* a prior phage infection, we took advantage of previous studies of the *E. coli* phage shock promoter (psp)[5]. This promoter is induced by a number of environmental signals, including infection by filamentous phage. Induction proceeds via expression of phage protein pIV, which inserts into the inner membrane, thereby initiating a signaling cascade that results in the binding of the transcriptional activator PspF to the phage shock promoter[6]. Thus, any phage that encodes pIV will induce this promoter in infected cells, and it has previously been shown that a phage deleted of gIII will form plaques on cells bearing a psp-gIII cassette[7]. We prepared cells with an AP carrying a psp-gIII cassette and observed pIV-dependent gene expression and robust phage production. Since psp has some basal expression of pIII, a mild defect with recipient cell infection is expected, and indeed observed in discrete assays (Figure 3.3). However, the defect does not seriously compromise activity-independent plaque formation (Figure 3.3) and continuous propagation (Figure 3.2).

We expected that induction from psp could be made ATc dependent by careful placement of TetR operators within critical regions of the promoter. Regulation of this promoter is known to be complex, so several variants were explored with TetR operators at candidate positions intended to disrupt either PspF or RNA polymerase binding (Figure

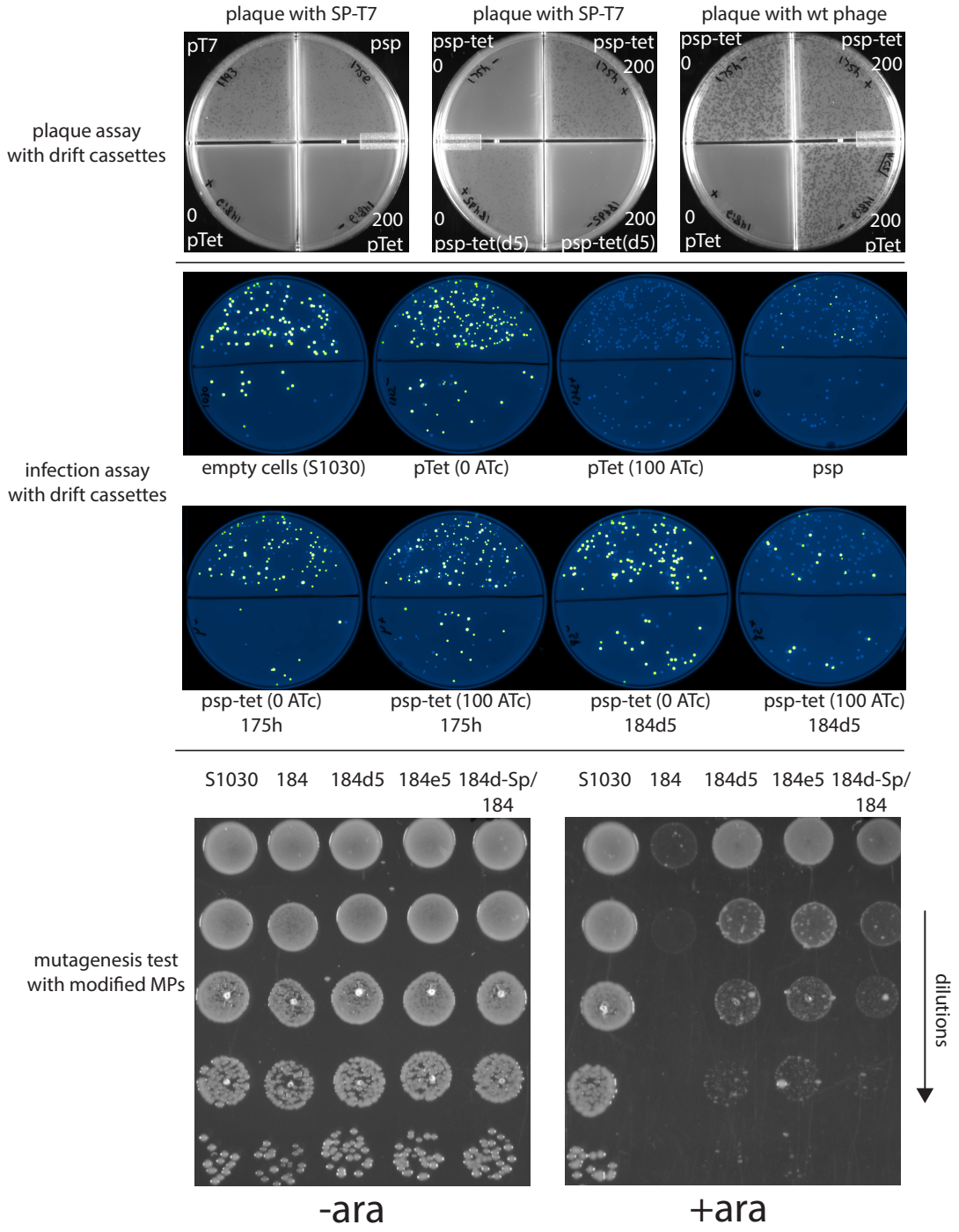


Figure 3.3. Top. Plaque assay with drift cassettes. Middle. Discrete infection assay using YFP-encoding phagemids. Bottom. Test of mutagenesis strength for drift-modified MPs as inferred by arabinose induced cell killing. pJC184 is the original MP, pJC184d5 combines mutagenesis and drift cassettes on a single plasmid, and pJC184d-Sp/184 harbors drift and mutagenesis cassettes on separate plasmids. All are compatible with simultaneous use of standard accessory plasmids.

3.2). We found that operators placed adjacent to binding sites for RNA polymerase, but not PspF, could effectively regulation psp induction. Placement of a TetR operator adjacent to the +1 transcription initiation site (in pJC175h) creates a promoter (herein referred to as psp-tet) that is maximally induced by the *combination* of ATc and pIV (as a surrogate for phage infection) and supports ATc dependent phage production (Figure 3.2). This drift cassette also performed well in discrete tests for plaque formation and infection competence (Figure 3.3). When an AP with this psp-tet-gIII cassette was tested in a PACE format, phage propagation was found to be robust and ATc dependent, thus making it possible to propagate a library with no starting activity (Figure 3-2).

3.2.3 The Stringency of a Selection Can Be Tuned With ATc

We next examined how the psp-tet-gIII cassette would influence the enrichment of active library members over inactive library members in the context of an actual PACE selection, in which an additional copy of gIII is controlled by an activity-dependent promoter. At intermediate concentrations of ATc, a phage encoding a new, active library member should enjoy a replicative advantage by inducing additional pIII expression from an activity-dependent promoter. This advantage should be inversely proportional to the concentration of ATc provided; as ATc induced pIII approaches saturating concentrations, the additional benefit of activity-induced pIII expression diminishes, as does the selection for active library members. Knowledge of this relationship is essential to maximize the speed with which an active library member can be generated *and* enriched.

To test the effect of psp-tet-gIII on activity-dependent enrichment, we moved this cassette onto a modified mutagenesis plasmid (pJC184d5) that can be used in conjunction with an activity dependent AP and verified ATc-dependent phage production (Figure 3.2), plaque formation, and discrete infection competence (Figure 3.3). We also examined the mutagenesis properties of this modified MP (as assayed by cell death upon arabinose induction) and found a weaker response suggestive of a reduced mutagenesis rate. The problem could be resolved by maintaining the drift and mutagenesis cassettes on two separate, compatible plasmids (pJC184d-Sp, pJC184, respectively), which preserved ATc-dependent phage production (Figure 3.2) and full induced mutagenesis (Figure 3.3), but the combined MP (pJC184d5) was used and found suitable for most experiments.

We set up lagoons with recipient cells containing both this modified MP and an AP with the T7 promoter driving gIII expression (Figure 3.4). Lagoons were seeded with a mixture of phage encoding an inactive T7 RNAP (SP-mut, active site mutation D812G)[8] and the WT T7 RNAP in a ratio of 100:1, induced with varying concentrations of ATc, and the phage concentrations were observed over time. As expected, without ATc, SP-T7 was rapidly enriched over the SP-mut. At an intermediate concentration of ATc (150 ng/mL), SP-T7 still enriched but at a slower rate. At the highest ATc concentration (400 ng/mL), enrichment of SP-T7 was not observed, suggesting that neutral selection had been achieved. When ATc was then withheld from this lagoon, SP-T7 rapidly enriched as expected. Taken together, these experiments show that psp-tet-gIII supports the propagation of inactive starting libraries while still allowing the selective enrichment of active mutants.

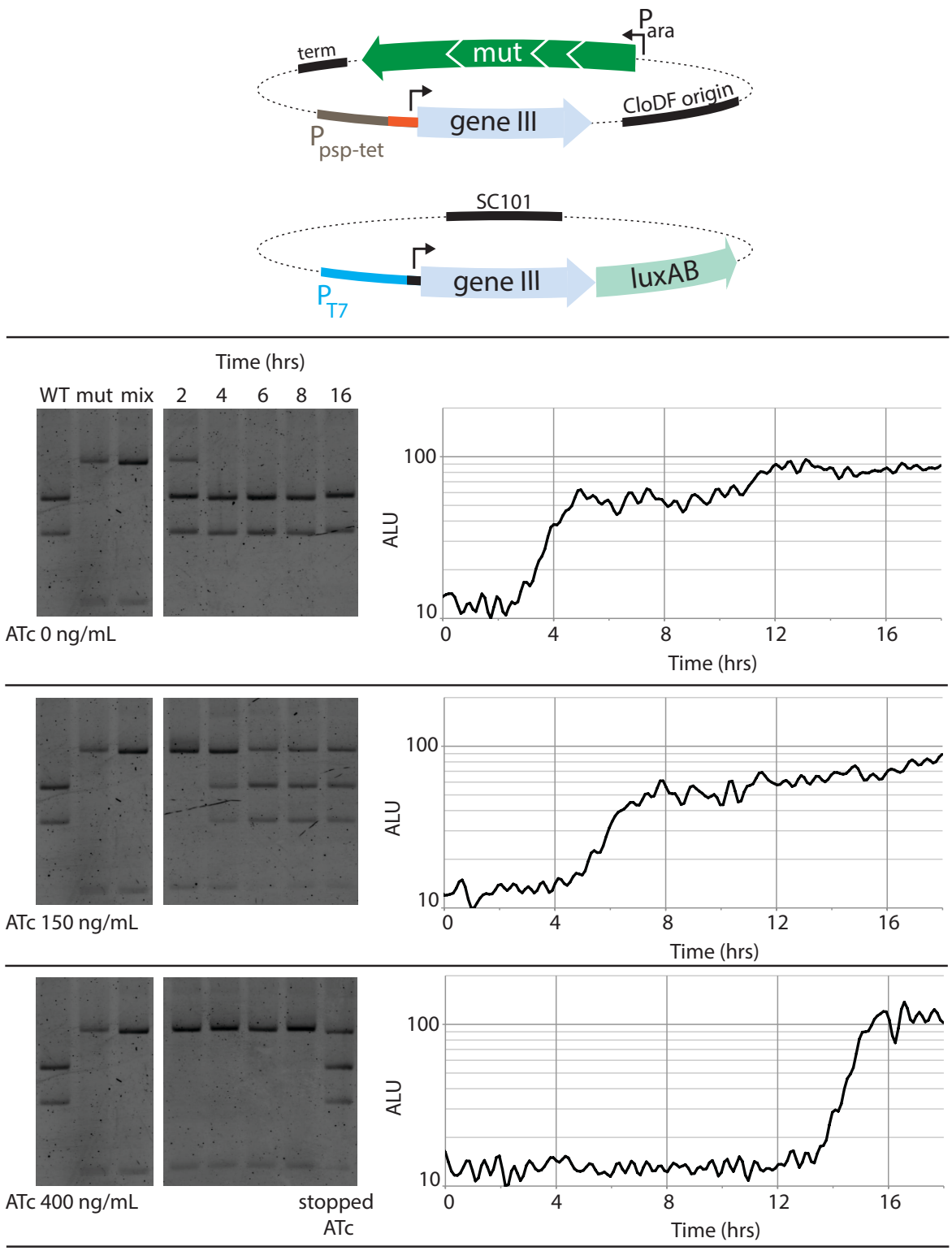


Figure 3.4. Top. Plasmids used in continuous drift experiment. Bottom left. Tracking of phage ratios over time by restriction digest. Bottom right. In-line luminescence monitoring of lagoons; high luminescence indicates an enrichment of the active SP-T7 phage.

3.3 Materials and Methods

3.3.1 General Materials and Methods

For all experiments in this chapter, the bacterial host strain is DH10B S1030. All continuous flow experiments are performed using recipient cells grown in a chemostat using the custom Davis rich media formulation. Discrete cultures are also grown in this same Davis rich media formulation, except where noted. All cultures were grown with antibiotic supplementation for the maintenance of exogenous plasmids (when a plasmid with the corresponding resistance gene was present) in the following concentrations: ampicillin (50 μg/mL), spectinomycin (50 μg/mL), chloramphenicol (40 μg/mL). Streptomycin (for selection of DH10B cells carrying the rpsL marker) and tetracycline (for selection of F plasmid) were not routinely included in culture media. Test strains were prepared by transformation of electrocompetent DH10B S1030 with appropriate plasmids, and plating on 2xYT agar supplemented with appropriate antibiotics.

For discrete experiments, 2 mL seed cultures were grown in 14 mL culture tubes. Cultures for gene expression measurements were typically grown as 0.5 mL volumes in 96-well deep well plates (Axygen, #391-01-061) seeded with a 0.2-1.0% inoculum of overnight culture.

3.3.2 Chemostat growth conditions and demonstration

Media for chemostat growth was prepared by autoclaving a 20 L carboy of MilliQ water containing: K_2HPO_4 anhydrous (140 g), KH_2PO_4 (40 g), $(NH_4)2SO_4$ (20 g), NaCl (58 g), casamino acids (100 g), Tween (20 mL), L-cysteine (1 g), L-tryptophan (0.5 g), adenine (0.5 g),

guanine (0.5 g), cytosine (0.5 g), uracil (0.5 g), CaCl2 (0.5 μM) final. This was allowed to cool, and amended with a 500 mL filter-sterilized solution of MilliQ water containing: $NaHCO_3$ (16.8 g), glucose (90 g), sodium citrate (10 g), $MgSO_4$ (1 g), $FeSO_4$ (56 mg), thiamine (134 mg), calcium pantothenate (94 mg), para-aminobenzoic acid (54 mg), para-hydroxybenzoic acid (54 mg), 2, 3 dihydroxybenzoic acid (62 mg), $(NH_4)_6Mo_7$ (3 nM final), H_3BO_3 (400 nM final), cobalt acetate (30 nM), $CuSO_4$ (10 nM final), $MnCl_2$ (80 nM final), $ZnSO_4$ (10 nM final).

For demonstration of chemostat based phage propagation, SP-T7 (1e7 pfu) was inoculated into a 60 mL lagoon (flow rate 2 vol/hr) that was receiving chemostat grown cells harboring an AP with pT7-gIII-luxAB. The chemostat dilution rate was ~1.6 vol/hr.

3.3.3 Construction and operation of real-time luminescence monitor

The in-line bioluminescence measurements were taken by a custom-built online luminometer. A TD-20e luminometer (Turner Designs) was installed inside a dark box in continuous reading mode. While in this mode, the luminometer outputs a DC voltage that varies between 0 and 4 V depending on the light level received by the photomultiplier fitted in the instrument (the range was set to the most sensitive one). This signal was sent to an Arduino Mega (Arduino) prototyping board through the analog input port. The Arduino board was controlled via Matlab (Mathworks), using a custom made Graphical User Interface (GUI). Measurements were taken at around 100 Hz and then integrated to 1 s before being written by the software into a text file.

To allow for multichannel recordings, the system was fitted with a custom-built tube holder that could hold up to four flow-through tubes with minimal cross talk between channels

(below noise level). The holder allows for only one tube to be observed at a time, and it is mounted on a stepper motor that is in turn controlled using the Matlab GUI via the Arduino board and a Pololu A4988 stepper motor driver Pololu R&E) . The system waits until a measurement is complete before rotating the sample holder until the next tube is in position. From there it repeats the cycle until all channels are covered and then it unwinds to start a new cycle so that the tubes do not entangle.

Matlab was the main tool used for data analysis. The purpose of the analysis is to extract the features of the luminescence fluctuation that are relevant to the timescale and purpose of the experiment, even when the time resolution of the instrument is much higher than that. The code loads the data from the text files and separates it in the three channels used in the experiments. Then the data is binned by a factor of 10 to get rid of most of the artifacts due to the fast measurements. Following this we transform the data into "timeseries" format and filter it with an ideal notch filter. This removes all the high frequency information that is of no relevance to the analysis. After filtering the data is again binned by a factor of 10 to render a very smooth trace that provides the low frequency features. Finally, the data is saved to an Excel spreadsheet that is used to generate plots.

3.3.4 Engineering of host strain

Starting from commercial DH10B *E. coli*, the Lambda Red method[9] was used to modify the genome as follows: 1) replace proBA with the pir-116 gene, enabling use of R6K origin plasmids and allowing proline auxotrophy to be used as a mechanism of F plasmid selection if needed, 2) delete lac IZYA, allowing use of LacI based mutagenesis screen, 3) create a scarless

mutation of rpoZ such that the start codon is mutated and a frameshift is introduced to the coding region. The scarless strategy was necessary to avoid a toxic polar effect with spoT. the mutation of rpoZ enables n-hybrid schemes that use a fusion to the N-terminus of the RNAP omega subunit (competition with endogenous omega subunit is problematic).

Starting from commercial ER2738, the Lambda Red method was used to modify the F plasmid as follows: 1) deletion of lacIZY, which allowed us of LacI based mutagenesis screens 2) integration of a cassette that overexpresses LacI and TetR, thus allowing the use of these inducible expression systems, 3) integration of a cassette that expresses luxCDE, which encode for enzymes in decanal biosynthesis. this modification enables luciferase measurements to be taken easily and continuously without the addition of exogenous substrate.

The modified F plasmid was mated into the modified DH10B strain to create S714. This strain was used in initial chemostat experiments, and severe clumping of cells and clogging of tubing was observed. The Lambda Red method was then used to delete: 1) the csg operon and 2) flu operon.

This strain performed better in chemostat experiments, and clumping was not observed until after several days. But, clumping became severe in this strain when any additional plasmid (apart from the AP) was included. pgaC was then disrupted, and biofilm formation was substantially abolished under any conditions. In addition to biofilm gene disruption, the AraE promoter was replaced with a constitutive promoter to enable titrable arabinose induction[10], and also possibly increase the inducible mutagenesis rate by increasing the kinetics of arabinose induction. Mating of this strain with the modified F plasmid yielded strain S1030.

3.3.5 Assay for infection-competence of recipient cells

Recipient cultures were grown to mid-log phase, and 100 μL of culture was mixed by pipetting with 2 μL of pJC126b phagemid prep (encodes R6K origin, f1 origin, pLac-YFP) totaling ~1e9 cfu. Reactions were incubated for 2 minutes and then stopped by transfer of 1 μL to 1 mL of fresh 2xYT media. Cells were pelleted, resuspended in 1 mL fresh 2xYT, and an aliquot was diluted 10-fold into fresh media. 20 uL was plated to 2xYT-agar plates with 1 mM IPTG and grown overnight. Colonies were scanned on a Typhoon laser imager (488 nm laser, 520/40 filter).

3.3.6 Discrete assay for phage production

Cultures with candidate drift plasmids contained: pT7 (pJC173b), pTet (pJC148i2), psp (pJC175e), psp-tet (pJC175g/h/k/m/n), psp-tet on MP (pJC184d5), psp-tet on colE1-spect plasmid (pJC184d-Sp). Cultures were grown to mid-log phase, infected with an excess of SP-T7, and cells were pelleted and washed to remove residual excess phage. Cells were re-inoculated into fresh media and grown to ~OD 0.8, and supernatants were saved for titering. ODs were measured to normalize titers later.

3.3.7 Mutagenesis assay

Arabinose induced cell killing is used as a surrogate for a functional mutagenesis plasmid. Cells were grown to mid-log phase, serial dilutions were prepared and spotted on 2xYT agar plates with or without 5 mM arabinose.

3.3.8 Test of pTet-p3 drift cassette in continuous flow experiment

A chemostat culture of DH10B S1030 carrying plasmid pJC148i2 was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, two lagoons (volume 40 mL, flow rate 100 mL/hr) were started and seeded with 10^7 pfu of SP-T7. Lagoons received supplements delivered by syringe pump (flow rate 1 mL/hr) consisting of either 20 μg/mL ATc (to make 200 ng/mL final concentration) or vehicle (H_2O). At 4 hr. intervals, 0.6 mL samples were taken from lagoons, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored in the freezer. Supernatants were later titered on cells of DH10B S1030 containing pJC173b (an AP carrying pT7-p3).

3.3.9 Test of psp-p3 drift cassette in continuous flow experiment

A chemostat culture of DH10B S1030 carrying plasmid pJC175e was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, a lagoon (volume 40 mL, flow rate 100 mL/hr) was started and seeded with 10^7 pfu of SP-T7. At 4 hour intervals, 0.6 mL samples were taken from the lagoon, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored in the freezer. Supernatants were later titered on cells of DH10B S1030 containing pJC173b (an AP carrying pT7-p3).

3.3.10 Test of psp-tet-p3 drift cassette in continuous flow experiment

A chemostat culture of DH10B S1030 carrying plasmid pJC175h was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, two lagoons (volume 40 mL, flow rate 100 mL/hr) were started and seeded with 10^7 pfu of SP-T7. Lagoons received supplements delivered by syringe pump (flow rate 1 mL/hr) consisting of either 20 μg/mL ATc (to make 200 ng/mL final concentration) or vehicle (H_2O). At 4 hour intervals, 0.6 mL samples were taken from the lagoon, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored at in the freezer. Supernatants were later titered on cells of DH10B S1030 containing pJC173b (an AP carrying pT7-p3).

3.3.11 Test of dose-dependent stringency tuning with drift cassette

A chemostat culture of DH10B S1030 carrying plasmids pJC184d5 and pJC173c was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, three lagoons (volume 40 mL, flow rate 100 mL/hr) were started, with the waste lines diverted toward the in-line luminescence monitor. Lagoons were seeded with a mixture containing 10^9 SP-T7-D812G and 10^7 SP-T7, along with ATc in an amount that instantly brings the 40 mL lagoon to the final concentration supplied by syringe. Lagoons received supplements delivered by syringe pump (flow rate 1 mL/hr) consisting of either 15 μg/mL ATc (to make 150 ng/mL final concentration), 40 μg/mL ATc (to make 400 ng/mL final

concentration), or vehicle (H_2O). At t = 8 hr, the lagoon receiving 40 μg/mL ATc was switched to receive vehicle. At timepoints t = 2, 4, 6, 8, 16 h, 0.6 mL samples were taken from the lagoon, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored in the freezer.

Supernatants were analyzed for relative phage ratios using a PCR and analytical digest. 1 μL of each supernatant was added to a 20 μL Q-PCR reaction (iQ SYBR Green Mastermix, Biorad) with primers 1480/JC1481 and amplified by PCR with SYBR Green fluorescence monitored following each amplification cycle. Samples were removed individually from the PCR block after crossing a pre-determined fluorescence threshold, and placed on ice. Samples were then quantitated again by SYBR Green fluorescence in a plate reader, and approximately equal amounts of DNA from the reactions (normalized based on the fluorescence readings) were added to restriction digest reactions. The digest reactions used enzymes EcoRV and HindIII in buffer NEB2 with BSA, prepared according to manufacturer instructions in 20 μL total volumes. Digestions were heat inactivated, combined with 0.2 volumes of 5X loading dye containing Orange G and xylene cyanol, and analyzed by agarose gel electrophoresis (1% agarose, 0.5X TBE, 120V, 80 min). The electrophoresed gel was stained with SYBR Gold (30 min, using manufacturer protocol, Invitrogen) and imaged on a Typhoon laser scanner (excitation 488, emission 520/40 nm).

3.4 References

1. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228**, 506-515 (1992).

2. Klock, G. & Hillen, W. Expression, purification and operator binding of the transposon Tn1721 encoded Tet repressor. *J. Mol. Biol.* **189**, 633-641 (1986).

3. Degenkolb, J., Takahashi, M., Ellestad, G. A. & Hillen, W. Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. *Antimicrob. Agents Chemother.* **35**, 1591-1595 (1991).

4. Boeke, J. D., Model, P. & Zinder, N. D. Effects of bacteriophage f1 gene III protein on the host cell membrane. *Mol. Gen. Genet.* **186**, 185-192 (1982).

5. Brissette, J. L., Russel, M., Weiner, L. & Model, P. Phage shock protein, a stress protein of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 862-866 (1990).

6. Jovanovic, G., Weiner, L. & Model, P. Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the Escherichia coli stress-induced psp operon. *J. Bacteriol.* **178**, 1936-1945 (1996).

7. Rakonjac, J., Jovanovic, G. & Model, P. Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3. *Gene* **198**, 99-103 (1997).

8. Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269**, 25120-25128 (1994).

9. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640-6645 (2000).

10. Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., & Keasling, J. D. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriology*, **182** (24), 7029–7034 (2000).

Chapter Four

A Method for Negative Selection in Continuous Evolution

Jacob Charles Carlson, Ahmed H. Badran, David R. Liu

Jacob Carlson designed and performed all experiments, with the excellent assistance of Ahmed Badran.

4.1 Introduction

In our first PACE evolution experiments, the T3-active polymerases still retain their original high activity on the T7 promoter; these evolved enzymes are substrate promiscuous. This result is not surprising, as we and others have observed that in the absence of explicit negative selection against undesired activity, the directed evolution of novel substrate recognition often results in enzymes with broadened, rather than truly altered specificity. We envision that many targets of PACE (for example, proteases, protein binders, recombinases, etc) will require exceptional substrate selectivity to operate, with orthogonality, in the context of complex environments with many potential off-target substrates. Motivated by this need for selectivity, we developed a general negative selection for PACE in which undesired activity (e.g. off-target substrate recognition) induces expression of pIII-neg, a competitive inhibitor of pIII-mediated infectious phage production. This scheme operates in parallel, and simultaneously, with pIII-based positive selections to effect the rapid enrichment of highly active and substrate-specific mutants. We demonstrate the power of this dual selection in a mock enrichment in which a substrate selective RNAP is selected out of a starting 1,000,000-fold excess of a substrate promiscuous RNAP.

4.2 Results and Discussion

4.2.1 Development of a PACE Compatible Negative Selection

A PACE compatible negative selection should link undesired or off-target activity to inhibition of infectious phage production, and would ideally have the following properties: *(i)* inhibition would be specific to phage production, and *not* influence host cell viability, *(ii)* the relationship between off-target activity and inhibition would be easily tuned, *(iii)* inhibition would be proportional to the *ratio* of on and off-target activity, rather that just the absolute level of off-target activity. These criteria led us to consider a negative selection in which undesired activity induces expression of a variant of pIII that antagonizes the wild type pIII induced by a positive selection. To this end, we searched for a mutant of pIII that inhibits phage production.

The role of pIII in filamentous phage production is complex; it is essential for both the beginning and end of the phage life cycle. The protein is comprised of three domains, N1, N2, and C that together mediate infection of a host cell through initial attachment to the F-pilus tip (N2 domain), subsequent docking with the TolA cell-surface receptor (N1 domain), and unlocking of the particle for genome entry (C domain)[1,2,3]. Following infection, new progeny are synthesized as the replicated phage genome is extruded out through the host cell membrane as a pVIII-coated ssDNA filament. As synthesis of a unit length particle nears completion, pIII, which resides anchored by its C-terminus to the periplasmic face of the inner membrane, is attached via it's C-domain to the end of the particle in a stoichiometry of five molecules of pIII per particle. A presumed conformational change in pIII then catalyzes excision from the inner

membrane and release of the particle from the host cell in a process thought to bear symmetry with the infection process described above[3]. A wealth of biochemical and genetic experiments have defined regions of each domain essential for these steps, and these studies have raised a set of corresponding mutant proteins deficient in each process. One mutant described in this literature, termed N1-N2-C83, was of particular interest[3].

N1-N2-C83 contains in-tact N1 and N2 domains but has an internal deletion of 70 residues within the C domain. The residual C-domain contains sufficient information to mediate attachment to (but not catalyze excision of) a phage particle. When co-expressed with a pIII variant comprised of only the intact C-domain (which *can* mediate particle release), phage particle *production* was found to be normal, but the *infectivity* of these particles was drastically reduced. The phage particles, which contained a mixture of N-C83 and intact C-domains, were deficient in a step of the infection process following initial cell surface attachment. The authors described a model in which infection was obstructed because a particular in *cis* relationship between N1, N2, and full length C of the same pIII molecule is required to mediate the unlocking of a particle. However, we considered an alternative explanation, in which N1-N2-C83 might obstruct the cascade of conformational changes otherwise transmitted through all the in-tact C domains. If the latter model were true, this variant should also function as a dominant negative of full length, wild type pIII.

To test N-C83 expression as a candidate negative selection scheme, we created a host cell line containing a gIII deleted phage (SP-T7), an AP with a pTet-gIII cassette (induced by ATc) and a second accessory plasmid (termed AP-neg) with a pLac-N1-N2-

C83 cassette (induced by IPTG). Discrete production cultures were grown at two concentrations of ATc, and with or without IPTG, and culture broths were assayed for infectious phage titer. The addition of IPTG, which induces N1-N2-C83 expression, dramatically reduced the titer of infectious phage. The fold change in infectious phage production was dependent on the concentration of ATc added. At 4 ng/mL ATc, a 20-fold decrease was observed, while at 20 ng/mL ATc, a 5-fold decrease was observed when IPTG was included (Figure 4.1). Alternative pIII-neg candidates, which included the full

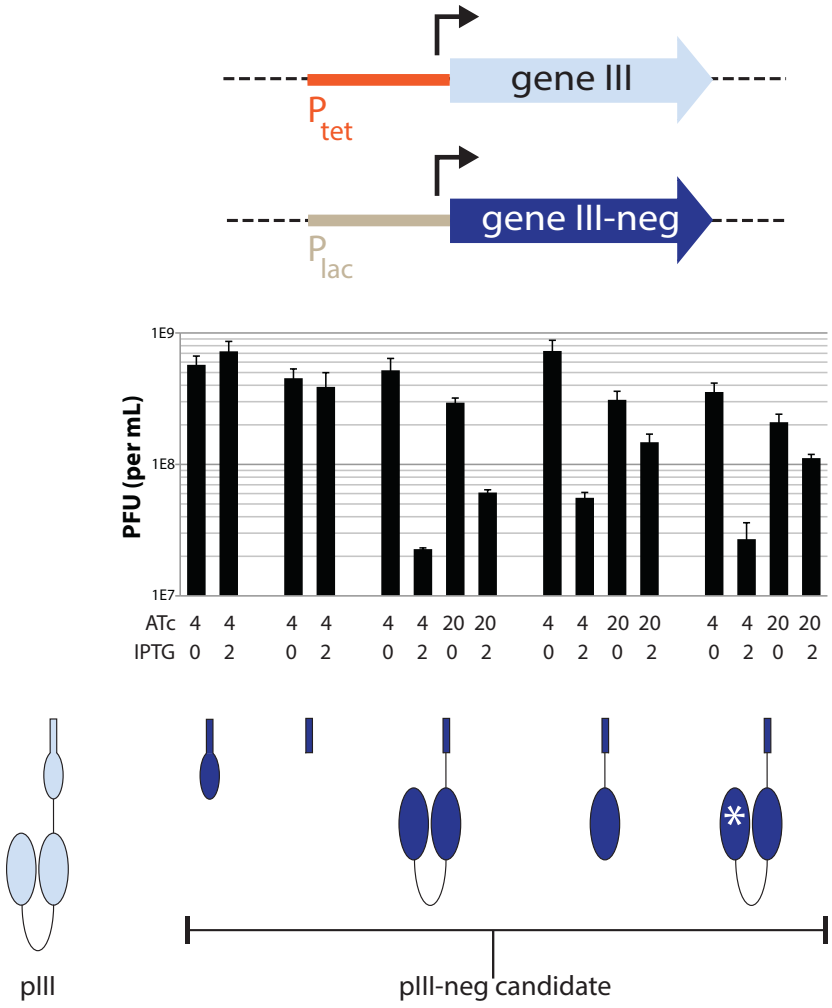


Figure 4.1. Effect on phage production of pIII-neg candidate expression. Constructs from left to right are 156a2 (C-domain), 156c2 (C83), 156j2 (N1-N2-C83), 156m2 (N2-C83), 156o2 (N1*-N2-C83)

C-domain, or C83 (which contains the 83 residues of the C-domain present in N1-N2-C83) had no effect on infectious phage production when expressed. Variants of N1-N2-C83 were also created with deletions of either the N1 domain or the YGT residues that causes F-pilus retraction and resistance to infection; these variants were found to be inhibitory to phage production, but the magnitude of the effect was reduced.

The dose dependence was verified by preparing recipient cells with pT7-gIII and pT3-ribo-N1-N2-C83 and observing phage production in the presence and absence of

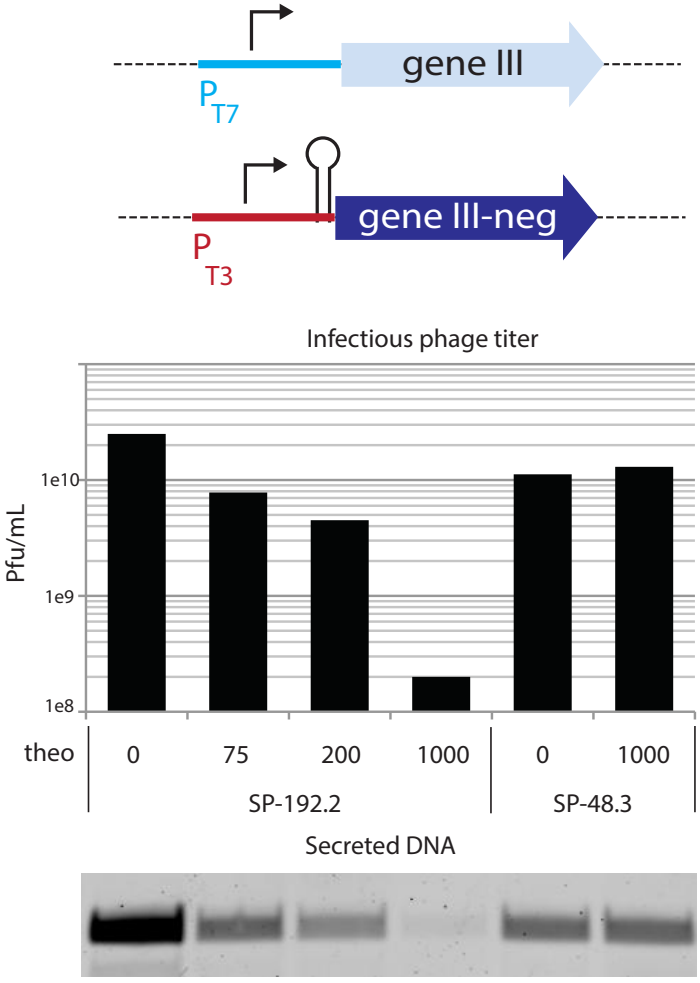


Figure 4.2 Dose dependent effect of pIII-neg expression on phage production of the promiscuous phage SP-192.2 or specific phage SP-48.3. SP-192.2 produces fewer progeny phage only when theophylline is added. The total secreted DNA was also measured and found to be reduced accordingly.

theophylline (Figure 4.2). Only promiscuous phage SP-192.2 is hindered by theophylline addition. Interestingly, the amount of secreted phage DNA in these cultures was also measured, and found to be reduced by an amount that explains the reduced infectious titer. This suggests that pIII-neg may inhibit infectious phage production by blocking the excision of extruding phage progeny, rather than influencing the infectivity of the resulting particles. This observation contradicts the original model. Nonetheless, these results demonstrate that expression of N1-N2-C83, herein referred to as pIII-neg, can form the basis of a very potent negative selection that is ratio-dependent with respect to pIII.

4.2.2 Demonstration of negative selection with continuous mock enrichment

Having determined that pIII-neg expression can reduce infectious phage production in discretely grown cultures, we next determined if, in a mock PACE experiment, this negative selection could be used to enrich for a substrate specific T7 RNAP over a promiscuous one. For this competition experiment, we cloned SPs encoding either of two RNAP variants from our previous study: L2-48.3 (specific for T7 promoter, SP-48.3) and L1-192.2 (promiscuous for T7 *and* T3 promoters, SP-192.2). We then prepared host cells containing an AP with the T7 promoter driving pIII expression, and an AP-neg with the T3 promoter driving pIII-neg transcription with pIII-neg translation controlled by a theophylline-activated riboswitch[4]. In these host cells, a phage encoding a promiscuous, T3-active RNAP should be penalized with pIII-neg expression, and *only* when theophylline is provided. Three lagoons containing these host cells were

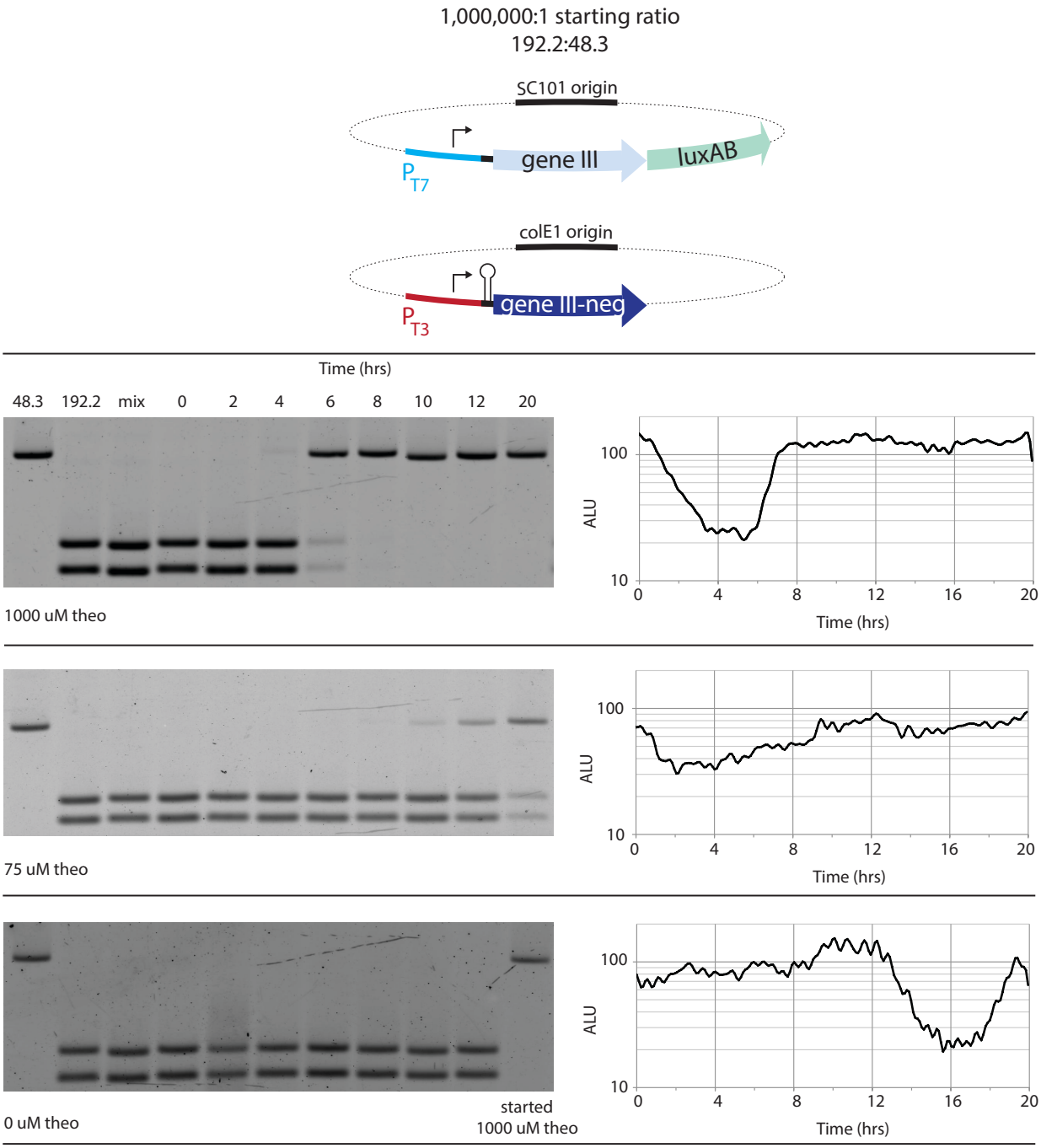


Figure 4.3. Mock enrichment of a pT7-selective RNAP (SP 48.3) over a pT7/pT3 promiscuous RNAP (SP-192.2). Recipient cells contained an AP with pT7-gene III-luxAB and an AP-neg with pT3-pIII- neg (top). A ratio of 1,000,000:1 SP-192.2:SP-48.3 was seeded into lagoons. At t = 0 hr, the indicated theophylline treatments were applied. Time points were taken and phage samples were analyzed by PCR and restriction digest to asses the relative abundance of the two phage species (left). Lagoons were also monitored with a real-time, in-line luminescence reader that reports population activity on the T7 promoter (right).

each seeded with a 10^6:1 ratio of SP-192.2:SP-48.3 and allowed to equilibrate for 4.5 hours. We then treated the lagoons with three different concentrations of theophylline (0, 75, 1000 uM) and followed the concentrations of the two SP species over time. We observed that, in the presence of theophylline, the promiscuous phage was rapidly depleted from the lagoon and replaced by SP-48.3 (Figure 4.3). This result demonstrates that p3-neg, serving as a potent and dose dependent negative selection, can effect the rapid enrichment of substrate selective RNAPs.

4.3 Materials and Methods

4.3.1 General Materials and Methods

For all experiments in this chapter, the bacterial host strain is DH10B S1030 and the chemostat media and growth conditions are those described in Chapter 3. Discrete cultures are grown in the Davis rich media formulation (Chapter 3), except where noted. All cultures were grown with antibiotic supplementation for the maintenance of exogenous plasmids (when a plasmid with the corresponding resistance gene was present) in the following concentrations: ampicillin (50 μg/mL), spectinomycin (50 μg/mL), chloramphenicol (40 μg/mL). Streptomycin (for selection of DH10B cells carrying the rpsL marker) and tetracycline (for selection of F plasmid) were not routinely included in culture media. Test strains were prepared by transformation of electrocompetent DH10B S1030 with appropriate plasmids, and plating on 2xYT agar supplemented with appropriate antibiotics.

For discrete experiments, 2 mL seed cultures were grown in 14 mL culture tubes. Cultures for gene expression measurements were typically grown as 0.5 mL volumes in 96-well deep well plates (Axygen, #391-01-061) seeded with a 0.2-1.0% inoculum of overnight culture.

4.3.2 Test of pIII-neg candidates

A host strain of DH10B S1030 containing pJC148i2 and either pJC156a2 (C-domain), pJC156c2 (C83), pJC156j2 (N1-N2-C83), pJC156m2 (N2-C83), or pJC156o2 (N1*-N2-C83) was prepared and grown to mid-log phase in Davis rich media. Cells were then infected with an excess of SP-T7, incubated in the warm room for 10 min, spun down and washed to remove residual phage, and resuspended in fresh 2xYT media. These infected cells were then used as inocolum for discrete 2xYT cultures grown in either ATc 4ng/mL or 20 ng/mL and in the presence or absence of 2 mM IPTG. Cultures were grown to OD ~0.9, cells were removed by centrifugation, and the supernatants were titered by plaque assay on DH10B S1030 cells carrying a pT7-AP.

4.3.3 Test of dose-dependent negative selection in discrete experiments

A host strain of DH10B S1030 containing pJC173c and pJC174c-R5 was prepared and grown to mid-log phase in Davis rich media. Cells were then infected with an excess of either SP-T7 192.2 or SP-T7 48.3, incubated in the warm room for 10 min, spun down and washed to remove residual phage, and resuspended in fresh Davis rich

media. These infected cells were then used as inocolum for discrete Davis rich media cultures grown with the indicated concentrations of theophylline.

4.3.4 Test of dose-dependent negative selection in continuous flow experiments

A chemostat culture of DH10B S1030 carrying plasmids pJC173c and pJC174c-R5 was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, three lagoons (volume 40 mL, flow rate 100 mL/hr) were started, with the waste lines diverted toward the in-line luminescence monitor. Lagoons were seeded with a mixture containing 10^{10} SP-192.2D and 10^4 SP-48.3 and allowed to equilibrate for 4.5 hrs; during this time the lagoons received 0.1 M NaOH at 0.5 mL/hr by syringe pump. After equilibration (t = 0 hr), lagoons received supplements delivered by syringe pump (flow rate 0.5 mL/hr) consisting of either 200 mM theophylline (to make 1 mM final concentration), 15 mM theophylline (to make 75 μM final concentration), or vehicle (0.1 M NaOH). After 16 hrs, the lagoon receiving vehicle was switched to receive 200 mM theophylline (1 mM final concentration). At timepoints t = 0, 2, 4, 6, 8, 10, 12, and 20 h, 0.6 mL samples were taken from the lagoon, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored in the freezer.

Supernatants were analyzed for relative phage ratios using a PCR and analytical digest. 1 μL of each supernatant was added to a 20 uL Q-PCR reaction (iQ SYBR Green Mastermix, Biorad) with primers JC1163/JC1488 and amplified by PCR with SYBR Green fluorescence monitored following each amplification cycle. Samples were

removed individually from the PCR block after crossing a pre-determined fluorescence threshold, and placed on ice. Samples were then quantitated again by SYBR Green fluorescence in a plate reader, and approximately equal amounts of DNA from the reactions (normalized based on the fluorescence readings) were added to restriction digest reactions. The digest reactions used AvaII in buffer NEB4, prepared according to manufacturer instructions in 20 μL total volumes. Digestions were combined with 0.2 volumes of 5X loading dye containing Orange G and xylene cyanol, and analyzed by agarose gel electrophoresis (1% agarose, 0.5X TBE, 120V, 80 min). The electrophoresed gel was stained with SYBR Gold (using manufacturer protocol, Invitrogen) and imaged on a Typhoon laser scanner (excitation 488, emission 520/40 nm).

4.4 References

1. Rakonjac, J. & Model, P. Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282**, 25-41 (1998).

2. Rakonjac, J., Feng, J. n. & 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.* **289**, 1253-1265 (1999).

3. Bennett, N. J. & Rakonjac, J. Unlocking of the filamentous bacteriophage virion during infection is mediated by the C domain of pIII. *J. Mol. Biol.* **356**, 266-273 (2006).

4. Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **37**, 184-192 (2009).

Chapter Five

Evolution of T7 RNA Polymerases with Specificity for the T3 Promoter

Jacob Charles Carlson, David R. Liu

Jacob Carlson designed and performed all experiments.

5.1 Introduction

To demonstrate the utility of the advances in Chapters 3 and 4, we integrated both developments by continuously evolving mutant T7 RNA polymerase enzymes with greatly altered substrate specificities in a single PACE experiment. The evolved enzymes exhibit selectivity for their target substrate that exceeds that of native phage RNA polymerases for their cognate substrates, while also maintaining wild-type-like levels of activity.

5.2 Results and Discussion

5.2.1 Evolution of T3-selective RNAPs from WT T7 RNAP

Our approach for evolving RNAPs with altered specificities is depicted in Figure 5.1. We divided the task into a series of stages designed to first acquire activity on the T3 promoter, and then gradually eliminate remaining activity on the T7 promoter. The expectation was that as we increased the stringency of the positive selection (by reducing pIII expression per unit activity), activity on the T3 promoter would increase. As we increased the stringency of the negative selection (increasing pIII-neg expression per unit activity), activity on the T7 promoter would decrease (Figure 5-1).

We began by preparing two lagoons with host cells containing a pT3-gIII AP and the modified MP encoding the psp-tet-gIII cassette, and seeded these lagoons with SP-T7, which has negligible starting activity on pT3. To one lagoon, we added no ATc and observed that phage quickly washed out, as expected. To the second lagoon, we added 200 ng/mL ATc, which was previously observed to allow inactive phage to propagate by

minimizing the selection stringency. After 12 hours (t = 12 hr), phage were present, but still encoded RNAPs with negligible activity on pT3. We then reduced the concentration of ATc to 20 ng/mL, thus increasing the selection stringency relative to the starting condition.

After six hours (t = 18 hr), the in-line luminescence monitor showed a clear increase in signal (Figure 5-1), suggesting that a T3-active population had overtaken the lagoon. This signal continued to increase, and RNAPs isolated from a later timepoint (t = 28 hrs) showed a ~100-fold increase in activity on pT3 and converged on two mutations, E222K and N748D, that have been observed by us and others to broaden the substrate scope of T7 RNAP (Figure 5-2)[1,2]. This result demonstrates that tunable selection stringency can be used to directly evolve a novel activity from inactive starting libraries without the use of intermediate substrates as evolutionary stepping stones.

As expected, the RNAPs recovered from this stage of the selection were promiscuous and retained high activity on pT7. To initiate the negative selection against recognition of pT7, we prepared host cells containing the same pT3-gIII AP, the MP (lacking the drift cassette), and a pT7-gIII-neg AP-neg with pIII-neg expression controlled by a theophylline dependent riboswitch[3]. The lagoon from the drift stage was then switched to receive these recipient cells and the lagoon was allowed to equilibrate for four hours, after which theophylline was added to trigger the counterselection (t = 32 hrs). After 24 hours at this intermediate stringency stage (t = 52 hrs), many of the RNAPs isolated had acquired a modest improvement in selectivity for pT3 (Figure 5.2) and these improved clones contained a variety of mutations proximal to the promoter

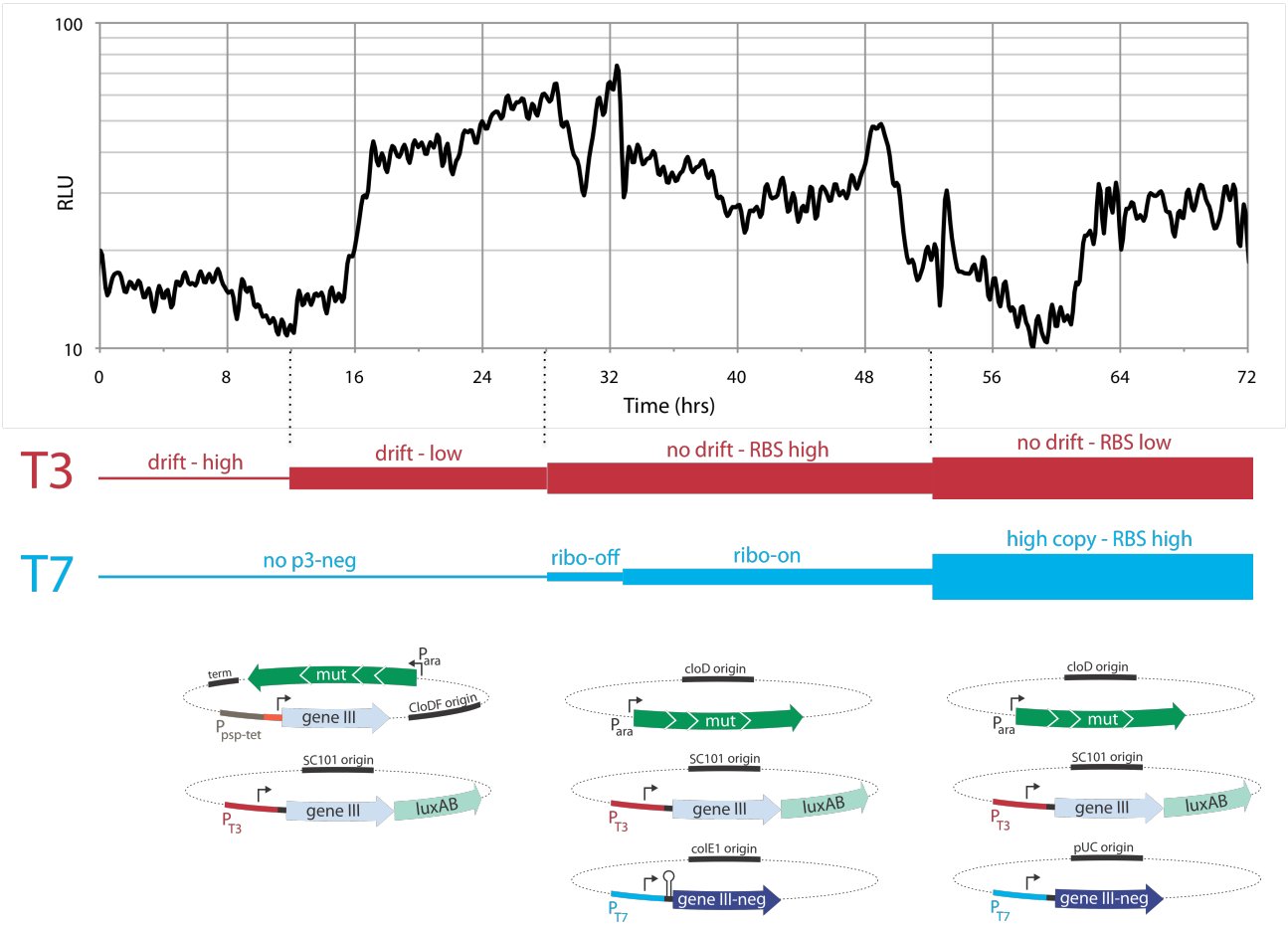


Figure 5.1. In-line luminescence monitoring (top), stringency schedule (middle), and constructs (bottom) used to evolve T3-selective enzymes starting from WT T7.

(R96L, K98F, E207K, P759L). In contrast, clones isolated from the no-theophylline lagoon did not show such changes in activity.

To further enhance the selectivity of our evolving RNAPs, we increased the stringency of the negative selection in two ways. First, we weakened the ribosome binding site driving translation of pIII from the AP, with the expectation that evolved RNAPs would compensate with increased (but pT3-selective) transcriptional activity. Second, we modified the AP-neg to contain a high copy pUC origin, and replaced the theophylline riboswitch with a very strong ribosome binding site driving pIII-neg

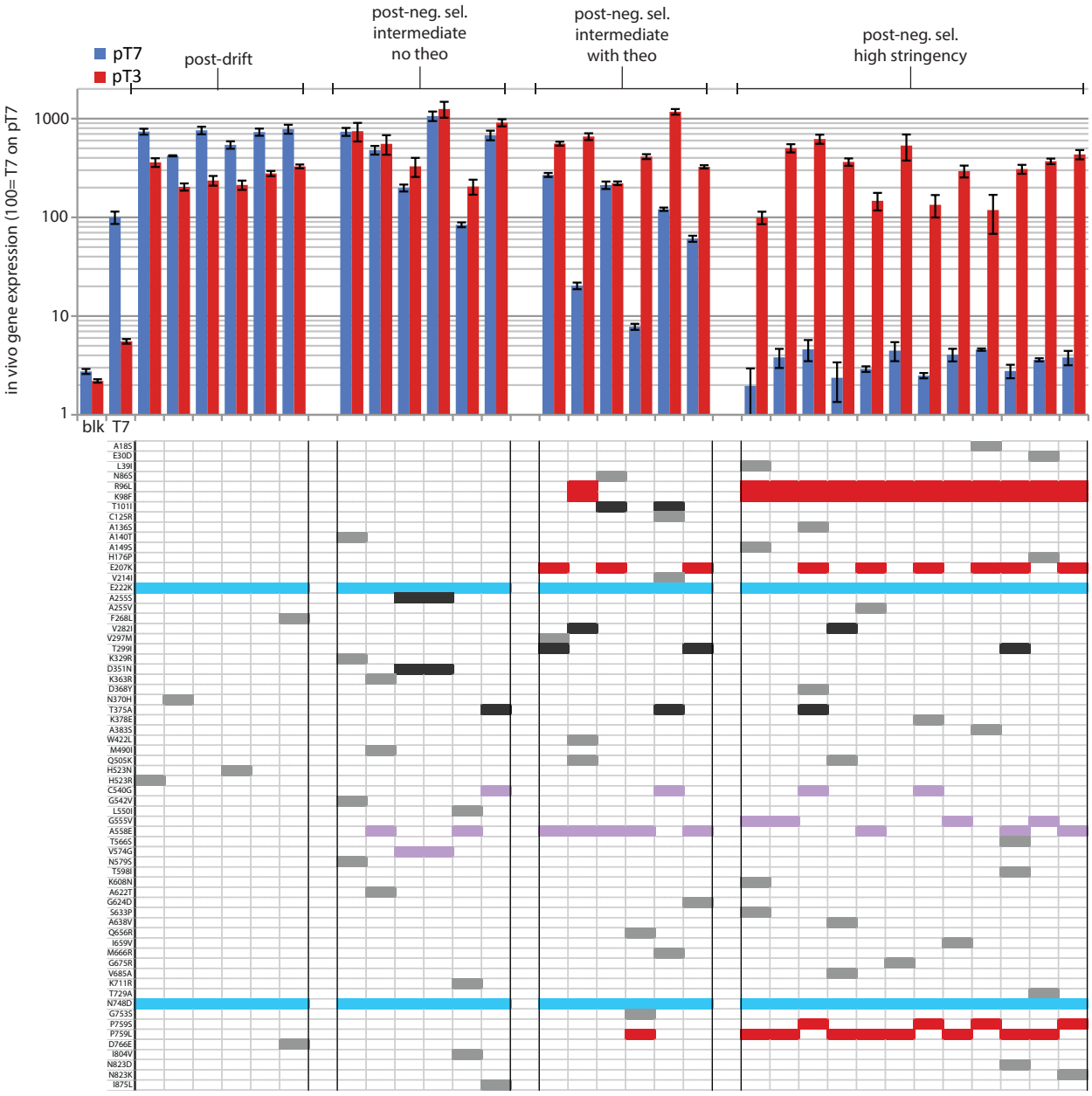


Figure 5.2. Activity and mutations present in clones isolated following the drift stage (left, t = 28 hrs), low stringency counterselection without theophylline (center left, t= 52 hrs), low stringency counterselection with theophylline (center right, t=52 hrs), and high stringency counterselection (right, t = 70.5 hrs). Mutations in blue are conserved in all clones and are found to confer activity on pT3. Mutations in red are conserved in high specificity clones. Mutations in purple are, as a group, conserved and physically clustered, but mutually exclusive of each other. Mutations in gray and black are isolated, and mildly conserved, respectively.

translation. Together, these modifications should increase the stringency of the counterselection by increasing the ratio of pIII-neg to pIII for a given ratio of activity on pT7:pT3. The phage population from the previous stage was then changed to receive these host cells, and the luminescence monitor indicated a dramatic decline and rebound of the population. RNAPs isolated following this recovery (t = 70 hrs) had dramatically improved specificity and had converged completely on promoter proximal mutations R96L, K98R, P759S/L (Figure 5.2).

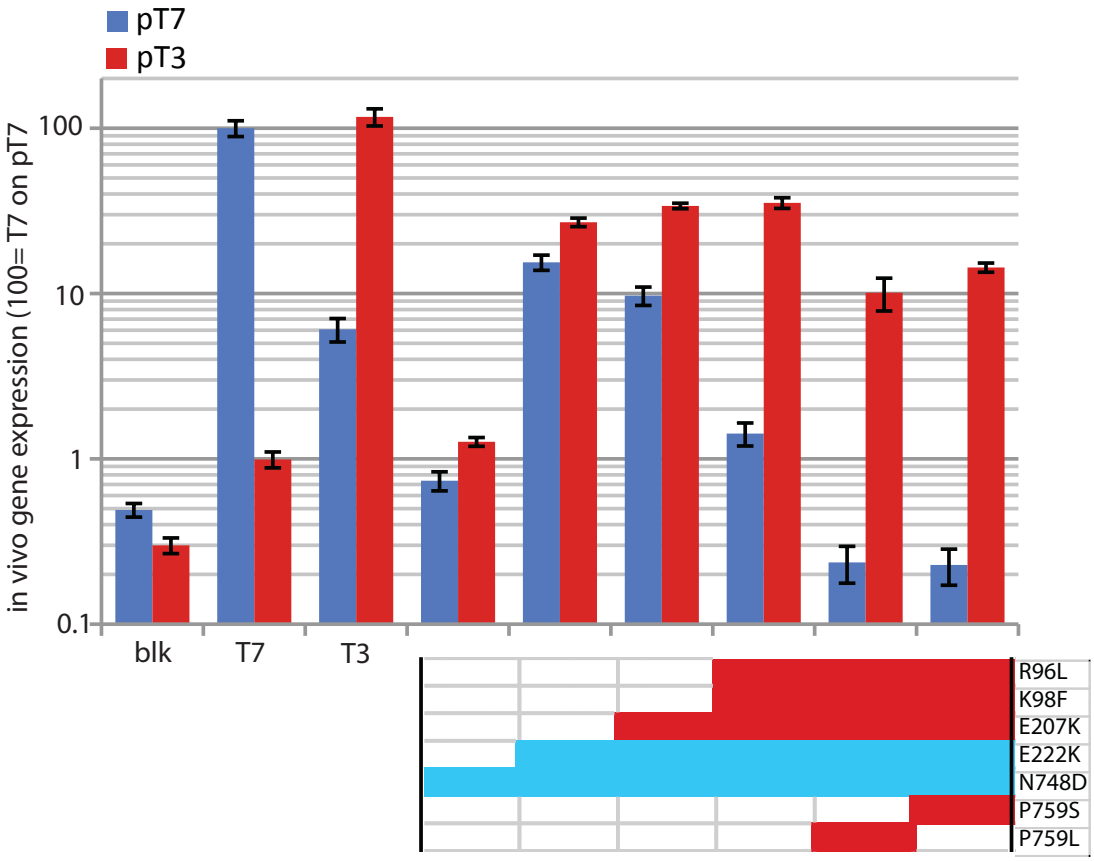


Figure 5.3. Forward mutation analysis of convergent changes. Changes in blue contribute to promiscuous activity on pT3, and changes in red contribute to selectivity for pT3 over pT7.

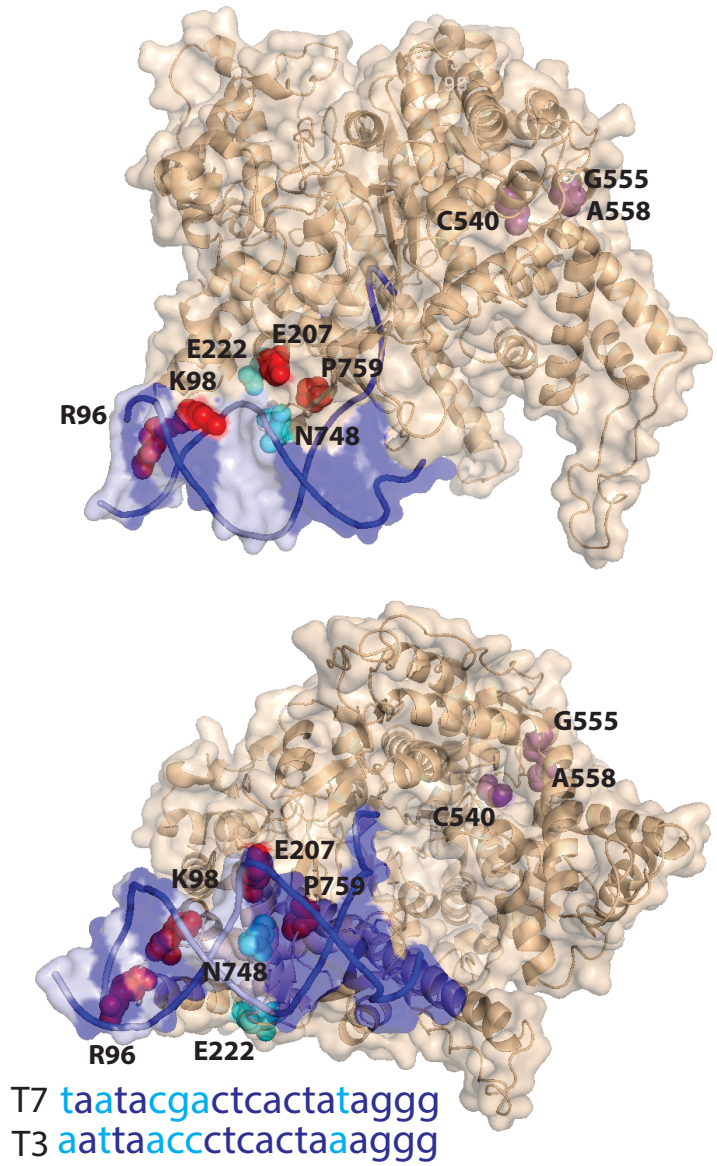


Figure 5.4. Convergent mutations in T7 RNAP-promoter complex[4]. Blue spheres are T7 promoter substrate, light blue spheres show changes between the T7 and T3 promoters. Cyan spheres are residues enabling pT3 recognition, red spheres contribute to *selective* pT3 recognition, purple spheres are a conserved cluster of exclusive mutations seen in final clones.

5.2.2 Dissection of specificity conferring mutations

A forward mutation study was performed (Figure 5.3) to examine the contribution of convergent mutations to the specificity of the final clones. The double mutation of E222K/N748D was found to be essential for acquiring activity on the T3 promoter, and

the R96L, K98R, and P759L mutations conferred high specificity toward T3 (and against T7).

5.3 Materials and Methods

5.3.1 General Materials and Methods

For all experiments in this chapter, the bacterial host strain is DH10B S1030 and the chemostat media and growth conditions are those described in Chapter 3. Discrete cultures are grown in the Davis rich media formulation (Chapter 3), except where noted. All cultures were grown with antibiotic supplementation for the maintenance of exogenous plasmids (when a plasmid with the corresponding resistance gene was present) in the following concentrations: ampicillin (50 μg/mL), spectinomycin (50 μg/mL), chloramphenicol (40 μg/mL). Streptomycin (for selection of DH10B cells carrying the rpsL marker) and tetracycline (for selection of F plasmid) were not routinely included in culture media. Test strains were prepared by transformation of electrocompetent DH10B S1030 with appropriate plasmids, and plating on 2xYT agar supplemented with appropriate antibiotics.

For discrete experiments, 2 mL seed cultures were grown in 14 mL culture tubes. Cultures for gene expression measurements were typically grown as 0.5 mL volumes in 96-well deep well plates (Axygen, #391-01-061) seeded with a 0.2-1.0% inoculum of overnight culture.

5.3.2 Evolution of T3 selective RNA polymerases

A chemostat culture of DH10B S1030 carrying plasmids pJC184d5 and pJC174f was prepared and growth was equilibrated overnight; the flow rate was 400 mL/hr in a chemostat volume of 250 mL. The next morning, two lagoons (volume 50 mL, flow rate 100 mL/hr) were started, with the waste lines diverted toward the in-line luminescence monitor. Lagoons were seeded with 10^5 pfu of SP-T7. Lagoons received supplements delivered by syringe pump (flow rate 1 mL/hr) consisting of either 20 μM ATc for Lagoon 1 or vehicle (500 mM L-arabinose) for Lagoon 2. At t = 12 hrs, Lagoon 1 was switched to receive 2 μM ATc. At t = 28 hrs, the first chemostat was discontinued and both lagoons began receiving cells from a second chemostat with DH10B cells containing plasmids pJC184, pJC174f, and pJC173f-R5 (prepared in the same way). Both lagoons received arabinose. At t = 32 hrs, half of Lagoon 1 was transferred to a new lagoon, Lagoon 3 and equilibrated for 40 min. Lagoons 1 and 2 then received 100 mM theophylline (dissolved in 0.1 M NaOH, flow rate 1 mL/hr) and Lagoon 3 received vehicle. At t = 52 hrs, the second chemostat was discontinued and all lagoons began receiving cells from a third chemostat with DH10B cells containing plasmids pJC184, pJC174k, and pJC173g-SD8 (prepared in the same way). All lagoon volumes were reduced to 40 mL (flow rate maintained at 100 mL/hr) and were supplemented with arabinose. At t = 70.5 hrs, lagoon volumes were reduced to 30 mL (flow rate maintained at 100 mL/hr). Periodically, 0.6 mL samples were taken from the lagoon, measured discretely for luminescence, and centrifuged to remove cells. Supernatants were combined with an equal volume of 50% glycerol and stored in the freezer.

5.3.3 Gene expression properties of evolved phage

Cells for luminescence assays were DH10B S1030 containing pJC184c-rrnB and either pJC173e4 (for T7 promoter activity) or pJC174e4 (for T3 promoter activity). Cells were grown to log phase and 20 μL aliquots were distributed into a deep-well plate. 10 μL of phage aliquots were added to these wells and plates were incubated in the warm room for 15 min. Wells were then supplemented with 500 μL media, grown to mid-log phase, and cells were transferred to a microplate for luminescence measurement.

5.3.4 Gene expression properties of forward mutants

Cells for luminescence assays were DH10B S1030 containing an expression plasmid (expressing a subcloned forward mutant of T7 RNAP) and either pJC173e4 (for T7 promoter activity) or pJC174e4 (for T3 promoter activity). Cultures were grown overnight and used to seed 500 uL cultures in 96-well blocks that were grown to mid-log phase, after which cells were transferred to a microplate for luminescence measurement.

5.4 References

1. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503 (2011).

2. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228**, 506-515 (1992).

3. Lynch, S. A. & Gallivan, J. P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **37**, 184-192 (2009).

4. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80-83 (1999).