The Dissertation Committee for Adam Joshua Meyer Certifies that this is the approved version of the following dissertation:

The Evolution and Engineering of T7 RNA Polymerase

Committee:

Andrew D. Ellington, Supervisor

Hal Alper

Jeffrey Barrick

James Bull

Walter Fast

The Evolution and Engineering of T7 RNA Polymerase

by

Adam Joshua Meyer, B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2014

Dedication

For my mother.

Acknowledgements

The work presented herein is the culmination of six years in the Ellington Lab. Over that time, I have engaged in countless discussions of varying degrees of appropriateness and scientific rigor. These conversations have shaped how I think as an investigator and who I am as a person. I owe so much to the members of the lab, past and present, and to Andy Ellington for fostering this unique corner of world.

Special thanks to Angel Syrett and Gwen Stovall for making me a buddy; to Eric Davidson and Matt Levy for laying solid groundwork; to Randall Hughes and Alex Miklos for keeping the lab functional in my early years; to Xi Chen and Peter Allen for loosening my notion of what is a crazy idea; to Michiko Kimoto and Laura Martini for making me feel lazy; to Michael Hammerling, Andre Maranhão, Jimmy Gollihar, Ross Thyer, Dan Garry, John Milligan, Bo Wang, and Drew Tack for carrying directed evolution into the future; to Tony Hwang and Steven Chirieleison for making the middle office ridiculous and to Christien Kluwe and Bingling Li for making it respectable; to Michelle Byrom, Arti Pothukuchy, and Paulina Dlugosz for making the trains run on time; and to Sarah Caton and Jorge Villafana for enabling all of us.

Thanks are especially due to Jared Ellefson, who has been my partner in science for the past five years. I would not have made it through this process without his friendship and collaboration.

Thanks to Fricano's Deli for fueling my efforts and to Crown and Anchor for hosting numerous symposia.

I thank my family, especially Jim, Karen, Eric and the Curls, and Dad for helping me through some rough years. Most importantly, I thank Allie, who loves me even when I am terrible, and Biscuits, whom I love even when he is terrible.

The Evolution and Engineering of T7 RNA Polymerase

Adam Joshua Meyer, Ph.D.

The University of Texas at Austin, 2014

Supervisor: Andrew D Ellington

T7 RNA polymerase is a single protein capable of driving transcription from a simple promoter in virtually any context. This has made it a powerful tool in a range of biotechnology applications. In this work, previous efforts to evolve or engineer T7 RNA polymerase are reviewed. This work is then expanded upon, first with the development of a method for the cell-free evolution of T7 RNA polymerase based on the functioning of an autogene. The autogene is a transcriptional feedback circuit in which active T7 RNA polymerase proteins transcribe their own gene, resulting in exponential amplification of their genetic information. While this system is doomed by an error catastrophe, this can be delayed by the use of *in vitro* compartmentalization. In response to the limits of the autogene, a novel directed evolution approach termed compartmentalized partnered replication (CPR) is presented. CPR couples the *in vivo* functionality of a gene to its subsequent *in vitro* amplification by emulsion PCR. The use of CPR to generate a panel of six versions of T7 RNA polymerase, each specific to one of six promoters, is described. Separately, a rational engineering approach, taken to facilitate the high-yield transcription of fully 2′-modified RNA, is detailed. Two sets of mutations to T7 RNA polymerase, previously known to confer thermal stability and enhance promoter clearance respectively, can be used to enhance the activity of existing T7 RNA polymerase mutants that utilize non-standard nucleotides as their substrates.

Next, CPR and random mutagenesis is used to populate the functional fitness landscape of T7 RNA polymerase. This neutral drift library is then challenged to increase the processivity of T7 RNA polymerase, enabling long-range transcription. Finally, the lessons that can be learned about T7 RNA polymerase specifically and molecular evolution and protein engineering generally are discussed.

Table of Contents

List of Tables xii

List of Figures xiii

Chapter 1: T7 RNA polymerase: biological context, use in biotechnology, and previous engineering and directed evolution efforts 1

T7 RNA polymerase biological context 1

T7 RNA polymerase in the bacteriophage life cycle 1

T7 RNA polymerase structure and mechanism 2

Wild type T7 RNA polymerase use in biotechnology 4

T7 RNA polymerase *in vivo* expression platform 4

In vitro transcription 6

In vivo and *in vitro* transcriptional circuitry 6

Previous T7 RNA polymerase engineering and evolution efforts 8

Promoter recognition 8

T7 promoter evolution 8

Rational T7 RNA polymerase and T7 promoter mutation studies 9

Expansion of T7 RNA polymerase promoter specificity with an autogene selection 9

Selection for altered T7 RNA polymerase promoter recognition with PACE 11

Rational engineering of promoter specificity mutants by phylogenetic part mining 13

Thermal stability 13

Promoter clearance 14

Substrate specificity 14

Rational mutations conferring expanded substrate specificity in T7 RNA polymerase 14

Directed evolution of expanded substrate specificity in T7 RNA polymerase 16

Conclusion 17

References.....17

Chapter 2: T7 RNA polymerase evolution using an *in vitro* autogene system.....25

Introduction.....25

Background and rationale.....25

Prior work toward implementation of the *in vitro* autogene.....27

Results and discussion.....29

In vitro autogene optimization.....29

In vitro autogene selection.....35

Conclusion.....38

Materials and methods.....40

Template preparation.....40

In vitro transcription and translation reactions in emulsion.....40

Template regeneration.....42

Cloning and sequencing.....42

T7 RNA polymerase activity assay.....43

Reverse transcription assay.....43

References.....44

Chapter 3: Directed evolution of a panel of orthogonal T7 RNA polymerase:promoter pairs using compartmentalized partnered replication 47

Introduction.....47

Background and rationale.....47

The need for orthogonal T7 RNA polymerase:promoter pairs...47

The limits of previous approaches.....48

Design of the compartmentalized partnered replication platform for directed evolution.....51

Results and discussion.....53

Implementation of the directed evolution of T7 RNA polymerase by CPR.....53

Enrichment for active T7 RNA polymerase mutants.....53

Library design.....57

Selection for the use of the wild type T7 promoter.....57

Selection for the use of a synthetic T7 promoter by CPR.....59
Selection for the use of P_{CGG}.....59
Maturation of CGG-R7-8.....62
Characterization of CGG-R12-KIRV 65
Comparison of CGG-R12-KIRV to other T7 RNA polymerase promoter specificity mutants.....66
Directed evolution of a panel of orthogonal T7 RNA polymerase:promoter pairs.....75
Selection for the use of P_{CTGA} 75
Selection for the use of P_{T3}, P_{K1F}, and P_{N4} 78
Assay of the most active mutant from each selection.....83
Cross-reactivity assays.....86
Assay of the most specific mutant from each selection.....91
Conclusion 96
Materials and methods 99
Library design and selection 99
In vivo GFP assay.....101
In vivo RNA polymerase cross-reactivity assays.....102
T7 RNA polymerase purification.....103
In vitro transcription assay.....103
References.....104
Chapter 4: Stabilization of T7 RNA polymerase substrate specificity mutants .107
Introduction.....107
Background and rationale 107
The need for RNA polymerases capable of transcribing RNAs with 2′-modified ribose moieties.....107
T7 RNA polymerase substrate specificity mutants.....108
Activity stability trade-offs in enzyme mutants.....110
Results and discussion 112
Analysis of the effects of stabilizing mutations on the activity of theT7 RNA polymerase substrate specificity mutant G542V, H784S112

Analysis of the effects of the M5 mutations on the activity of T7 RNA polymerase substrate specificity mutants....116
Analysis of the effects of the M5 and M6 mutations on the incorporation of 2′-*O*-methyl NTPs....120
Analysis of the ability of the T7 RNA polymerase RGVG-M6 to transcribe heavily-modified RNA....124
Conclusion....127
Materials and methods....128
Preparation of T7 RNA polymerase variants....128
In vitro transcription assays....129
References....131
Chapter 5: Lessons, ongoing work, and future directions....135
Lessons from the evolution and engineering of T7 RNA polymerase....135
Mutation rate and selective advantage....135
Activity and specificity....137
Assay conditions....138
Activity and stability....139
Ongoing work and future directions....140
A neutral network of T7 RNA polymerases....140
T7 RNA polymerase processivity....145
Thermal stability....152
Concluding remarks....153
Materials and methods....154
Compartmentalized partnered replication....154
Error-prone PCR....155
Gene shuffling....155
In vivo activity assay....156
References....156
References....160

List of Tables

Table 2-1. Mutational analysis from an *in vitro* autogene selection. 36

Table 2-2. Sequence of clones from Round 4 of an *in vitro* autogene selection.37

Table 3-1. Comparison of T7 RNA polymerase promoter specificity mutants.67

Table 3-2. Comparison of T7 RNA polymerase promoter specificity mutants.74

Table 3-3. A list of the most active T7 RNA polymerase promoter specificity mutants from each selection......86

Table 3-4. A list of the most specific T7 RNA polymerase promoter specificity mutants from each selection......95

Table 4-1. T7 RNA polymerase nucleotide specificity mutants......109

Table 4-2. Potential stabilization mutants of the T7 RNA polymerase nucleotide specificity mutant, VRS......114

Table 4-3. T7 RNA polymerase nucleotide specificity mutants with M5 mutations. 118

Table 5-1. Conditions used in neutral drift experiments. 143

Table 5-2. Conditions used in the selection for processivity. 146

List of Figures

Figure 1-1. The structure of T7 RNA polymerase. 2

Figure 1-2. T7 RNA polymerase interaction with the T7 promoter. 3

Figure 1-3. Antibiotic based selection. 8

Figure 1-4. Autogene based selection. 10

Figure 1-5. Phage-assisted continuous evolution. 12

Figure 1-6. T7 RNA polymerase interaction with ribose moieties. 16

Figure 2-1. *In vitro* autogene. 28

Figure 2-2. *In vitro* autogene based selection. 30

Figure 2-3. Failed propagation of the *in vitro* autogene selection. 31

Figure 2-4. DNA polymerase comparison for PCR amplification of autogene cDNA. 32

Figure 2-5. Replacement of the EMCV IRES. 34

Figure 2-6. Recoding the T7 RNA polymerase open reading frame. 35

Figure 2-7. Location of mutations found in Round 4 of an *in vitro* autogene selection. 37

Figure 3-1. Activity and specialization. 50

Figure 3-2. Compartmentalized partnered replication selection scheme. 52

Figure 3-3. CPR mock selection recovery. 54

Figure 3-4. CPR mock selection reamplification. 55

Figure 3-5. CPR mock selection digestion. 56

Figure 3-6. Sequence analysis of the T7 RNA polymerase specificity loop selection for the use of P_{T7}. 58

Figure 3-7. Activity assay of the T7 RNA polymerase specificity loop selection for the use of P_{T7}......59

Figure 3-8. Analysis of Round 7 of the T7 RNA polymerase specificity loop selection for the use of P_{CGG}......61

Figure 3-9. Sequence analysis of Round 12 of the T7 RNA polymerase specificity loop selection for the use of P_{CGG}......63

Figure 3-10. *In vivo* activity assay for the use of P_{CGG}......64

Figure 3-11. *In vivo* activity and cross-reactivity assay of CGG-R12-KIRV.65

Figure 3-12. *In vitro* activity and cross-reactivity assay of CGG-R12-KIRV.66

Figure 3-13. *In vivo* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants......68

Figure 3-14. Quantitative i*n vivo* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants......69

Figure 3-15. *In vivo* orthogonality of T7 RNA polymerase promoter specificity mutants......71

Figure 3-16. *In vitro* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants......73

Figure 3-17. Analysis of Round 7 of the T7 RNA polymerase specificity loop selection for the use of P_{CTGA}......76

Figure 3-18. Sequence analysis of Round 13 of the T7 RNA polymerase specificity loop selection for the use of P_{CTGA}......77

Figure 3-19. *In vivo* activity assay for the use of P_{CTGA}......78

Figure 3-20. Sequence analysis of Round 5 of the selection for the use of P_{T3}.80

Figure 3-21. Sequence analysis of Round 5 of the selection for the use of P_{K1F}.
.....81

Figure 3-22. Sequence analysis of Round 5 of the selection for the use of P_{N4}. 81

Figure 3-23. *In vivo* activity assay for the use of P_{T3}. 82

Figure 3-24. *In vivo* activity assay for the use of P_{K1F}. 82

Figure 3-25. *In vivo* activity assay for the use of P_{N4}. 83

Figure 3-26. *In vivo* activity and cross-reactivity assay of the most active T7 RNA polymerase mutants from each selection. 84

Figure 3-27. Quantitative *in vivo* activity and cross-reactivity assay of the most active T7 RNA polymerase mutants from each selection. 85

Figure 3-28. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{CGG} selection. 87

Figure 3-29. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{CTGA} selection. 88

Figure 3-30. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{T3} selection. 89

Figure 3-31. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{K1F} selection. 90

Figure 3-32. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{N4} selection. 91

Figure 3-33. *In vivo* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection. 92

Figure 3-34. Quantitative *in vivo* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection. 93

Figure 3-35. *In vitro* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection. 94

Figure 3-36. Selective pressure and the evolution of enzyme specificity. 98

Figure 4-1. Activty stability trade-offs in TEM1 β-lactamase. 111

Figure 4-2. The effects of the H772R mutation on the activity of the VRS mutant. 113

Figure 4-3. The effects of stabilizing mutations on the activity of the VRS mutant. 115

Figure 4-4. Structure of the transcribing "M5" RNA polymerase initiation complex. 117

Figure 4-5. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe canonical NTPs. 119

Figure 4-6. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe 2′-F-pyrimidines.120

Figure 4-7. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe 2′-*O*-methyluridine.121

Figure 4-8. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methyluridine......122

Figure 4-9. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methylpyrimidines......123

Figure 4-10. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methyladenosine and 2′-*O*-methylpyrimidines. 123

Figure 4-11. Transcription of various nucleotide compositions by RGVG-M6. 124

Figure 4-12. Transcription of 2′-*O*-methylnucleotides by RGVG-M6 in various buffers. 126

Figure 4-13. Transcription of 2′-*O*-methylnucleotides (mN) in a permissive buffer......127
Figure 5-1. Activity and specialization......137
Figure 5-2. Compartmentalized partnered replication selection scheme....142
Figure 5-3. 5′ UTR length and protein yield. 147
Figure 5-4. *In vivo* activity assay of Round 25 of the T7 RNA polymerase selection for processivity......151

Chapter 1: T7 RNA polymerase: biological context, use in biotechnology, and previous engineering and directed evolution efforts

The T7 phage-encoded RNA polymerase (T7 RNA polymerase) is a 98-kilodalton (883 amino acid) protein capable of recognizing a short 17-base pair promoter with high specificity and subsequently initiating transcription *de novo* (Chamberlin *et al*., 1970; Studier and Moffatt, 1986). In contrast to most known RNA polymerases, phage-encoded polymerases comprise a single subunit. The relative simplicity of T7 RNA polymerase has made it suitable as a model for the mechanism of transcription. It functions well in prokaryotic (Studier and Moffatt, 1986; Conrad *et al*., 1996; Gamer *et al*., 2009) and eukaryotic contexts (Fuerst *et al*., 1986; Elroy-Stein and Moss, 1990; McBride *et al*., 1994; Wirtz *et al*., 1998; Tokmakov *et al*., 2006; Davidson *et al*., 2012) and it does not appear to interact with host machinery or promoters. It requires no additional proteins or cofactors to function *in vitro* (Schenborn and Mierendorf, 1985; Milligan *et al*., 1987). These qualities have contributed to it being the workhorse of nucleic acid research and protein overexpression and the subject of numerous engineering efforts.

T7 RNA POLYMERASE BIOLOGICAL CONTEXT

T7 RNA polymerase in the bacteriophage life cycle

The T7 bacteriophage is a linear, double-stranded DNA phage about 40 kilobases in length enclosed in a polyhedral capsid (Krüger and Schroeder, 1981; Dunn and Studier, 1983). The tail adsorbs to the bacterial host and the genome is injected through the tail into the host. Shortly thereafter, gene 1 is expressed using host machinery to produce T7 RNA polymerase (Studier, 1972). T7 RNA polymerase then recognizes 17 distinct 17 base pair promoters, from which it initiates *de novo* transcription (Krüger and

Schroeder, 1981; Bull *et al*., 2007). This facilitates the expression of all middle and late phase genes, allowing phage propagation (Zhang and Studier, 1995).

T7 RNA polymerase structure and mechanism

T7 RNA polymerase is structurally related to family A DNA polymerase (*e.g.* the Klenow Fragment and HIV reverse transcriptase) (Zhang and Studier, 1995; Sousa, 1996; Kochetkov *et al*., 1998). Like other such nucleic acid polymerases, T7 RNA polymerase resembles a right hand with domains called the palm, thumb, and fingers as well as a large N-terminal domain (Sousa *et al*., 1993; Steitz *et al*., 1994; Jeruzalmi and Steitz, 1998; Cheetham *et al*., 1999). (**Figure 1-1**).

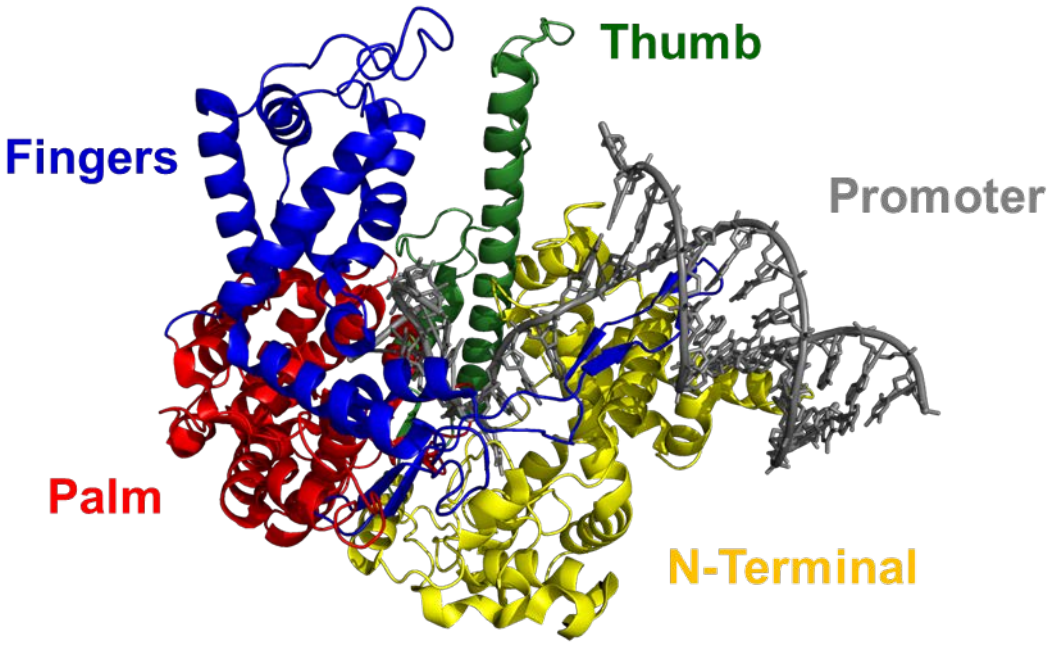


Figure 1-1. The structure of T7 RNA polymerase.
T7 RNA polymerase catalyzes the polymerization of RNA (gray) in an active site formed by the palm (red), fingers (blue), and thumb (green) domains. The N-terminal (yellow) and fingers (blue) domains interact with the promoter (gray). This figure is based on the crystal structure of T7 RNA polymerase in the initiation phase, PBD accession number 1QLN (Cheetham and Steitz, 1999).

The first step in transcription is the binding of the promoter and the formation of the closed complex (Kochetkov *et al*., 1998). This is accomplished by three distinct protein-DNA interactions (Cheetham and Steitz, 1999) (**Figure 1-2**). The AT-rich region from -17 to -13 of the promoter is contacted in the minor groove by residues 85 to 106 of the N-terminal domain. The -4 to +3 region of the promoter is contacted and the double helix is melted by residues 230 to 243 of the N-terminal domain. These regions are less important to the high degree of promoter specificity observed in T7 RNA polymerase. Promoter substitutions in these regions, especially A to T and T to A mutations, are well tolerated (Imburgio *et al*., 2000). The primary specificity-determining contacts are made by the "specificity loop" (residues 742 to 761 of the fingers domain), specifically by interactions made by residues R746, N748, R756, and Q758 with the major groove from -7 to -11. Mutations to this part of the promoter usually result in complete loss of promoter recognition (Imburgio *et al*., 2000). The specific molecular recognition of the T7 promoter by T7 RNA polymerase is discussed extensively in **Chapter 3**.

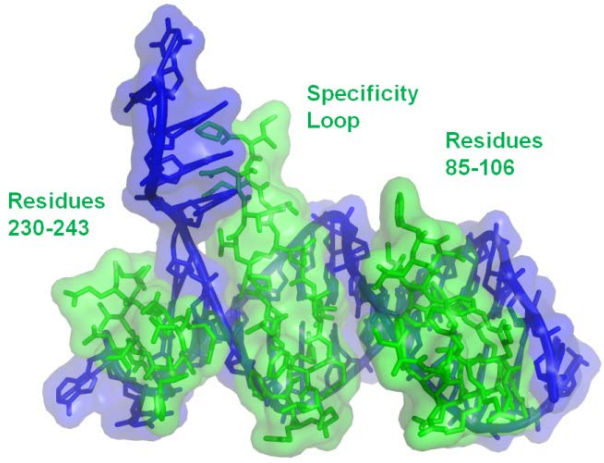


Figure 1-2. T7 RNA polymerase interaction with the T7 promoter.
3 loops of T7 RNA polymerase (green) make contacts with the T7 promoter (blue). This figure is based on the crystal structure of T7 RNA polymerase in the initiation phase, PBD accession number 1QLN (Cheetham and Steitz, 1999).

Upon promoter binding, an N-terminal β-hairpin inserts between the -4 base pair, and V237 disrupts hydrogen bonding and facilitates formation of the open promoter complex (Cheetham *et al*., 1999). The template strand is guided into the active site by the base of the specificity loop, and the +1 nucleotide is flipped out by W422, thus orienting it toward the incoming +1 GTP. Natural T7 transcripts are heavily biased toward a pppGpG start, likely because the open complex is most stabilized by the strong G:C base pairing and purine:purine base stacking in the nascent transcript (Kennedy *et al*., 2007). This *de novo* initiation phase of transcription is unstable, often leading to abortive cycling, which releases short transcripts fewer than 10 nucleotides in length. After this phase, there is a large-scale refolding of the N-terminal domain, eliminating contact with promoter, and forming a tunnel for the exit of the nascent transcript (Yin and Steitz, 2002). This moves the polymerase into the elongation phase in which the thumb and finger guide the downstream DNA into the active site and allow for processive transcription (Tahirov *et al*., 2002).

WILD TYPE T7 RNA POLYMERASE USE IN BIOTECHNOLOGY

T7 RNA polymerase *in vivo* expression platform

T7 RNA polymerase elongates five times faster than *E. coli* RNA polymerase (Studier and Moffatt, 1986), and is not impeded by host terminators (McAllister *et al*., 1981) or transcriptional polarity (Studier, 1972) allowing it to make predominantly full-length mRNAs. Upon induction, mRNA concentrations reach levels comparable to ribosomal RNAs, translation machinery becomes saturated, and the target protein can account for 50% of the total protein in only a few hours (Studier and Moffatt, 1986).

The T7 promoter is not recognized by the host's transcriptional machinery. This means that lethal genes under the control of a T7 promoter can be cloned, provided they

are oriented in the opposite direction of the *E. coli* promoters on the plasmid (Studier and Rosenberg, 1981). It was shown that these silent genes could be activated by the infection of T7 phage, which provides T7 RNA polymerase to the cell (McAllister *et al*., 1981). This was proposed as a possible method for controlled overexpression of cloned genes in *E. coli* (McAllister *et al*., 1981). However, this proved ineffective because transcription from the T7 promoters on the T7 phage genome diverted resources and led to lysis of the host (Studier and Moffatt, 1986).

This limitation was overcome by the cloning of T7 RNA polymerase into a heat inducible plasmid (Tabor and Richardson, 1985) and the development of lactose inducible systems (Studier and Moffatt, 1986). A mutant of the λ phage called DE3 was created, in which T7 RNA polymerase was placed under the control of the lacUV5 promoter. DE3 infection could provide T7 RNA polymerase for expression of target genes under the control of the T7 promoter. For convenience, and to avoid phage-based translation inhibition, a stable cell line was created, called BL21(DE3). BL21(DE3) contains a stable DE3 lysogen which provides lactose-inducible T7 RNA polymerase (Studier and Moffatt, 1986). Protein overexpression is now routinely performed with IPTG-induced BL21(DE3) cells and the companion pET expression plasmids.

T7 RNA polymerase is also the basis of many cell-free expression systems, including commercially available *E. coli*, wheat germ, or rabbit reticulocyte lystate. These systems have proven invaluable in expressing toxic proteins. T7 RNA polymerase-mediated cell-free transcription and translation of proteins from degenerate DNA libraries has enabled the *in vitro* evolution of binding species (Levy and Ellington, 2008) DNA modifying enzymes (Tawfik and Griffiths, 1998; Doi *et al*., 2004), and other enzymes (Griffiths and Tawfik, 2003; Mastrobattista et al., 2005; Lu et al., 2014).

In vitro transcription

The relative simplicity of phage RNA polymerases makes them ideal for the generation of an abundance of RNA not only *in vivo*, but also *in vitro*. Purified T7 or SP6 RNA polymerases were shown to transcribe RNA *in vitro* with digested plasmid containing their respective cognate promoters as the template (Melton and Krieg, 1984; Schenborn and Mierendorf, 1985). Importantly, no additional proteins or cofactors were needed for prolific transcription. The full power of *in vitro* transcription was then realized with the chemical synthesis of DNA oligonucleotides containing a T7 promoter followed by an arbitrary sequence. This enabled the production of milligram quantities of any short RNA sequence. The transcription of PCR products (Stoflet *et al*., 1988) or cDNA (Kwoh *et al*., 1989; Guatelli *et al*., 1990) only expanded the utility of T7 RNA polymerase-based *in vitro* transcription. The use of degenerate DNA templates and *in vitro* transcription by T7 RNA polymerase has served as the foundation of the *in vitro* selection of aptamers (Tuerk and Gold, 1990; Ellington and Szostak, 1990), ribozyme ligases (Bartel and Szostak, 1993; Levy *et al*., 2005; Wochner *et al*., 2011), and other ribozymes (Wilson and Szostak, 1995; Agresti *et al*., 2005; Lau and Ferré-D'Amaré, 2013).

In vivo and *in vitro* transcriptional circuitry

The high degree of activity, reliability, and specificity of T7 RNA polymerase has made it not only the core of directed evolution experiments, but also of increasingly complex transcriptional circuitry. It has allowed robust and predictable gene expression for metabolic engineering and RNA production for nucleic acid computation.

The Voigt Lab took advantage of the orthogonality and tight regulation of the T7 RNA polymerase expression system to learn about the native regulation of a gene cluster. Specifically, they rebuilt the *Klebsiella oxytoca* nitrogen fixation gene cluster from the

bottom up such that each gene was driven from a T7 promoter (Temme *et al*., 2012b). This removed any known or as yet undiscovered gene regulation from cryptic binding sites and promoters, pause sites, small RNAs, or other regulatory mechanisms. The 20 genes were arranged into four operons, each with a different strength T7 promoter (as noted above, mutations in the -4 to +4 region of the promoter can reduce activity without ablating it). Amazingly, the synthetic cluster retained about 8% of the activity of the native cluster.

The Church Lab took advantage of the predictable, specific, and robust gene expression characteristics of T7 RNA polymerase in their metabolic engineering efforts. They used multiplex automated genome engineering (MAGE) to deliver T7 promoters upstream of up to 12 genes in the indigo biosynthetic pathway (Wang *et al*., 2012). Upon addition of a plasmid expressing T7 RNA polymerase, indigo production was increased two to three-fold over strains with no added promoters. This approach allowed them to quickly optimize indigo production, and the semi-random insertion allowed them to uncover important epistatic and synergistic relationships in the metabolic pathway. It should be noted that this process could be further optimized by introducing degeneracy into the T7 promoters, thus allowing the relative gene dosage to be tuned.

The Winfree Lab has made extensive use of T7 RNA polymerase, using it as the core of *in vitro* transcriptional circuitry. They have devised several systems based on the notion that a RNA transcript may inhibit or activate the formation of an active T7 promoter. For example, the RNA output from one transcriptional unit could inhibit a second unit, whose product could activate the first unit. This results in oscillatory behavior (Kim and Winfree, 2011). Similar approaches have yielded bistable switches (Kim *et al*., 2006), as well as inverter or repeater circuits (Subsoontorn *et al*., 2012).

PREVIOUS T7 RNA POLYMERASE ENGINEERING AND EVOLUTION EFFORTS

The above demonstrations illustrate the power and versatility of T7 RNA polymerase, and, recognizing this, several groups have endeavored to evolve or engineer T7 RNA polymerase to enhance or alter its properties. Below are some of the most prominent of such efforts.

Promoter recognition

T7 promoter evolution

The directed evolution of the T7 RNA polymerase:promoter system started with a probing of the sequence space of the promoter. *E. coli* cells were co-transformed with two plasmids. The first plasmid contained T7 RNA polymerase and the second contained a mutagenized library of the T7 promoter controlling the expression of chloramphenicol acetyltransferase (Ikeda *et al*., 1992) (**Figure 1-3**). Any bacteria growing on chloramphenicol plates, therefore, harbor T7 promoters that can be recognized by T7 RNA polymerase. This work uncovered several promoters of varying strengths and revealed the importance of some residues that could never vary.

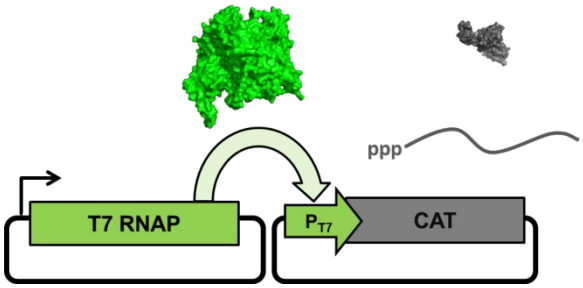


Figure 1-3. Antibiotic based selection.
T7 RNA polymerase (T7 RNAP) and its promoter (P_{T7}) must function together in order to drive chloramphenicol acetyltransferase (CAT), allowing its host to survive in the presence of chloramphenicol.

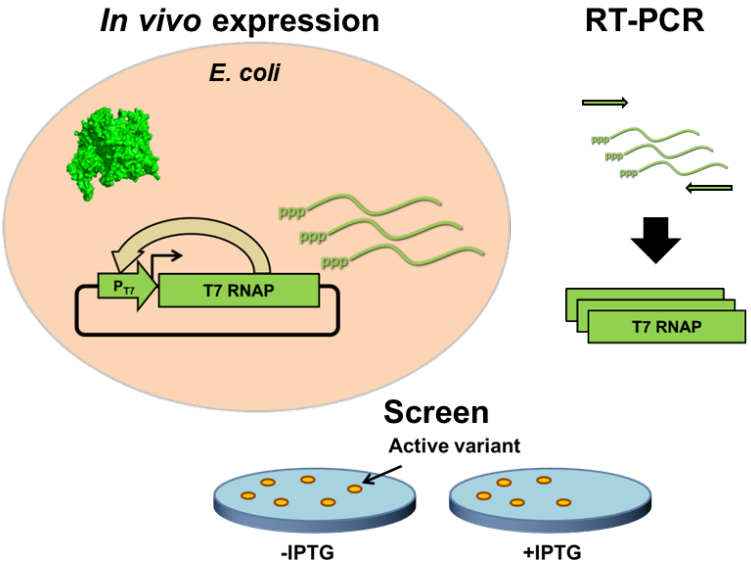
Rational T7 RNA polymerase and T7 promoter mutation studies

These efforts were followed by a suite of studies, in which defined mutations were made to the promoter, the polymerase, or both, and the strength of the resulting pair was assayed *in vitro*. It was shown that the promoter specificity of T7 RNA polymerases could be changed with a single mutation. While the closely related T3 and T7 RNA polymerases only weakly recognize one another's promoters, single amino acid mutations can switch their respective preferences. That is, an N748D mutation in T7 RNA polymerase enables it to transcribe from the T3 promoter (Raskin *et al*., 1992). This was followed by an extensive analysis, in which T7 RNA polymerase and the 19 possible mutants at the N748 position were tested for their ability to transcribe from the T7 promoter or one of the nine possible mutants that differ from the T7 promoter by a single base pair at either the -12, -11, or -10 position. This analysis highlighted the importance of the contact between N748 and the -11 and -10 position, and generated several somewhat orthogonal polymerase:promoter pairs (Raskin *et al*., 1993). Similar work was done with Q758 and the -8 position (Rong *et al*., 1998) and R756 with the -11, -10, and -9 positions (Imburgio *et al*., 2000).

Expansion of T7 RNA polymerase promoter specificity with an autogene selection

The Ellington Lab looked to evolve not only the promoter, but the RNA polymerase itself (Chelliserrykattil *et al*., 2001). In this directed evolution platform the T7 promoter drives the expression of the T7 RNA polymerase gene itself, a so called autogene (**Figure 1-4**) (Dubendorff and Studier, 1991). In the autogene circuit, an active polymerase:promoter pair would form a positive-feedback loop in which the concentration of T7 RNA polymerase mRNA and protein would increase exponentially. *E. coli* cells are transformed with a promoter or polymerase library, and, after a growth period, mRNA is collected and amplified by RT-PCR. Since active autogenes create

thousands of copies of their mRNA, the recovered RT-PCR product is enriched for active autogenes. Several cycles of cloning, electroporation, and RT-PCR yields a population of active autogenes. A final screening step is then performed, in which autogene clones are replica-plated onto a fully-inducing media. Some colonies are unable to grow on inducing plates, presumably because they are burdened by the metabolic load associated with a fully-induced, active autogene. These colonies are picked from the original (non-inducing) plate and sequenced.



T7 RNA polymerase mutants	Mutant T7 promoters recognized (-11 to -8)
Q758C	GACT (WT), GTTA, GTCA, GATA, GACG, TGTA
R3-17 (N748A, R756M, Q758S)	GACT (WT), GGTA, TATA, TGTA

Figure 1-4. Autogene based selection.
T7 RNA polymerase (T7 RNAP) is driven by its own promoter (P_{T7}), resulting in a positive feedback loop yielding high concentrations of mRNA. Active variants were screened for their ability to inhibit host growth when fully induced. A selection in which a library of T7 RNA polymerase mutants had to use a different promoter each round of selection yielded two generalist polymerases, each capable of using multiple promoters (Chelliserrykattil *et al*., 2001).

In one such autogene selection, the critical -11 position was mutated from G to C (similar to the related T3 promoter), and the amino acid residues R746, L747, and N748

were randomized. Sequencing from Round 3 showed a convergent population in which R746 and L747 reverted to wild type, while the N748D mutation fixed in the population. The T7 RNA polymerase N748D mutant had been previously shown to recognize T3-like promoters, thus illustrating the power of this method (Chelliserrykattil *et al*., 2001).

This initial success was followed with a more ambitious selection, in which the polymerase and promoter were both randomized. Specifically, -8 to -11 of the promoter was randomized, as were N748, R756, and Q758. As above, active autogene mRNA was isolated and re-cloned. Since mRNA does not contain the promoter sequence, a given polymerase was cloned under the control of a different promoter each round of selection. The dominant polymerase that emerged, Q758C, was a generalist. It was found paired with GACT (wild type) promoter as well as GTTA, GTCA, GATA, GACG, and TGTA promoters. Another variant, "R3-17" (N748A, R756M, Q758S) was paired with the GGTA promoter, but could also utilize GACT (wild type), TATA, and TGTA. Clearly, this autogene selection is an effective strategy for the evolution of promoter specificity generalist mutants of T7 RNA polymerase (Chelliserrykattil *et al*., 2001).

Selection for altered T7 RNA polymerase promoter recognition with PACE

The Liu Lab has recently developed a novel platform for directed evolution, called phage-assisted continuous evolution (PACE) (Esvelt *et al*., 2011). PACE ties a gene's function to the propagation of bacteriophage M13 (**Figure 1-5**). M13 is both deprived of its natural gene III, and also provided with a library of another gene. This library must execute some function in order to drive the expression of gene III from a plasmid carried by the host. The gene III product must be expressed in order to create infectious progeny phage particles. Fresh media and hosts containing the gene III

plasmid are continually added and a fraction of culture and hosts are continually eliminated, thus M13 must propagate in order to remain in the population.

The pilot experiment placed gene III on a plasmid under the control of a T7 promoter variant, and the T7 RNA polymerase gene in the M13 genome. Passage in this system enriched for M13 phages containing T7 RNA polymerase mutants capable of recognizing this synthetic promoter. This system has produced T7 RNA polymerase mutants capable of recognizing a T3-like promoter (Esvelt *et al*., 2011), a more divergent promoter resembling a hybrid of the T3 and SP6 promoters (Dickinson *et al*., 2013), and more recently has been tuned to increase stringency (Carlson *et al*., 2014).

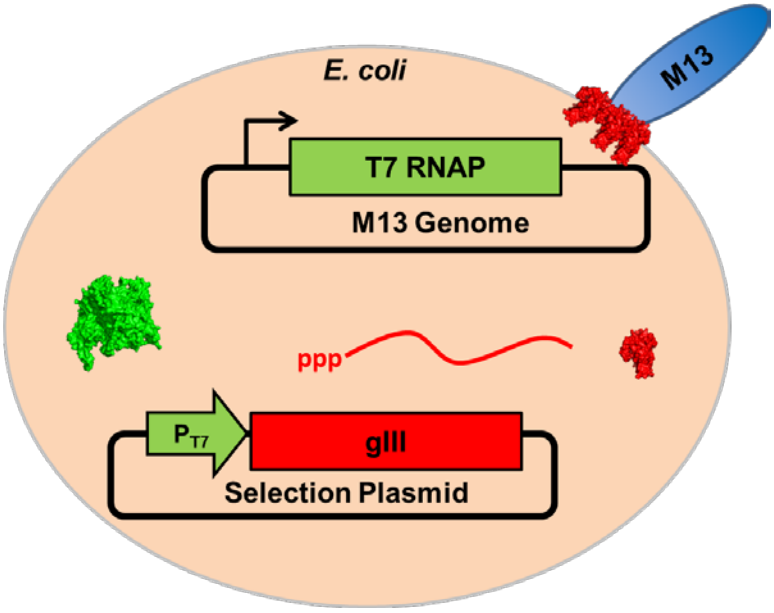


Figure 1-5. Phage-assisted continuous evolution.
A library of M13 bacteriophage, lacking gene III (gIII), but possessing a T7 RNA polymerase (T7 RNAP) library infects *E. coli* host cells. The phage-provided T7 RNA polymerase must recognize a synthetic T7 promoter variant (P_{T7}) in order to produce the gene III product (red), which is needed to create infectious progeny (Esvelt *et al*., 2011).

Rational engineering of promoter specificity mutants by phylogenetic part mining

The Voigt Lab took a part-mining approach to the problem of T7 RNA polymerase substrate specificity (Temme *et al*., 2012a). The promoter specificity loops (amino acid residues 745 to 761) from three related bacteriophage-encoded RNA polymerases (T3, K1F, and N4) were grafted in place of the T7 RNA polymerase specificity loop. The resulting chimeras were able to use promoters resembling the promoter used by those bacteriophage polymerases with about a 10-fold specificity over non-cognate promoters.

Thermal stability

The utility of T7 RNA polymerase in various *in vitro* transcription assays has driven interest in its ability to be used as part of an isothermal nucleic acid detection system based on cycles of transcription and reverse transcription (Kwoh *et al*., 1989; Guatelli *et al*., 1990; Compton, 1991). This has, in turn, driven a need for a thermal-stable RNA polymerase, so that these techniques can be performed at elevated temperatures, thus reducing mispriming and problematic secondary structure. This problem was addressed using an antibiotic-based selection system similar to that depicted in **Figure 1-3** where active T7 RNA polymerase variants are enriched based on their ability to confer chloramphenicol resistance to their host. The distinctive elements were the nature of the host, *Geobacillus stearothermophilus* (formerly *Bacillus stearothermophilus*), and the plating conditions (45 °C to 50 °C) (Sugiyama *et al*., 2009). The resulting mutants were sequenced and the mutations S430P, S633P, F849I, and F880Y were found to predominate. The reported oligonucleotide primers code for a fifth mutation N433T, but the mutation is not referenced in the text. The mutant reported is active at 37 °C and 50 °C.

Promoter clearance

Like other RNA polymerases, T7 RNA polymerase goes through cycles of abortive transcription before transitioning to the elongation phase. This leads to a heterogeneous mixture of RNA products in an *in vitro* transcription reaction. This can be especially problematic when a non-ideal initially transcribed sequence (ITS) is utilized. As noted above, initiation with purines (especially GTP) is quite favorable for *de novo* transcription. Pyrimidine rich initiation can lead to low yields and excessive accumulation of short products. To overcome this limitation, a directed evolution approach was taken (Guillerez *et al*., 2005). A library of T7 RNA polymerase mutants (made by hydroxylamine mutagenesis) was transformed into an *E. coli* strain in which the T7 promoter (with a non-ideal ITS) drives the lacZ gene. Plating on X-gal revealed clones of T7 RNA polymerase capable of initiating transcription with a poor ITS. The P266L mutation, especially, mitigated the ill effects of a T-rich ITS, facilitating promoter clearance and a transition to elongation.

Substrate specificity

Nucleic acid with modified bases can be chemically synthesized, but it is often preferable to produce modified RNA by *in vitro* transcription, especially for *in vitr*o selection (Knudsen *et al*., 2002). T7 RNA polymerase is an ideal enzyme for the production of large quantities of RNA *in vitro*, and has previously been engineered and evolved to have an expanded substrate range.

Rational mutations conferring expanded substrate specificity in T7 RNA polymerase

The Y639F mutation in T7 RNA polymerase was shown to result in a lower protein yield per transcript, suggesting a role in transcript homogeneity (Makarova, 1995) and the corresponding residue in structurally homologous DNA polymerase was known

to be important for fidelity (Carroll *et al*., 1991) and for binding dNTPs (Polesky *et al*., 1990). These insights led multiple groups to investigate the substrate specificity of the Y639F mutant. Y639F (now available commercially as "durascribe") allows for the *in vitro* transcription of RNA containing dNTPS as well as nucleotides with 2′-fluoro and 2′-amino modified ribose (Kostyuk *et al*., 1995; Sousa and Padilla, 1995; Padilla and Sousa, 1999).

The above relationship between Y639F and the 2′-OH group of the NTP substrate was confirmed by several crystal structures (Cheetham and Steitz, 1999; Tahirov *et al*., 2002; Temiakov *et al*., 2004) (**Figure 1-6**). The hydroxyl group of Y639 makes direct contact with the 2′-OH group of the incoming nucleotide, and is thus associated with the incorporation of unnatural substrates. The structures also reveal that H784 makes contact with the 2′-OH group of the most recently added nucleotide. H784A mutations were therefore investigated. The Y639F, H784A ("FA") double mutant can incorporate nucleotides with bulky modifications at the 2′ position (*e.g.* 2′-*O*-methyl) (Brieba and Sousa, 2000; Padilla and Sousa, 2002) presumably because the H784A mutation eliminates premature termination following the incorporation of a modified nucleotide.

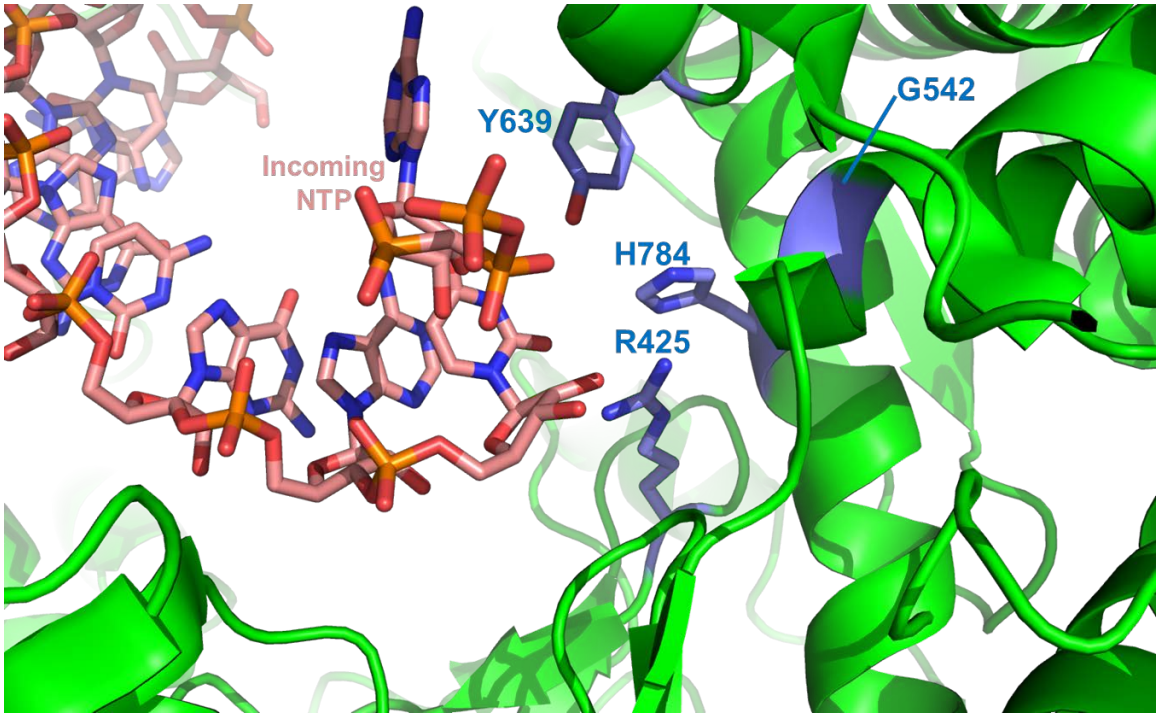


Figure 1-6. T7 RNA polymerase interaction with ribose moieties.
The nascent RNA and incoming nucleotide (pink) shown inside the T7 RNA polymerase active site (green and blue). DNA and some protein loops have been omitted for clarity. Y639 makes contact with the 2′-OH group of the incoming nucleotide and H784 makes contact with the 2′-OH group of the most recently added nucleotide. R425, G542, Y639, and H784 (blue) have been the subject of directed evolution studies. This figure is based on the crystal structure of T7 RNA polymerase in the elongation phase, PBD accession number 1SOV (Temiakov *et al*., 2004).

Directed evolution of expanded substrate specificity in T7 RNA polymerase

The Ellington Lab has previously taken a directed evolution approach, in which the aforementioned Y639 and H784 residues, as well as the important R425 and G542 were randomized (Chelliserrykattil and Ellington, 2004). The resulting library was subjected to an antibiotic-based selection similar to that described above (**Figure 1-3**) in which hosts harboring active polymerases could survive on chloramphenicol plates. Thus the population was enriched for T7 RNA polymerase variants that retained the ability to transcribe RNA *in vivo* (with natural ribose). This population was then screened for

altered substrate specificities *in vitro*. A mutant, termed "RGVG," (R425, G542, Y639V, H784G plus additional E593G and V685A mutations that arose organically during the selection) showed strong activity with 2′-*O*-methyl UTP. A second mutant, termed "VRS," (G542V and H784S as well as an additional H772R mutation) was able to incorporate 2′-fluoro modified pyrimidines.

More recent works include a high throughput screen for derivatives of RGVG with enhanced 2′-*O*-methyl recognition. This approach uncovered the “2P16” mutant, a mutant of RGVG with seven additional mutations (Siegmund *et al*., 2012). A similar approach, with R425 randomized yielded the “R425C” mutant (Ibach *et al*., 2013). Each of these mutants is reported to enable the incorporation of 2′-*O*-methylnucleotides.

CONCLUSION

The above biological insights into, laboratory uses of, and attempts to improve T7 RNA polymerase have informed and shaped experiments that are the subject of the remainder of this dissertation. Some of the limits of the directed evolution schemes previously described provided lessons into the nature of molecular evolution. Some of the mutants created in the works described in this chapter are directly compared to the mutants described in the ensuing chapters. The strengths and weaknesses of various approaches will be explored in an attempt to lay out the best strategies for future directed evolution undertakings.

REFERENCES

Agresti, J. J., Kelly, B. T., Jäschke, A., and Griffiths, A. D. 2005. Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America* 102:16170–5.

Bartel, D., and Szostak, J. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418.

Brieba, L. G., and Sousa, R. 2000. Roles of histidine 784 and tyrosine 639 in ribose discrimination by T7 RNA polymerase. *Biochemistry* 39:919–23.

Bull, J. J., Springman, R., and Molineux, I. J. 2007. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. *Molecular Biology and Evolution* 24:900–8.

Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. 2014. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* 10:216–22.

Carroll, S. S., Cowart, M., and Benkovic, S. J. 1991. A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* 30:804–813.

Chamberlin, M., McGrath, J., and Waskell, L. 1970. New RNA polymerase from Escherichia coli infected with bacteriophage T7. *Nature* 228:227–231.

Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. 1999. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399:80–3.

Cheetham, G. M., and Steitz, T. A. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286:2305–9.

Chelliserrykattil, J., Cai, G., and Ellington, A. D. 2001. A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnology* 1:13.

Chelliserrykattil, J., and Ellington, A. D. 2004. Evolution of a T7 RNA polymerase variant that transcribes 2’-O-methyl RNA. *Nature Biotechnology* 22:1155–60.

Compton, J. 1991. Nucleic acid sequence-based amplification. *Nature* 350:91–92.

Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. 1996. A T7 promoter-specific, inducible protein expression system for Bacillus subtilis. *Molecular & General Genetics* 250:230–236.

Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. 2012. An in vitro Autogene. *ACS Synthetic Biology* 1:190–196.

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*.

Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N., and Yanagawa, H. 2004. In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Research* 32:e95.

Dubendorff, J. W., and Studier, F. W. 1991. Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. *Journal of Molecular Biology* 219:61–8.

Dunn, J. J., and Studier, F. W. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology* 166:477–535.

Ellington, A. D., and Szostak, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.

Elroy-Stein, O., and Moss, B. 1990. Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 87:6743–7.

Esvelt, K. M., Carlson, J. C., and Liu, D. R. 2011. A system for the continuous directed evolution of biomolecules. *Nature* 472:499–503.

Fuerst, T. R., Niles, E. G., Studier, F. W., and Moss, B. 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 83:8122–6.

Gamer, M., Fröde, D., Biedendieck, R., Stammen, S., and Jahn, D. 2009. A T7 RNA polymerase-dependent gene expression system for Bacillus megaterium. *Applied Microbiology and Biotechnology* 82:1195–203.

Griffiths, A. D., and Tawfik, D. S. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *The EMBO Journal* 22:24–35.

Guatelli, J. C., Whitfield, K. M., Kwoh, D. Y., Barringer, K. J., Richman, D. D., and Gingeras, T. R. 1990. Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proceedings of the National Academy of Sciences of the United States of America* 87:7797.

Guillerez, J., Lopez, P. J., Proux, F., Launay, H., and Dreyfus, M. 2005. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America* 102:5958–63.

Ibach, J., Dietrich, L., Koopmans, K. R. M., Nöbel, N., Skoupi, M., and Brakmann, S. 2013. Identification of a T7 RNA polymerase variant that permits the enzymatic synthesis of fully 2'-O-methyl-modified RNA. *Journal of Biotechnology* 167:287–95.

Ikeda, R. A., Ligman, C. M., and Warshamana, S. 1992. T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Research* 20:2517–24.

Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* 39:10419–30.

Jeruzalmi, D., and Steitz, T. A. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal* 17:4101–13.

Kennedy, W. P., Momand, J. R., and Yin, Y. W. 2007. Mechanism for de novo RNA synthesis and initiating nucleotide specificity by T7 RNA polymerase. *Journal of Molecular Biology* 370:256–68.

Kim, J., White, K. S., and Winfree, E. 2006. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology* 2:68.

Kim, J., and Winfree, E. 2011. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology* 7:465.

Knudsen, S. M., Robertson, M. P., and Ellington, A. D. 2002. In vitro selection using modified or unnatural nucleotides. *Current Protocols in Nucleic Acid Chemistry* Chapter 9:Unit 9.6.

Kochetkov, S. N., Rusakova, E. E., and Tunitskaya, V. L. 1998. Recent studies of T7 RNA polymerase mechanism. *FEBS Letters* 440:264–7.

Kostyuk, D. A., Dragan, S. M., Lyakhov, D. L., Rechinsky, V. O., Tunitskaya, V. L., Chernov, B. K., and Kochetkov, S. N. 1995. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Letters* 369:165–8.

Krüger, D. H., and Schroeder, C. 1981. Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. *Microbiological Reviews* 45:9–51.

Kwoh, D. Y., Davis, G. R., Whitfield, K. M., Chappelle, H. L., DiMichele, L. J., and Gingeras, T. R. 1989. Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proceedings of the National Academy of Sciences of the United States of America* 86:1173–7.

Lau, M. W. L., and Ferré-D'Amaré, A. R. 2013. An in vitro evolved glmS ribozyme has the wild-type fold but loses coenzyme dependence. *Nature Chemical Biology* 9:805–10.

Levy, M., and Ellington, A. D. 2008. Directed evolution of streptavidin variants using in vitro compartmentalization. *Chemistry & Biology* 15:979–89.

Levy, M., Griswold, K. E., and Ellington, A. D. 2005. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* 11:1555–62.

Lu, W. C., Levy, M., Kincaid, R., and Ellington, A. D. 2014. Directed evolution of the substrate specificity of biotin ligase. *Biotechnology and Bioengineering* 111:1071–81.

Makarova, O. 1995. Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proceedings of the National Academy of Sciences of the United States of America* 92:12250–12254.

Mastrobattista, E., Taly, V., Chanudet, E., Treacy, P., Kelly, B. T., and Griffiths, A. D. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & Biology* 12:1291–300.

McAllister, W. T., Morris, C., Rosenberg, A. H., and Studier, F. W. 1981. Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *Journal of Molecular Biology* 153:527–44.

McBride, K. E., Schaaf, D. J., Daley, M., and Stalker, D. M. 1994. Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 91:7301–5.

Melton, D., and Krieg, P. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research* 18:38–44.

Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research* 15:8783–98.

Padilla, R., and Sousa, R. 2002. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Research* 30:e138.

Padilla, R., and Sousa, R. 1999. Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutant T7 RNA polymerase (RNAP). *Nucleic Acids Research* 27:1561–1563.

Polesky, A. H., Steitz, T. A., Grindley, N. D., and Joyce, C. M. 1990. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from Escherichia coli. *The Journal of Biological Chemistry* 265:14579–91.

Raskin, C. A., Diaz, G. A., and McAllister, W. T. 1993. T7 RNA polymerase mutants with altered promoter specificities. *Proceedings of the National Academy of Sciences of the United States of America* 90:3147–51.

Raskin, C. A., Diaz, G., Joho, K., and McAllister, W. T. 1992. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *Journal of Molecular Biology* 228:506–15.

Rong, M., He, B., McAllister, W. T., and Durbin, R. K. 1998. Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 95:515–9.

Schenborn, E. T., and Mierendorf, R. C. 1985. A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research* 13:6223–36.

Siegmund, V., Santner, T., Micura, R., and Marx, A. 2012. Screening mutant libraries of T7 RNA polymerase for candidates with increased acceptance of 2'-modified nucleotides. *Chemical Communications* 48:9870–2.

Sousa, R. 1996. Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences* 21:186–90.

Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B. C. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature* 364:593–9.

Sousa, R., and Padilla, R. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *The EMBO Journal* 14:4609–21.

Steitz, T. A., Smerdon, S. J., Jäger, J., and Joyce, C. M. 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266:2022–5.

Stoflet, E., Koeberl, D., Sarkar, G., and Sommer, S. 1988. Genomic amplification with transcript sequencing. *Science* 239:491–494.

Studier, F. W. 1972. Bacteriophage T7. *Science* 176:367–76.

Studier, F. W., and Moffatt, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189:113–30.

Studier, F. W., and Rosenberg, A. H. 1981. Genetic and physical mapping of the late region of bacteriophage T7 DNA by use of cloned fragments of T7 DNA. *Journal of Molecular Biology* 153:503–25.

Subsoontorn, P., Kim, J., and Winfree, E. 2012. Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch. *ACS Synthetic Biology* 1:299–316.

Sugiyama, A., Nishiya, Y., and Kawakami, B. 2009. RNA Polymerase Mutants with Increased Thermostability. U. S. Patent. US 7,507,567 B2.

Tabor, S., and Richardson, C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America* 82:1074–1078.

Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., and Yokoyama, S. 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* 420:43–50.

Tawfik, D. S., and Griffiths, A. D. 1998. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* 16:652–656.

Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassylyev, D. G. 2004. Structural basis for substrate selection by T7 RNA polymerase. *Cell* 116:381–91.

Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. 2012a. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* 40:8773–81.

Temme, K., Zhao, D., and Voigt, C. A. 2012b. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America* 109:7085–90.

Tokmakov, A. A., Matsumoto, E., Shirouzu, M., and Yokoyama, S. 2006. Coupled cytoplasmic transcription-and-translation--a method of choice for heterologous gene expression in Xenopus oocytes. *Journal of Biotechnology* 122:5–15.

Tuerk, C., and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.

Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. 2012. Genome-scale promoter engineering by coselection MAGE. *Nature Methods* 9:591–3.

Wilson, C., and Szostak, J. 1995. In vitro evolution of a self-alkylating ribozyme. *Nature*.

Wirtz, E., Hoek, M., and Cross, G. A. 1998. Regulated processive transcription of chromatin by T7 RNA polymerase in Trypanosoma brucei. *Nucleic Acids Research* 26:4626–34.

Wochner, A., Attwater, J., Coulson, A., and Holliger, P. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332:209–212.

Yin, Y. W., and Steitz, T. A. 2002. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* 298:1387–95.

Zhang, X., and Studier, F. W. 1995. Isolation of transcriptionally active mutants of T7 RNA polymerase that do not support phage growth. *Journal of Molecular Biology* 250:156–68.

Chapter 2: T7 RNA polymerase evolution using an *in vitro* autogene system

Biological life likely started with simple molecular replications, which, over time, elaborated function becoming increasingly complex multi-molecular consortia (Meyer *et al*., 2012). The study of abiotic replication systems can therefore lend insight into the origins of life. Among the simplest molecular replications are small nucleic acids that can catalyze their own formation from smaller parts (von Kiedrowski, 1986; Zielinski and Orgel, 1987; Lincoln and Joyce, 2009). These systems are simple enough that they are resistant to passively replicating parasites, but more complicated systems that include enzymes in the replication process are not so lucky. As was seen with the Qβ bacteriophage genome, continuous nucleic acid replication by an enzyme leads to the accumulation of shorter nucleic acids (Haruna and Spiegelman, 1965) because they can be replicated faster than their full-length competitors (Bull and Pease, 1995). This problem can be mitigated by the introduction of compartments to confine enzymatic function (linking genotype to phenotype) and exclude parasites. The abiotic self-replication of the Qβ replicase (Urabe *et al*., 2010) and *Taq* DNA polymerase (Ghadessy *et al*., 2001) with *in vitro* compartmentalization has previously been shown. In this chapter, the creation of an *in vitro* T7 RNA polymerase replicator, its properties, and its lessons for molecular replication and evolution are discussed.

INTRODUCTION

Background and rationale

The directed evolution systems described in the previous chapter each possess their own strengths and limitations. Selections in which T7 RNA polymerase mutants are tasked with the expression of chloramphenicol acetyltransferase (Ikeda *et al*., 1992;

Sugiyama *et al*., 2009; Chelliserrykattil and Ellington, 2004) are limited by a problem of thresholding. That is, if a mutant is active enough to produce a minimal amount of chloramphenicol acetyltransferase it will survive the selection. Any mutant that is more active gets no benefit from this added activity, and also faces a penalty because it puts a higher metabolic load on its host, lowering its growth rate. Selections with this mechanism are therefore prone to generating RNA polymerases with moderate activity.

The *in vivo* autogene system (**Figure 1-4**), in contrast, is not limited by thresholding (Dubendorff and Studier, 1991; Chelliserrykattil *et al*., 2001; Finn *et al*., 2004). A more active T7 RNA polymerase can, in theory, make more copies of its own mRNA across all levels of activity. In practice however, the most active autogenes likely inhibited host growth enough to eliminate themselves from the population. Thus, these selections also produced RNA polymerases of moderate activity. Indeed, the Q758C mutant was found using the GACT (wild type) promoter far more often than the GACG promoter in this experiment, but Q758C will be shown to prefer the GACG promoter by at least 2.5 fold (**Table 3-2**). Thus, the selection favored an inferior polymerase:promoter pair when a more active pair was possible.

In order to eliminate the effects of thresholding and host growth, it was proposed to eliminate the hosts altogether. This could be accomplished by replacing some of the functions of the host (*i.e.* translation machinery) with a cell-free translation lysate. This would allow a large library of autogenes (in the form of degenerate PCR products) to be transcribed (kick-started by a different RNA polymerase) and then translated by the lysate machinery. The characteristic positive feedback would then ensue, allowing functional autogenes to be exponentially amplified, without regard to the health of any host.

Bacterial hosts, while fragile, provide a directed evolution system with the critical element of compartmentalization. Since each host cell usually receives only one copy of the gene under selection, it is physically isolated from all other competitors. Thus, its genotype and phenotype are effectively linked. The use of cell-free lysate as the basis of selection removes this inherent advantage to cell-based systems. A new solution can be found in *in vitro* compartmentalization (IVC) by water-in-oil emulsion. IVC has been shown to be effective in confining a gene product's function (Tawfik and Griffiths, 1998) and in preventing shorter parasites from overrunning an entire population (Urabe *et al*., 2010). Thus a T7 autogene in cell-free lysate with *in vitro* compartmentalization should, in theory, be evolvable, and the most functional variants should thrive.

Prior work toward implementation of the *in vitro* autogene

The *in vitro* autogene was first implemented by Eric Davidson of the Ellington Lab. His early efforts were based on the *E. coli* S30 Extract System for Linear Templates. This lysate contained the *E. coli* RNA polymerase, which was used to kick-start the auto-feedback system by initially transcribing T7 RNA polymerase from a tetA promoter. The positive feedback transcription would proceed from a T7 promoter located upstream of the tetA promoter, thus "feedback transcript" mRNA would contain a unique sequence not found in the "initial transcript" mRNA. While protein production could be monitored in this system, mRNA (from either promoter) could not be detected. It was determined after many trials that a eukaryotic system (rabbit reticulocyte lysate; RRL) initiated by the SP6 RNA polymerase was the most suitable implementation of the *in vitro* autogene (Davidson *et al*., 2012) (**Figure 2-1**).

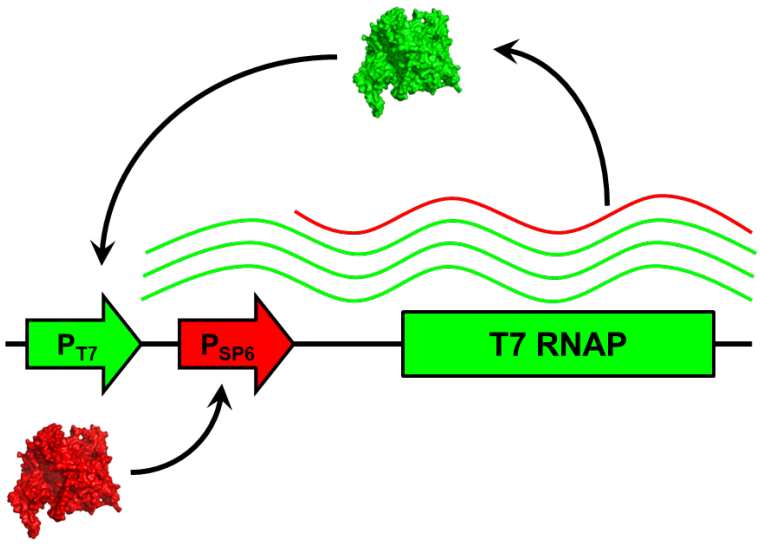


Figure 2-1. *In vitro* autogene.
SP6 RNA polymerase (red) drives initial transcription from its cognate promoter (P_{SP6}; red arrow), producing "initial transcript" (red line). This is translated into T7 RNA polymerase (green), which then transcribes "feedback transcript" (green line) from the T7 promoter (P_{T7}; green arrow).

While feedback transcripts could be detected from the transcription and translation (TNT) reaction in bulk solution, the *in vitro* autogene needed to be able to function in emulsion in order to allow evolution to occur. Several factors were optimized in order to improve the function of the system. An oil and surfactant mixture of 96% mineral oil (Sigma) and 4% Abil EM90 (Evonik) had previously been shown to be compatible with RRL (Ghadessy and Holliger, 2004), and was found to be suitable for this system. The inclusion of 5′ biotin groups stabilized the DNA template and increased mRNA yield. An F880Y mutation, previously shown to aid in thermal stability (Sugiyama *et al*., 2009), was added to the T7 RNA polymerase open reading frame. Finally, the inclusion of an EMCV IRES facilitated translation in the eukaryotic system. With the improved formulation, feedback transcripts accounted for 40% of mRNA recovered.

Before attempting a selection with a complex library, a two-member library "mock" selection was performed by Eric Davidson. Initially, 10% of the population was active autogene (WT) and 90% was inactive autogene (containing premature stop codons; XX). After TNT in emulsion, recovery of mRNA, and RT-PCR, eight of 10 clones sequenced were WT, suggesting at least moderate enrichment of active sequences. However, these efforts were not built upon because multiple rounds of selection could not be performed. An initial "library" comprising only WT templates could go through one round of emulsion TNT, recovery, and RT-PCR, but the reamplification of the second such round produced a smeared product. Sequencing of this population showed at least two mutations per kilobase as well as large deletions (Davidson *et al*., 2012). No RT-PCR band could be seen from Round 3 reamplification. At this point, the author and Jared Ellefson of the Ellington Lab began work on the implementation of the *in vitro* autogene as a platform for directed evolution.

RESULTS AND DISCUSSION

In vitro autogene optimization

Rounds of selection of the *in vitro* autogene are depicted in **Figure 2-2.** Essentially, purified autogene template DNA is added to rabbit reticulocyte lysate (RRL) on ice (to prevent premature reactions), and the mixture is emulsified such that each compartment is expected to contain zero or one DNA template. SP6 RNA polymerase initiates transcription, and the initial transcript is translated by RRL machinery. The newly translated T7 RNA polymerase then transcribes from the T7 promoter. After an incubation period, the emulsion is broken with chloroform, mRNA is recovered, and template DNA is digested. Transcription from the T7 promoter is specifically amplified

by RT-PCR. Additional PCR steps use primers that reintroduce the T7 promoter and 5′ biotin groups. The template DNA is then prepared for further rounds of selection.

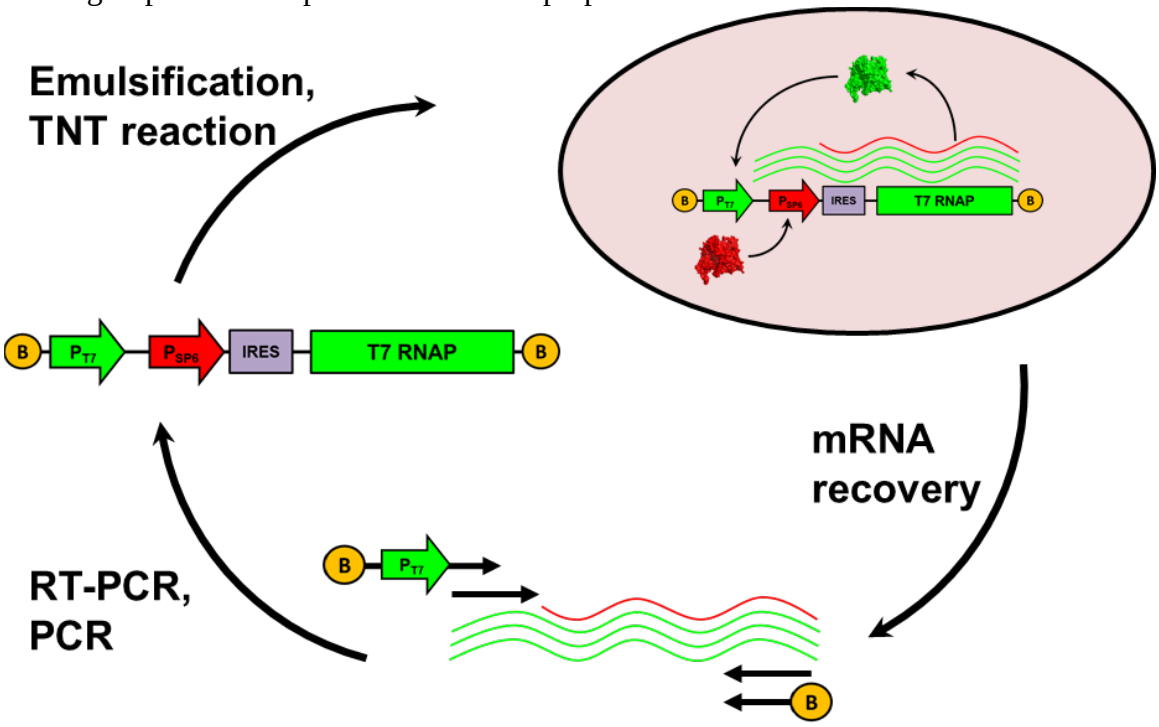


Figure 2-2. ***In vitro*** **autogene based selection.**
The autogene template DNA is emulsified in rabbit reticulocyte lysate, wherein a transcription and translation (TNT) reaction occurs. After an incubation period, mRNA is recovered, and feedback transcripts are specifically amplified by RT-PCR. Subsequent PCR steps reintroduce the T7 promoter (P_{T7}) and 5′ biotin groups, preparing the DNA for further rounds of selection.

The above schema, however, cannot be propagated past a second round even when the initial library input is 100% active autogenes (WT), and by the second round, products are truncated and heavily mutated. **Figure 2-3** shows typical RT-PCR products from a first or second round of selection. The two groupings on the left show failed cycle course RT-PCRs of two independent Round 2 mRNAs. The middle grouping is a negative control in which the input DNA was all non-function autogenes with two

premature stop codons (XX). The two groupings on the right show successful cycle course RT-PCRs of two independent Round 1 mRNAs.

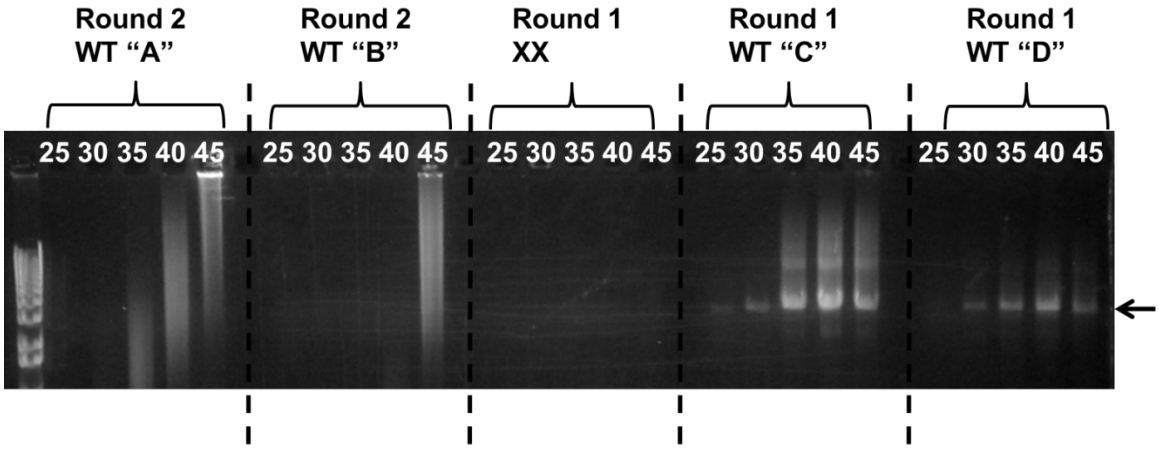


Figure 2-3. Failed propagation of the *in vitro* autogene selection.
For WT A-D, the initial library input DNA is 100% active autogenes. For XX, the initial library input DNA is 100% non-function autogenes with two premature stop codons (XX). A and B are similar, C and D are similar. A and B show the cycle course RT-PCR of the mRNA recovered from a second round of selection, which is smeared. C and D show the cycle-course RT-PCR of the mRNA recovered from a first round of selection, which amplifies at the correct size (arrow). XX does not show a band, as expected.

One of the major problems seen in the prior attempts at turning serial rounds of selection with the WT autogene was a high error rate (about one mutation per kilobase per round). This is partially due to the error associated with transcription by T7 RNA polymerase, which has an error rate of about 6×10^{-5} in cells (this may be higher in emulsion) (Brakmann and Grzeszik, 2001), and reverse transcription by SuperScript III, which has an error rate of about 3×10^{-5} (Goldsmith and Tawfik, 2009). A larger contributor was likely the error rate of Platinum *Taq* DNA polymerase High Fidelity (a mixture of hot start *Taq* DNA polymerase and Deep Vent DNA polymerase), which was used in all PCR steps. Platinum *Taq* DNA polymerase High Fidelity is used for "tricky" amplification and has been seen to work when other DNA polymerases do not. Thus, there was concern that a move away from Platinum *Taq* DNA polymerase High Fidelity

would hamper efforts to amplify cDNA from the emulsion (which is at a low concentration and possibly co-purified with inhibitors). As a test, autogene cDNA was amplified by Accuprime Pfx DNA polymerase (KOD DNA polymerase), Pfx50 DNA polymerase (from *Thermococcus zilligii*), Phusion DNA polymerase (a modified *Pfu* DNA polymerase), and PfuUltra II Hotstart (a modified *Pfu* DNA polymerase) as well as Platinum *Taq* DNA polymerase High Fidelity. Only Platinum *Taq* DNA polymerase High Fidelity and PfuUltra II Hotstart could effectively amplify the cDNA (**Figure 2-4**). The fidelity of PfuUltra II Hotstart was assayed by PCR amplifying a plasmid of known sequence and sequencing the resulting amplicon. After 50 cycles of PCR (the number commonly faced between rounds), no errors were found in several kilobases of sequence. It should be noted that PfuUltra II Hotstart is often used for general cloning and is considered to have one of the lowest error rates among commercially available DNA polymerases. PfuUltra II Hotstart was therefore used in all subsequent autogene selections.

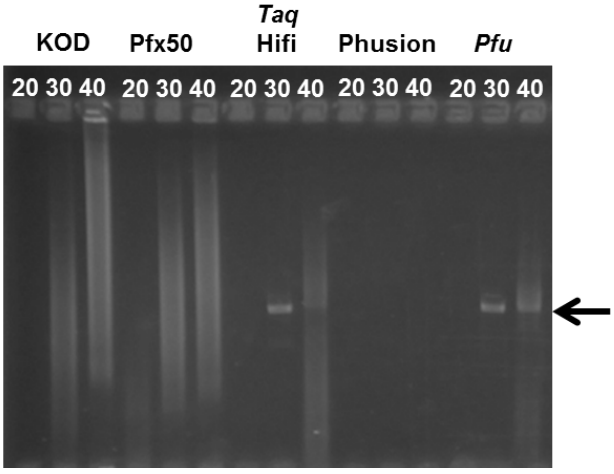


Figure 2-4. DNA polymerase comparison for PCR amplification of autogene cDNA. Accuprime Pfx DNA polymerase (KOD), Pfx50 DNA polymerase (Pfx50), Platinum *Taq* DNA polymerase High Fidelity (*Taq* Hifi), Phusion DNA polymerase (Phusion), and PfuUltra II Hotstart DNA polymerase (*Pfu*) were used to PCR amplify autogene cDNA. A variable number of cycles were performed.

The other major problem in the previous instantiation of the autogene selection was the accumulation of large deletions, which presumably arise during reverse transcription. One possible explanation is excessive secondary structure in the EMCV IRES and the T7 RNA polymerase coding sequence. To address the issue of structure in the IRES, the "IRESite" database was searched for suitable alternatives (Mokrejs *et al*., 2010). The database contained 10 IRESes that had been experimentally validated in rabbit reticulocyte lysate (RRL). Three of these were eliminated from consideration due to their lengths in excess of 900 nucleotides. An additional two IRESes were considered after a literature search suggested their compatibility with RRL (Kang *et al*., 2009; Vallejos *et al*., 2010). The nine IRES sequences under consideration were inputted into NUPAC folding software, and analyzed for secondary structure at several temperatures. All IRESes have considerable secondary structure at 30 °C, the temperature at which they function. Three IRESes (SWSS, crTMV, and Gypsy) were markedly less structured than the EMCV IRES at 55 °C, the temperature at which they serve as the template for reverse transcription (**Figure 2-5**). Each of these IRESes were built from oligodeoxynucleotides using thermodynamically balanced inside-out PCR (Gao, 2003). These IRESes and a random "N50" sequence were each assembled into autogene constructs and tested for translation in RRL using [^{35}S]-methionine for detection by autoradiography. SWSS appeared to be non-functional and was removed from consideration. EMCV, N50, Gypsy, and crTMV were then tested for their compatibility with reverse transcription. mRNA was transcribed, purified, and reverse transcribed using [α^{32}P]dATP for detection by autoradiography (**Figure 2-5**). crTMV was not well reverse-transcribed. Full length product from EMCV-mRNA could be seen, but shorter products were apparent. Gypsy-mRNA was slightly better and N50 was by far the best. Using an N50 pool as an IRES could be problematic (enhancing or reducing translation and forming cryptic promoters

or primer binding sites). However, reports have suggested that unstructured regions are all that is needed for translation initiation in lysate (Mureev *et al*., 2009), and the exact N50 sequences were unlikely to affect translation. Cryptic promoters and primer binding sites would be disfavored and lost from the population. It was therefore judged that these considerations were outweighed by the improved reverse transcription. The N50 leader sequence was used in all subsequent autogene selections performed by Jared Ellefson, and the Gypsy IRES was used in selections performed by the author.

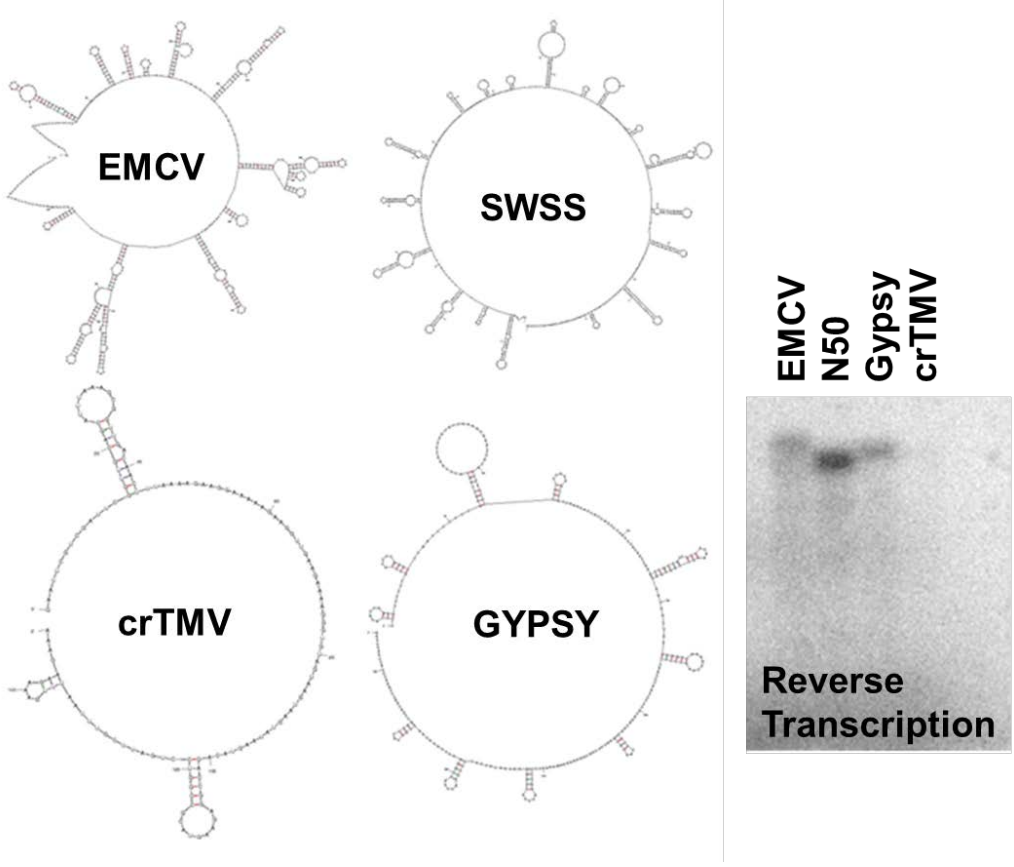


Figure 2-5. Replacement of the EMCV IRES.
Left) The NUPACK generated secondary structure of the least structured IRESes considered, as well as the EMCV IRES. The predicted folding of RNA at 55 °C is shown. **Right**) Reverse transcription of IRES-T7 RNA polymerase mRNA. The N50 "IRES" gave the most robust signal, followed by the Gypsy IRES.

Aside from IRESes, the natural T7 RNA polymerase coding sequence is heavily structured. Some of the truncation likely came from improper reverse transcription of the open reading frame itself. To address this, Ben Braun of the Ellington Lab devised an algorithm to make iterative "mutations" to the mRNA and "evolve" toward a sequence of minimal secondary structure. The resulting sequence of T7 RNA polymerase with reduced secondary structure (T7 RSS) contains fewer and shorter stems and was expected to be less problematic than the wild type sequence (**Figure 2-6**). The T7 RSS gene was cloned, PCR amplified to create the linear autogene, and used in all subsequent selections.

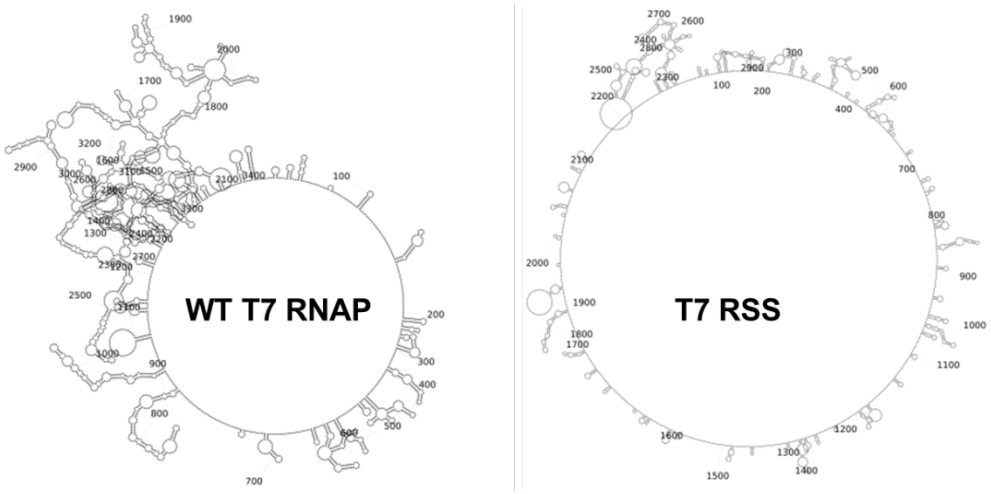


Figure 2-6. Recoding the T7 RNA polymerase open reading frame.
The NUPACK generated secondary structure of the wild type T7 RNA polymerase (WT T7 RNAP) coding sequence and that of the T7 RNA polymerase with reduced secondary structure (T7 RSS). The predicted folding of RNA at 55 °C is shown.

In vitro autogene selection

Having implemented the described changes, both Jared Ellefson and the author were able to routinely turn two to three rounds before smearing occurred. Jared Ellefson was able to turn four successful rounds (with a smear on the fifth round). Several clones from each round of this selection were sequenced, assayed, and analyzed by the author.

Eight to 10 variants from each round of selection were cloned and sequenced (**Table 2-1**). Analysis of the sequencing data from each round of selection revealed that that while the number of point mutations increased with each round (about 2.5 per gene per round), the number of catastrophic mutations (nonsense and frameshift) did not increase and the proportion of synonymous mutations increased each round.

Round	1	2	3	4
Point mutations per kb	0.8	1.7	2.3	3.4
Synonymous mutations per kb	0.2	0.6	0.8	1.4
Percent of mutations that are synonymous	20	33	35	40
Percent predicted viable sequences	70	67	70	88

Table 2-1. Mutational analysis from an *in vitro* autogene selection.

Although the Round 4 population was heavily mutated (averaging about 11 mutations per gene), many of the mutations mapped to the peripheral loops of the fingers domain and the less catalytically relevant N-terminal domain (**Figure 2-7**). There were few mutations in the promoter contact region (residues 85 to 106, 742 to 761, and 230 to 243) or active site (residues 420 to 442, 536 to 543, 631 to 641, 778 to 788, and 808 to 814) (**Table 2-2**).

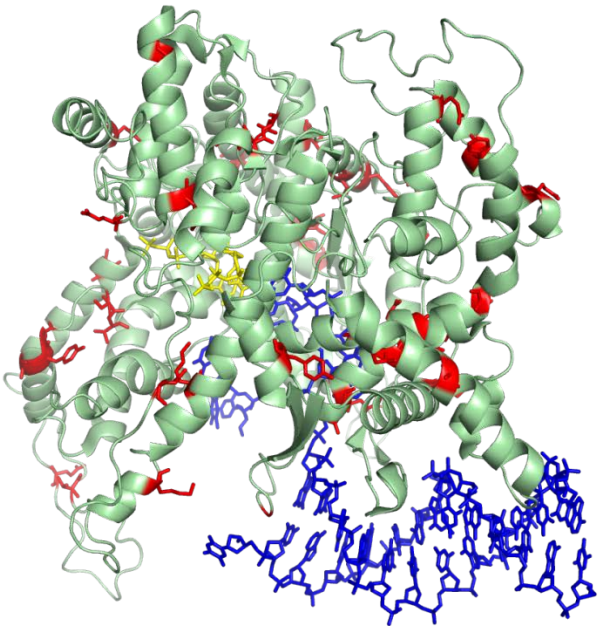


Figure 2-7. Location of mutations found in Round 4 of an *in vitro* autogene selection.

T7 RNA polymerase (pale green), DNA including the T7 promoter (blue), and incoming nucleotides (yellow) are shown along with mutations found in any of the eight clones from Round 4 of the *in vitro* autogene selection (red).

Rd4.1	Rd4.2	Rd4.3	Rd4.4	Rd4.5	Rd4.6	Rd4.7	Rd4.8
G235E	Q135H	F151L	T121A	I141V	F21L	A61V	E15G
Y250R	A622T	K378R	E358G	R143G	E30G	A83T	H205R
E504G	A708V	K387E	P533L	R150W	G142R	L446P	Y317H
V833A	5 silent	F475L	I573V	T729A	I145T	I482V	T445A
D844G		Y623H	E593Q	I320T	I281T	V796D	6 silent
D847N		D847N	N762D	F509L	V521A	2 silent	
4 silent		2 silent	6 silent	V609A	V629G		
		1 deletion		G624D	T729M		
		1 stop		S628G	2 silent		
				C723R			
				6 silent			

Table 2-2. Sequence of clones from Round 4 of an *in vitro* autogene selection.

A functional assay showed that six out of eight clones from Round 4 were active, despite being highly mutated (**Figure 2-8**). All of this suggests that "purifying selection" (Bershtein *et al*., 2008) was occurring, preferentially purging the deleterious mutations at the expense of neutral mutations. However, the eventual collapse of the system suggests the low activity and high mutation rate of the system could not avoid mutating itself to death.

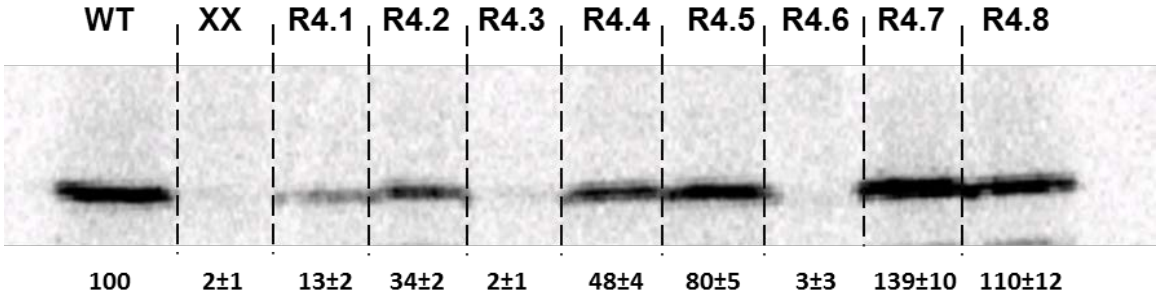


Figure 2-8. Function assay of eight clones from Round 4 of an *in vitro* autogene selection.
T7 RNA polymerase variants were expressed in RRL, along with a P_{T7}-GFP construct. PAGE and autoradiography (with [^{35}S]-methionine) shows that six of the variants could produce GFP. Wild type T7 RNA polymerase (WT) and negative control with premature stop codons (XX) are shown for comparison. The WT value is defined as 100. Numbers shown are the average and standard error from three independent measurements.

CONCLUSION

The limitations of the autogene can be understood as an example of "lethal mutagenesis" (Bull *et al*., 2007). Essentially, a nucleic acid replicator must replicate itself faster than it mutates itself in order to survive. More formally, to avoid extinction the following must be true:

$$f\,(e^{-Ud}) > 1 \text{ or } e^{Ud} < f$$

Where f is the fecundity of a replicator (the number of progeny per replicative cycle) and Ud is the average number of deleterious mutations that arise during a replicative cycle. In the context of the *in vitro* autogene selection, analysis of the sequences recovered from Round 1 indicated that there are about 1.5 non-synonymous mutations per gene. Other models of protein mutation from the literature suggest that about 37% of mutations are deleterious (Bershtein *et al.*, 2006, 2008; Soskine and Tawfik, 2010). Ud can then be assumed to be about 37% of 1.5, or about 0.56. The fact that extinction occurred suggests that fecundity must have been less than about 1.7.

Since the same number of input DNA molecules goes into each round, the overall fecundity of the population is 1.0. The firm constraint on the population size therefore means that the fecundity of a replicator is not merely an intrinsic property, but exists in the context of the library. That is, a fecundity of 1.7 does not mean that the active replicator makes 1.7 transcripts, but that it makes enough transcripts to increase its proportion of the population by 1.7-fold.

Interestingly, the lethal mutagenesis model with a constrained population size begins to resemble the "error catastrophe" (Eigen, 1971; Eigen and Schuster, 1977). Error catastrophe argues that a wild type replicator must have enough of a selective advantage over mutants to compensate for its error rate (*i.e.* the rate at which it creates mutants). Specifically, to avoid error catastrophe:

$$e^{\varepsilon} < \sigma$$

where σ is the selective advantage of a given sequence over its mutant and ε is the error rate per base pair multiplied by the number of base pairs (similar to Ud, above). As above, at least a 1.7 fold selective advantage is required to avoid catastrophe.

The limits of the autogene as a platform for directed evolution can be summed up as two problems: 1) the error rate is too high and 2) the selective advantage of the best

species is too low. In the next chapter, a new selection scheme is introduced that addresses both of these problems.

MATERIALS AND METHODS

Template preparation

Plasmids were generated in which the SP6 promoter was placed upstream of either wild type T7 RNA polymerase or a truncated version with two premature stop codons. The T7 promoter was not included in the plasmid, due to toxic effects on the host. PCR was then carried out with primers designed to extend the template to include the T7 promoter, a T7 terminator, and 5′ biotin groups. PCR was carried out with either Platinum *Taq* DNA polymerase High Fidelity (Life Technologies) or PfuUltra II Hotstart (Agilent). Intermediate PCR products were gel purified using Wizard Gel and PCR Clean-Up System (Promega). The final PCR product was further purified by phenol:chloroform extraction. In short, 100 µl of PCR reaction was mixed with 100 µl 25:24:1 phenol:chloroform:isoamylalcohol pH 7.9 (Ambion). After centrifugation, the (upper) aqueous phase was removed and mixed with 100 µl chloroform. After centrifugation, the (upper) aqueous phase was removed and mixed with 10 µl 3 M sodium acetate, and 300 µl 100% ethanol. The solution was stored at -20 °C for 30 minutes, and then centrifuged at 13,000 g at 4 °C for 30 minutes. The supernatant was removed and the pellet was washed with 70% ethanol before being air-dried and resuspended in 50 µl water. This yield of PCR product was quantified on an agarose gel using DNA quantitation standards.

In vitro transcription and translation reactions in emulsion

Selections were performed in TNT SP6 Quick Coupled Transcritpion Translation System (Promega) referred to as "RRL" hereafter. Reactions contained 3 ng – 30 ng (10^9

- 10^{10} molecules) of full-length autogene template in 50 µl RRL. Reactions were assembled on ice until emulsification.

Individual tubes (Sarstedt, 95 x 16.8 mm polypropylene) were set up that contained 96 µl mineral oil and 4 µl Abil EM90 (Evonik) and placed on ice. A stir bar (Spinplus 9.5 x 9.5 mm Teflon) was added to the tube containing the oil:surfactant mixture. The tube was moved into a beaker containing ice water on top of a magnetic stir plate (Corning) and stirred on the "high" setting (1150 rpm) for one minute. While stirring the oil:surfactant mixture, the 50 µl *in vitro* transcription and translation reaction was added drop-by-drop over one minute and then stirred for four additional minutes. The fully emulsified reaction was incubated at 30 °C for an additional two hours.

The emulsion reaction was stopped by placing the tube on ice for 10 minutes and adding Stop Solution (50 mM Tris, pH 7.0, 50 mM EDTA) sufficient to bring the aqueous volume to 100 µl total. The emulsion was broken by vortexing the reaction with one volume chloroform. The reaction was centrifuged at 13,000 g to separate the aqueous phase from the organic phase. The aqueous phase was removed for further purification.

The aqueous phase (after breaking an emulsion reaction as above) was added to 400 µl of Trizol reagent (Life Technologies). The tube was shaken and incubated for two minutes at room temperature. 150 µl of chloroform was added, the tube was again shaken and incubated for two minutes to allow the phases to separate. The reaction was centrifuged at 13,000 g at room temperature for two minutes. The aqueous phase was removed and precipitated by addition of 1 µg of glycogen and 0.7 volumes of isopropanol. After incubation at room temperature for 15 minutes the reaction was again centrifuged at 4 °C for 30 minutes. The supernatant was removed and the pellet was washed with 70% ethanol before being air-dried.

The pellet was resuspended in 43 µl water and 0.1 volume 10x DNase buffer, and then 2 Units Turbo DNase (Ambion) and 40 Units RNasin Plus RNase inhibitor (Promega) were added, bringing the final volume to 50 µl. The DNA digestion reaction was incubated at 37 °C for 30 minutes, and then 0.1 volumes DNase Inactivation Reagent was added. The solution was mixed gently for two minutes and the DNase Inactivation Reagent was removed by centrifugation at 13,000 g for two minutes at 4 °C. The aqueous supernatant was removed and could be stored at -80 °C or carried into further amplification reactions.

Template regeneration

Reverse transcription and PCR amplification of full-length mRNAs were performed in a single-step reaction using SuperScript III One-Step RT-PCR System with Platinum *Taq* High Fidelity (Life Technologies). The RT-PCR product was gel purified and used for regeneration of the autogene template. This protocol is the same as described above. Briefly, there were two regenerative PCR steps. The first PCR product was gel-purified and the second product was phenol:chloroform-extracted and ethanol-precipitated. The PCR steps add back the T7 promoter, 3′ UTR, and 5′ biotin groups.

Cloning and sequencing

Gel-purified RT-PCR products were amplified using primers that added appropriate restriction sites for cloning into a pASK sequencing vector. Eight to 10 clones from each round were sequenced using multiple primers that spanned the full length of the gene. Sequences were assembled and analyzed using Geneious (Biomatters).

T7 RNA polymerase activity assay

A SP6 promoter and EMCV IRES were appended to the genes for T7 RNA polymerase variants isolated from Round 4 via overlap PCR. DNA templates were gel-purified, and 3 x 10^9 templates were mixed with RRL, 200 pM [^{35}S]-methionine (1175 Ci /mmol; Perkins Elmer), and a plasmid containing the T7 promoter, EMCV IRES, and GFP gene. The reaction mixtures were incubated at 30 °C for two hours, followed by addition of Stop Solution, 10 mM DTT, and 4x LDS loading dye (Life Technologies). Samples were analyzed on a 4-12% NuPAGE gel (Life Technologies) with 1x MOPS, transferred to a 0.45 μm nitrocellulose membrane (Life Technologies), and exposed to a storage phosphor screen (Molecular Dynamics) before imaging on a STORM 840 Phospoimager (GE Healthcare). Autoradiographs were analyzed using ImageQuant (GE Healthcare). A band at 100 kDa indicated proper expression of the T7 RNA polymerase variants while a band at 27 kDa indicated expression of GFP, and thus T7 RNA polymerase activity.

Reverse transcription assay

To determine the suitability of IRES constructs, various autogene constructs were subject to *in vitro* transcription reactions using AmpliScribe-T7 High Yield Transcription Kit (Epicentre) under standard conditions. RNA was purified as above (trizol/chloform extraction, isopropanol precipitation, and turbo DNAse.) 5 μl RNA was reverse transcribed with SuperScript III reverse transcriptase (Life Technologies) under standard 20 μl reaction conditions. Reactions included 2 μl [α^{32}P]dATP (3000 Ci/mmol; 10 mCi/ml). Products were run on denaturing PAGE and gels were exposed to a storage phosphor screen (Molecular Dynamics) before visualization on a STORM 840 Phospoimager (GE Healthcare). Autoradiographs were analyzed using ImageQuant (GE Healthcare).

REFERENCES

Bershtein, S., Goldin, K., and Tawfik, D. S. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of Molecular Biology* 379:1029–44.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–32.

Brakmann, S., and Grzeszik, S. 2001. An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem* 2:212–9.

Bull, J. J., and Pease, C. M. 1995. Why Is the Polymerase Chain Reaction Resistant to In Vitro Evolution? *Journal of Molecular Evolution*:1160–1164.

Bull, J. J., Sanjuán, R., and Wilke, C. O. 2007. Theory of Lethal Mutagenesis for Viruses. *Journal of Virology* 81:2930–2939.

Chelliserrykattil, J., Cai, G., and Ellington, A. D. 2001. A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnology* 1:13.

Chelliserrykattil, J., and Ellington, A. D. 2004. Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nature Biotechnology* 22:1155–60.

Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. 2012. An in vitro Autogene. *ACS Synthetic Biology* 1:190–196.

Dubendorff, J. W., and Studier, F. W. 1991. Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. *Journal of Molecular Biology* 219:61–8.

Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523.

Eigen, M., and Schuster, P. 1977. The Hypercycle. *Naturwissenschaften* 64:541–565.

Finn, J., Lee, A. C. H., MacLachlan, I., and Cullis, P. 2004. An enhanced autogene-based dual-promoter cytoplasmic expression system yields increased gene expression. *Gene Therapy* 11:276–83.

Gao, X. 2003. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research* 31:143e–143.

Ghadessy, F. J., and Holliger, P. 2004. A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Engineering, Design & Selection* 17:201–4.

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

Goldsmith, M., and Tawfik, D. S. 2009. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proceedings of the National Academy of Sciences of the United States of America* 106:6197–202.

Haruna, I., and Spiegelman, S. 1965. Autocatalytic Synthesis of a Viral RNA in vitro. *Science* 150:884–886.

Ikeda, R. A., Ligman, C. M., and Warshamana, S. 1992. T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Research* 20:2517–24.

Kang, S. T., Leu, J. H., Wang, H. C., Chen, L. L., Kou, G. H., and Lo, C. F. 2009. Polycistronic mRNAs and internal ribosome entry site elements (IRES) are widely used by white spot syndrome virus (WSSV) structural protein genes. *Virology* 387:353–63.

Von Kiedrowski, G. 1986. A Self-Replicating Hexadeoxynucleotide. *Angewandte Chemie International Edition in English* 25:932–935.

Lincoln, T. A., and Joyce, G. F. 2009. Self-sustained replication of an RNA enzyme. *Science* 323:1229–1232.

Meyer, A. J., Ellefson, J. W., and Ellington, A. D. 2012. Abiotic Self-Replication. *Accounts of Chemical Research* 45:2097–2105.

Mokrejs, M., Masek, T., Vopálensky, V., Hlubucek, P., Delbos, P., and Pospísek, M. 2010. IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Research* 38:D131–6.

Mureev, S., Kovtun, O., Nguyen, U. T. T., and Alexandrov, K. 2009. Species-independent translational leaders facilitate cell-free expression. *Nature Biotechnology* 27:747–52.

Soskine, M., and Tawfik, D. S. 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews. Genetics* 11:572–82.

Sugiyama, A., Nishiya, Y., and Kawakami, B. 2009. RNA Polymerase Mutants with Increased Thermostability. U. S. Patent. US 7,507,567 B2.

Tawfik, D. S., and Griffiths, A. D. 1998. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* 16:652–656.

Urabe, H., Ichihashi, N., Matsuura, T., Hosoda, K., Kazuta, Y., Kita, H., and Yomo, T. 2010. Compartmentalization in a water-in-oil emulsion repressed the spontaneous amplification of RNA by Q beta replicase. *Biochemistry* 49:1809–1813.

Vallejos, M., Ramdohr, P., Valiente-Echeverría, F., Tapia, K., Rodriguez, F. E., Lowy, F., Huidobro-Toro, J. P., Dangerfield, J. A., and López-Lastra, M. 2010. The 5'-untranslated region of the mouse mammary tumor virus mRNA exhibits cap-independent translation initiation. *Nucleic Acids Research* 38:618–32.

Zielinski, W. S., and Orgel, L. E. 1987. Autocatalytic synthesis of a tetranucleotide analogue. *Nature* 327:346–347.

Chapter 3: Directed evolution of a panel of orthogonal T7 RNA polymerase:promoter pairs using compartmentalized partnered replication

The engineering and evolution of T7 RNA polymerase with altered promoter specificity has been a goal of protein engineers and synthetic biologists for decades. Rational engineering and directed evolution approaches have been utilized, but in all cases the resulting T7 RNA polymerases transcribe from their cognate promoters weakly or indiscriminately transcribe from multiple promoters. If T7 RNA polymerase mutants are to be useful in the development of complex transcriptional circuitry they will have to be both active and specific. This chapter details a novel approach to the directed evolution of promoter recognition by T7 RNA polymerase. This directed evolution platform, termed compartmentalized partnered replication (CPR), ties the *in vivo* recognition of a synthetic promoter by a T7 RNA polymerase mutant to the *in vitro* amplification of that mutant. The product of these efforts is a panel of six T7 RNA polymerase:promoter pairs that are both active and orthogonal to one another.

INTRODUCTION

Background and rationale

The need for orthogonal T7 RNA polymerase:promoter pairs

The desire to create alternate versions of T7 RNA polymerase stems from its inherent utility in a range of applications. T7 RNA polymerase's robust activity in prokaryotic, eukaryotic, and *in vitro* systems and its orthogonality to all host machinery have made it uniquely qualified to serve as the core of synthetic transcriptional circuits, both *in vivo* and *in vitro.* Examples of such circuits include the control of the nitrogen fixation cluster (Temme *et al*., 2012b) and pigment production *in vivo* (Wang *et al*., 2012;

Temme *et al*., 2012a) and nucleic acid computation *in vitro* (Kim *et al*., 2006; Kim and Winfree, 2011; Franco *et al*., 2011; Subsoontorn *et al*., 2012). As synthetic biological circuits become increasingly complex, it is desirable to precisely control modular synthetic transcriptional networks that function independently. This can be achieved by engineering a repertoire of T7 RNA polymerase variants that each specifically recognizes a synthetic promoter. Each orthogonal polymerase:promoter pair then acts as a transcriptional controller, analogous to bacterial sigma factors. Similarly to sigma factors, they enable the precise and flexible temporal, spatial, or signal dependent control of RNA expression.

The limits of previous approaches

Previous attempts to alter the molecular recognition of T7 RNA polymerase have been overviewed in **Chapter 1**. Single amino acid mutations that facilitate the recognition of mutant promoters have been found by screening (Raskin *et al*., 1993; Rong *et al*., 1998; Imburgio *et al*., 2000). These mutant promoters were only one base pair removed from the wild type T7 promoter; thus they were not divergent enough to prevent recognition of multiple promoters by a given polymerase.

Phylogenic part mining has been used to graft the specificity loop of related bacteriophage encoded RNA polymerase on the T7 RNA polymerase scaffold to allow it recognize chimeric promoters (Temme *et al*., 2012a). These promoters are sufficiently divergent to allow for the specific recognition of a single promoter. However, the grafted specifically loops were likely not optimally positioned in relation to the promoter, and thus the overall activity the polymerases were weak.

A large change in the polymerase is likely required in order to recognize a promoter divergent enough from the wild type promoter to allow for orthogonality.

However, such a large scale change in protein architecture cannot (currently) be made using phylogenic and structural considerations alone. Directed evolution is therefore uniquely able to achieve a change in promoter specificity sufficient to allow the specific recognition of a synthetic promoter.

Existing methods for directed evolution of T7 RNA polymerase tend to link the transcriptional output of a mutant to the fitness of its host. For example a T7 RNA polymerase mutant could drive expression of chloramphenicol acetyltransferase from a synthetic promoter and confer chloramphenicol resistance to its host (Ikeda *et al*., 1992). The T7 RNA polymerase would then be passively replicated based on its host's fitness. Similarly, phage assisted continuous evolution (PACE) ties expression of M13's gene III to the fitness of an M13 bacteriophage carrying a T7 RNA polymerase mutant (Esvelt *et al*., 2011; Dickinson *et al*., 2013). In each case, the evolutionary success of the T7 RNA polymerase gene is dependent not on its actual transcriptional output, but on the fitness of its host.

Tying directed evolution to host fitness is problematic for two reasons. First, there is usually a threshold of gene output required to confer maximum advantage to the host (*i.e.* a given number of chloramphenicol acetyltransferase proteins are required to allow for growth on a given concentration of chloramphenicol). Second, highly active polymerases that lead to the overexpression of a protein impart a costly metabolic load on their host. T7 RNA polymerase mutants may find an optimal strategy of mediocrity. Mediocre polymerases can produce just enough protein to survive the selection, while imposing minimal metabolic load. Thus, tying evolutionary success of a T7 RNA polymerase gene to the fitness of its host results in the evolution of polymerases of moderate activity. This makes intuitive sense, as organisms must strike a balance between activity and resource allocation.

Moderately active polymerases are a problem because they have relatively weak activity, but are even more problematic because they tend to be promiscuous. It has been shown that evolving enzymes go through a "generalist" phase in which they are moderately active, and after further evolution they become highly active specialists (**Figure 3-1**) (Matsumura and Ellington, 2001; Khersonsky and Tawfik, 2010). Selections limited by thresholding and host fitness effects may be unable to progress beyond the weak generalist phase, thus limiting their utility.

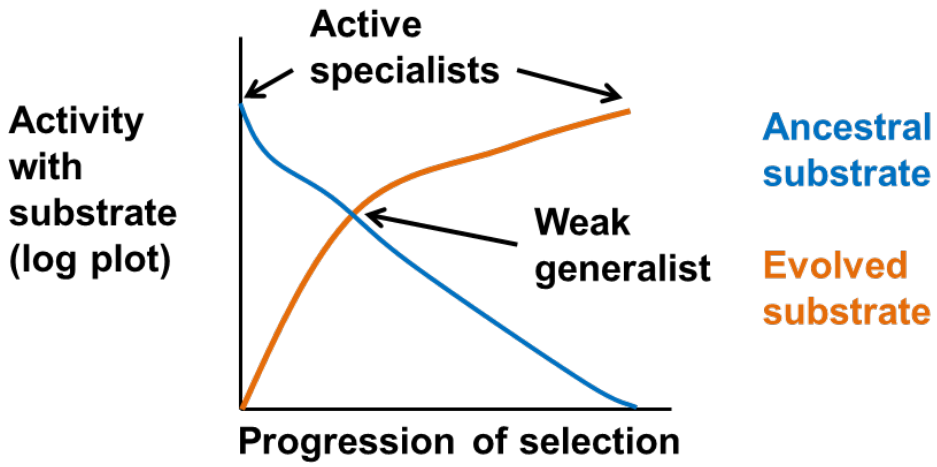


Figure 3-1. Activity and specialization.
In the evolution of a new substrate specificity, an enzyme may proceed through a "generalist" stage. Strong selective pressure on the new substrate pushes the substrate to become an active specialist. Selections with only moderate selective pressure often give rise to promiscuous enzymes of moderate active. This figure is based on data from the directed evolution of a hydrolase for new substrate specificity (Tokuriki *et al*., 2012).

In an attempt to decouple RNA polymerase activity from host fitness, the *in vivo* autogene couples T7 RNA polymerase activity to the transcription of its own gene; thus more active RNA polymerases are better represented in the total mRNA population (Chelliserrykattil *et al*., 2001). This system does not suffer the problem of thresholding, as there is no (theoretical) limit to the mRNA produced by a strong mutant. However, the problem of metabolic overtaxing of the host is magnified by the positive feedback loop.

The *in vitro* autogene removes consideration of the host organism and should therefore be ideally suited for the evolution of highly active polymerases (Davidson *et al*., 2012). However, as previously discussed, the mutation rate and the limitations of the *in vitro* transcription and translation systems doomed it in practice.

Design of the compartmentalized partnered replication platform for directed evolution

The problems associated with host fitness effects, thresholding effects, and excessive mutation rate have hampered previous efforts to engineer T7 RNA polymerase promoter recognition. The creation of an active and specific RNA polymerase should therefore involve a directed evolution scheme in which (1) selection is not reliant on host fitness, (2) increased performance scales with increased success over a wide range of fitness, and (3) the mutation rate can be controlled. Compartmentalized partnered replication (CPR, **Figure 3-2**) (Ellefson *et al*., 2014) is such a directed evolution scheme.

CPR couples the *in vivo* function of a protein, nucleic acid, operon, or circuit to its subsequent *in vitro* amplification. In the case of the directed evolution of T7 RNA polymerase promoter specificity, the *in vivo* function is the transcription of *Taq* DNA polymerase from a synthetic promoter. A library of T7 RNA polymerase mutants is transformed into *E. coli* cells that contain a plasmid in which the *Taq* DNA polymerase gene is directly downstream of the synthetic promoter. Cells that are transformed with T7 RNA polymerase mutants capable of driving expression from the synthetic promoter will produce *Taq* DNA polymerase protein upon induction, with the most active mutants producing the most DNA polymerase. Cells are then compartmentalized in a water-in-oil emulsion. Each compartment is expected to contain zero or one cell as well as buffer, dNTPs, and primers specific to the T7 RNA polymerase gene. Upon thermal cycling, cells are lysed, and their innards are released into the broader compartment. PCR

amplification occurs in compartments that contain *Taq* DNA polymerase protein, and these compartments also contain the gene encoding the T7 RNA polymerase mutant responsible for the expression of the *Taq* DNA polymerase protein. Thus, the active T7 RNA polymerase genes are amplified by emulsion PCR while inactive ones are not.

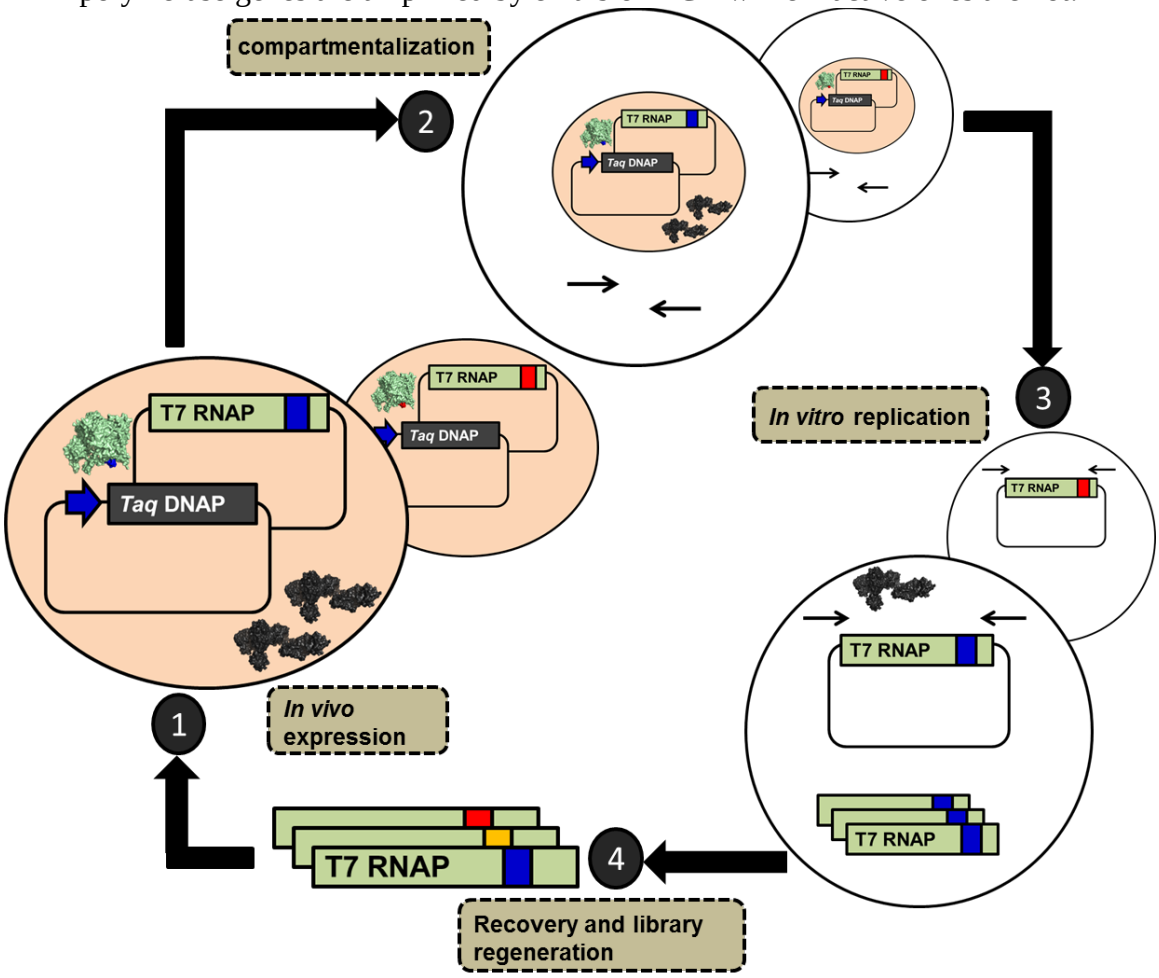


Figure 3-2. Compartmentalized partnered replication selection scheme.
(**1**) A library of T7 RNA polymerase mutants is transformed into *E. coli* cells containing the *Taq* DNA polymerase gene under the control of a synthetic T7 promoter variant (blue arrow). Variants capable of recognizing the promoter produce *Taq* DNA polymerase protein. (**2**) Whole cells are compartmentalized in a water-in-oil emulsion. The aqueous droplets also contain primers, dNTPs, and *Taq* DNA polymerase buffer. (**3**) Emulsions are thermal cycled, leading to cell lysis and PCR amplification of the T7 RNA polymerase genes that led to the production of *Taq* DNA polymerase during the *in vivo* step. (**4**) PCR product is recovered and prepared for the next round of selection.

RESULTS AND DISCUSSION

Implementation of the directed evolution of T7 RNA polymerase by CPR

Enrichment for active T7 RNA polymerase mutants

To demonstrate CPR as a suitable technique for the directed evolution of T7 RNA polymerase, three input "mock libraries" were created. The first library was composed entirely of wild type T7 RNA polymerase (WT), the second was entirely composed of inactive T7 RNA polymerase with two premature stop codons (XX), and the third comprised a 10:1 mixture of XX and WT (10:1). In each case, the T7 RNA polymerase open reading frame was under the control of the strong T5-lac promoter. All three libraries were transformed into a Bl21 *E. coli* strain in which the T7 promoter (P_{T7}) controlled *Taq* DNA polymerase expression. Each of the three transformed populations was subjected to selection by CPR, with the WT and 10:1 libraries each performed under two different induction conditions (1.0 mM IPTG and 0.1 mM IPTG). The primers used in the emulsion PCR step amplified roughly 500 base pairs, encompassing the majority of the fingers domain of T7 RNA polymerase. The DNA was recovered from the emulsion and a portion was run on an agarose gel (**Figure 3-3**). The two WT CPR recovery products can be easily seen, and there seems to be more product recovered when less IPTG is added. This is likely because the same volume of culture (not the same number of cells) was added to the emulsion and because overexpressing T7 RNA polymerase, which then overexpresses *Taq* DNA polymerase, is not ideal for cell growth. The 10:1 recovery products can only barely be seen.

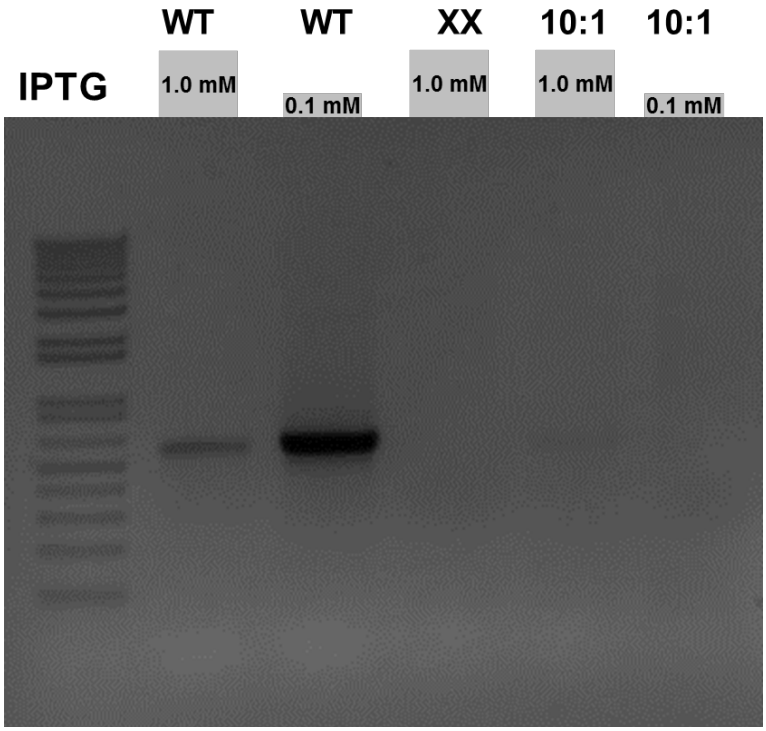


Figure 3-3. CPR mock selection recovery.
Mock libraries comprising either 100% wild type T7 RNA polymerase (WT), 100% inactive polymerase (XX) or 90% XX and 10% WT (10:1) were subject to a single round of CPR. The products of each emulsion PCR are shown.

The column-purified recovery products were also subject to reamplification with the same primers used in the emulsion PCR. These products were analyzed by agarose gel electrophoresis (**Figure 3-4**). The reamplification products for each reaction (even the negative controls) can be seen easily. The most pronounced recovery product bands (**Figure 3-3**) give rise to the most pronounced reamplification product bands (**Figure 3-4**) suggesting that selection for active polymerases is possible. The substantial reamplification product of the truncated T7 RNA polymerase genes suggests that parental plasmid is amplified in the reamplification PCR. This problem is addressed in subsequent implementations of the selection protocol.

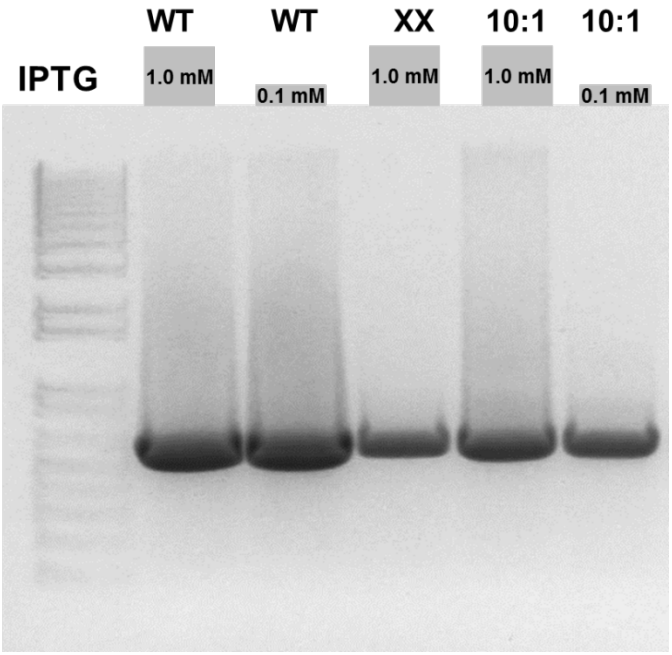


Figure 3-4. CPR mock selection reamplification.
The recovery products seen in **Figure 3-3** were reamplified using the same primers used in the emulsion PCR.

In order to determine the enrichment capability of the CPR selection, the reamplification products were digested with EcoRV. The amplicon from the WT sequences does not contain an EcoRV recognition site, but the XX amplicon does. Agarose gel electrophoresis of the digested reamplification product of the 10:1 population can therefore be used to estimate an enrichment factor (**Figure 3-5**). It appears that roughly 90% of the digestion product corresponds to the wild type sequence. WT's fecundity (f_{WT}) would need to be about 90 times higher than XX's fecundity (f_{XX}) in order to increase from the 10% of the input to 90% of the output in a single round:

$$1 * f_{WT} / (10 * f_{XX} + 1* f_{WT}) = 0.90 \text{ or}$$

$$f_{WT} = 90*f_{XX}$$

This implies a roughly 90-fold selective advantage of WT over XX.

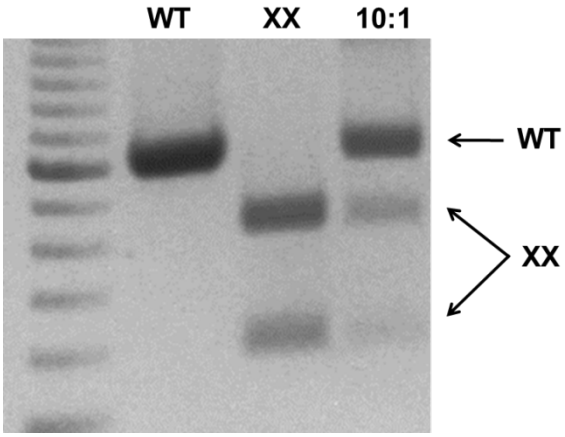


Figure 3-5. CPR mock selection digestion.
The reamplification products seen in **Figure 3-4** were digested with EcoRV. The wild type T7 RNA polymerase (WT) sequence is undigested while the inactive T7 RNA polymerase (XX) sequence is cut into two distinct fragments. The ratio of the three bands can be used to estimate an enrichment factor for the selection.

The successful enrichment of WT at the expense of XX suggests that the CPR selection method could be used for the evolution of T7 RNA polymerase. Lessons learned in this experiment were applied to subsequent selections. First, the induction strength chosen was 0.1 mM IPTG, which should reduce metabolic load and prevent saturation of *Taq* DNA polymerase production. Second, using the same primers in the emulsion PCR and reamplification PCR steps results in the amplification of the parental plasmid in the reamplification PCR. There are at least two possible solutions to this problem. In one strategy, the emulsion PCR primers can append an extraneous sequence tag to each end of the amplicon and this tag can serve as the primer binding site in the reamplification PCR. This strategy is employed in **Chapter 5**. A second strategy, used in the remainder of this chapter, utilizes biotinylated primers in the emulsion PCR step and purification of the PCR product (away from the parental plasmid) using ferromagnetic beads coated with streptavidin before the reamplification. Both strategies have proven to be effective.

Library design

T7 RNA polymerase makes three distinct contacts with the promoter, but the primary determinants of promoter specificity are made by the specificity loop (residues 742 to 761) (Cheetham *et al*., 1999). Hydrogen bonding between residues R746, N748, R756, and Q758 and the major groove from -7 to -11 of the promoter are especially critical (Raskin *et al*., 1993; Rong *et al*., 1998; Imburgio *et al*., 2000). The critical R746, N748, R756, and Q758 residues as well as the intervening L747 and L757 were therefore randomized.

The specificity loop of T7 RNA polymerase with reduced secondary structure (T7 RSS) was generated by PCR using a long oligodeoxynucleotide as the template. This long oligodeoxynucleotide was synthesized in-house by Randall Hughes of the Ellington Lab on an Expedite 8900 synthesizer at a 40 nmol synthesis scale. Degeneracy was introduced into the oligonucleotides at position 746 to 748 and 756 to 758 via the use of trimer phoshoramidites containing a mixture of 20 trimer (codon) phosphoramidites encoding all 20 amino acids.

The degenerate specificity loop was used in an overlap PCR to generate the full length T7 RSS open reading frame. The T7 RSS library was digested and ligated into the pQE backbone (in which a strong T5-lac promoter drives T7 RNA polymerase) (Ellefson *et al*., 2014). This library serves as the basis for several selections detailed in this chapter and is referred to as Codon Randomized Specificity Loop Round 0 (CRSL R0).

Selection for the use of the wild type T7 promoter

In order to test the capability of CPR to evolve promoter recognition, a genetic circuit was constructed in which transcription of *Taq* DNA polymerase relies on T7 RNA polymerase binding to and activating the wild type T7 promoter (P_{T7}). CRSL R0 was transformed into a strain containing this plasmid and four iterative rounds of selection by

CPR were performed. The primers used in the emulsion PCR amplified roughly 500 base pairs. As noted above, these primers were biotinylated, facilitating clean up after the emulsion was broken. The primers used in the reamplification PCR yielded only a small amplicon, encompassing the randomized regions (**Figure 3-6**).

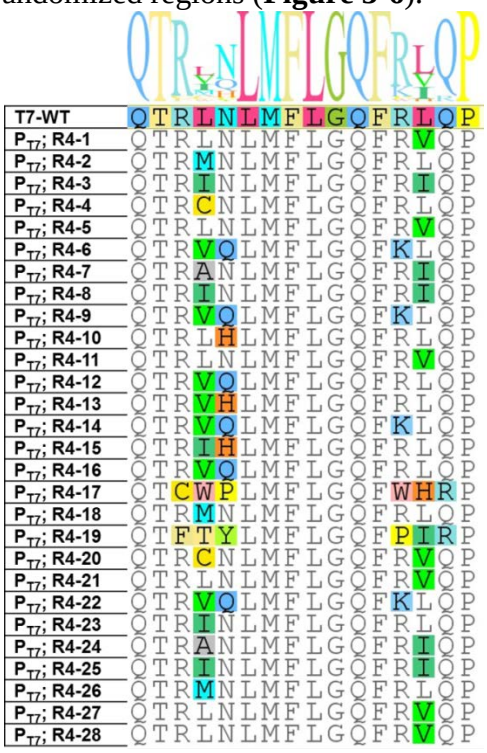


Figure 3-6. Sequence analysis of the T7 RNA polymerase specificity loop selection for the use of P_{T7}.

28 variants from Round 4 of the selection for the use of P_{T7} were sequenced. Mismatches from the wild type sequence (shown near the top) are indicated by colored blocks. The position weight matrix of the population is shown at the top.

28 clones from Round 4 were sequenced and compared to the wild type RNA polymerase sequence (**Figure 3-6**). While no perfect match to the wild type sequence was selected and significant diversity remained in the pool, the consensus sequence clearly converged on the wild type RLN…RLQ motif (R746, L747, N748 … R756, L757, Q758). Clones 17 and 19 are likely inactive, while the other 26 clones are likely

quite active. It should be noted that considerable diversity was retained at the non-critical positions L747 and L757.

The pooled library from each round was transformed into a strain in which GFP expression is driven from P_{T7} (**Figure 3-7, top**). Cultures were induced and images of GFP emission were taken with a digital camera (**Figure 3-7, bottom**). The Round 0 population appeared to be roughly as active as the inactive T7 RNA polymerase (XX). Each successive round increased in activity, with the Round 4 population having comparable activity to the wild type T7 RNA polymerase (WT).

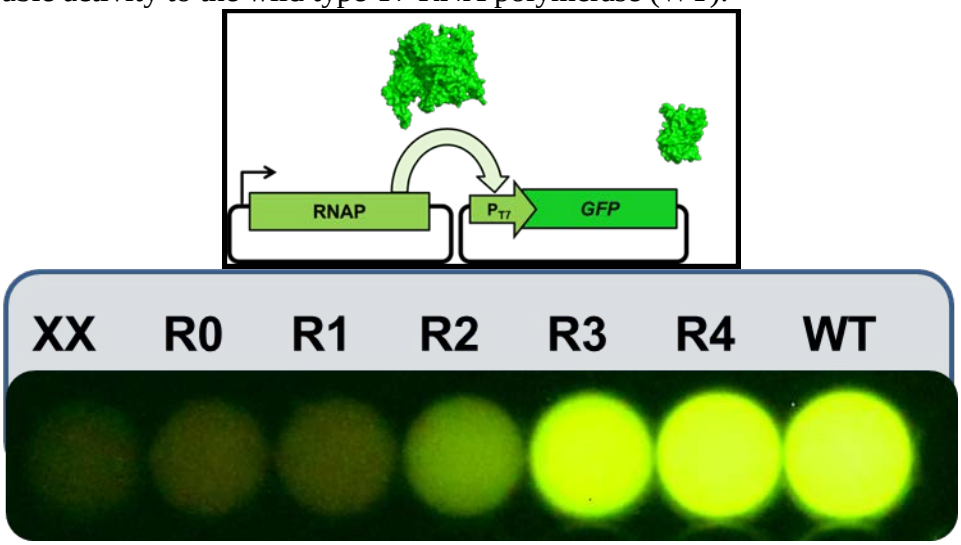


Figure 3-7. Activity assay of the T7 RNA polymerase specificity loop selection for the use of P_{T7}.
The library from each round, as well as an inactive T7 RNA polymerase (XX) and wild type T7 RNA polymerase (WT) were transformed into a P_{T7}-GFP reporter strain. GFP output is a measure of activity of the round.

Selection for the use of a synthetic T7 promoter by CPR

Selection for the use of P_{CGG}

The success of the selection for recognition of the wild type promoter suggested a selection for the use of a synthetic promoter was possible. A completely synthetic

promoter was designed with the following sequence: TAATACCGGTCACTATA. This promoter, named P_{CGG}, contains the sequence CGGT from -11 to -8. This was decided upon because it quite different from P_{T7} (GACT) as well as the promoter of the related bacteriophages SP6 (GTGA), T3 (CCCT), K11 (GGCA), K1F (CTAT), and N4 (CCCA). It should be noted that -11 is G or C and -8 is A or T in nearly all bacteriophage promoters known (Rong *et al.*, 1998; Temme *et al.*, 2012a). It is unclear if there is an evolutionary basis for this, or if it is merely a coincidence. However, A and T were avoided at the -11 position and G and C were avoided at the -8 position.

P_{CGG} was cloned upstream of *Taq* DNA polymerase such that T7 RNA polymerase mutants that can recognize this new promoter would be preferentially amplified in the CPR reaction. The Codon Randomized Specificity Loop Round 0 (CRSL R0) was transformed into a strain containing the P_{CGG}-*Taq* plasmid and rounds of selection using CPR were performed as above. After seven rounds of CPR, a single sequence "RVH…EMQ" dominated the population (**Figure 3-8, top**). A single clone bearing this motif (CGG-R7-8) was analyzed. The library from each round and CGG-R7-8 were transformed into cells containing a plasmid in which GFP is driven from P_{CGG}. The ability to drive GFP from P_{CGG} was quantified using a Tecan Safire monochromator and compared to the fluorescence output of wild type T7 RNA polymerase driving GFP from P_{T7}. The activity of the CGG-R7-8:P_{CGG} pair was equivalent to only about 2% of the activity of wild type T7 RNA polymerase:P_{T7} (WT:P_{T7}) pair (**Figure 3-8, bottom left**). In an assay for cross-reactivity, only background levels of fluorescence for the WT:P_{CGG} and CGG-R7-8:P_{T7} pairs were seen (**Figure 3-8, bottom right**).

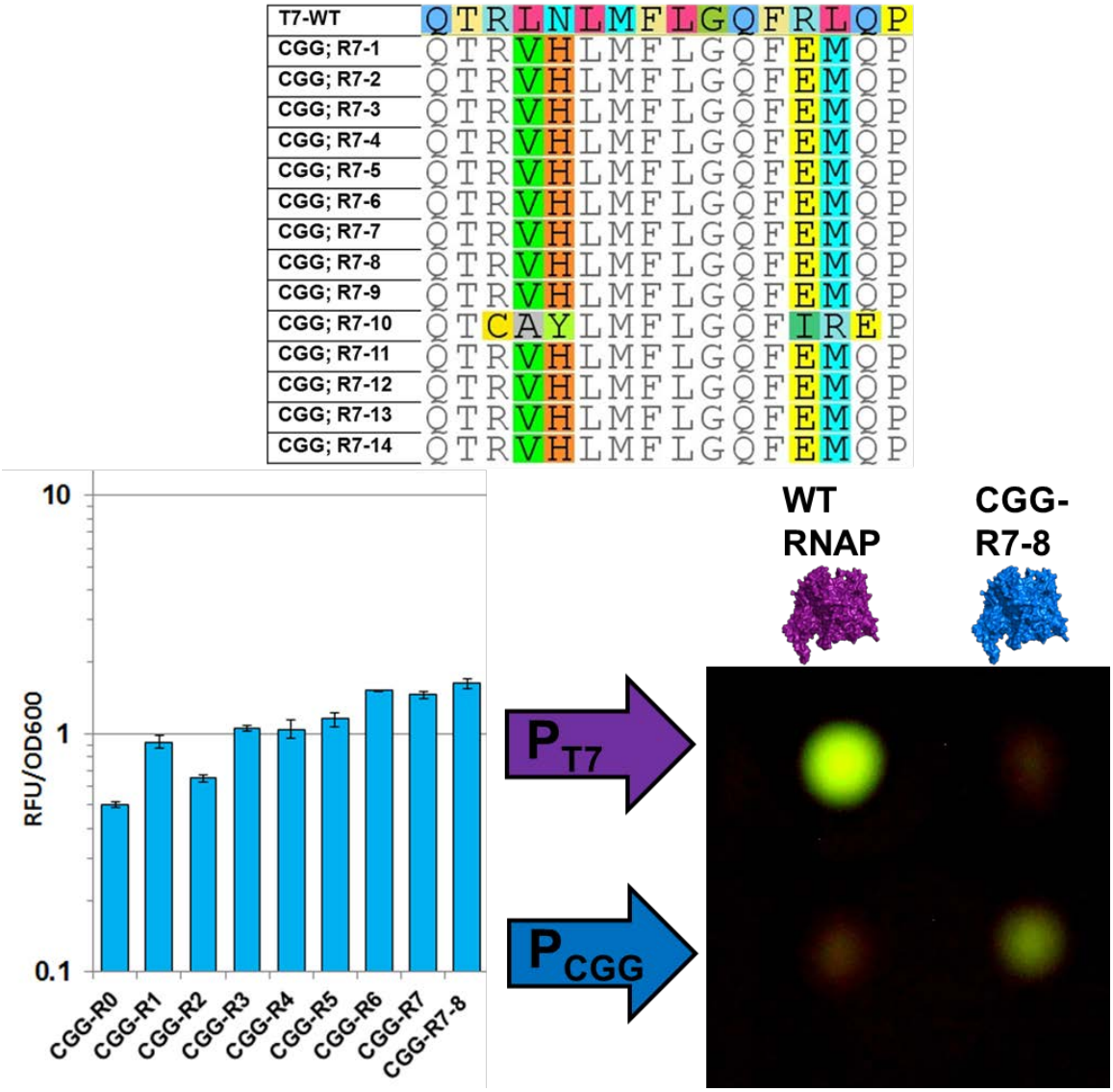


Figure 3-8. Analysis of Round 7 of the T7 RNA polymerase specificity loop selection for the use of P_{CGG}.

Top) 14 variants from Round 7 of the selection for the use of P_{CGG} were sequenced. Mismatches from the wild type sequence (shown at the top) are indicated by colored blocks. **Bottom left**) The library from each round of the P_{CGG} selection, as well as CGG-R7-8 were transformed into a P_{CGG}-GFP reporter strain. GFP output / OD600 is a measure of activity of the round, relative to the activity of the WT:P_{T7} pair. Values are the average from three independent cultures; error bars represent standard error. **Bottom right**) Wild type T7 RNA polymerase and CGG-R7-8 were transformed into P_{CGG}-GFP and P_{T7}-GFP reporter lines. Cells were grown, induced, and imaged to obtain a visual depiction of activity and cross-reactivity.

Maturation of CGG-R7-8

The modest activity of CGG-R7-8 suggested that additional T7 RNA polymerase mutations, not present in the initial library, were needed to allow the polymerase to recognize the synthetic promoter sequence. To further improve the activity of the CGG-R7-8 polymerase on the P_{CGG} promoter, an additional five rounds of selection by CPR were performed. In these rounds the reamplification PCR was performed with primers that yielded a larger amplicon (about 500 base pairs) flanking the specificity loop and encompassing much of the fingers domain. This allowed the entire promoter specificity domain to evolve, instead of being held constant. Additionally this larger region of was subjected to random mutagenesis via error-prone PCR before the first and fourth additional rounds.

The Round 12 population was transformed into the P_{CGG}-GFP reporter strain and the activity (GFP fluorescence / OD600) of 93 clones were quantified. The eight T7 RNA polymerase variants that drove the highest GFP expression from P_{CGG} were sequenced (**Figure 3-9**).

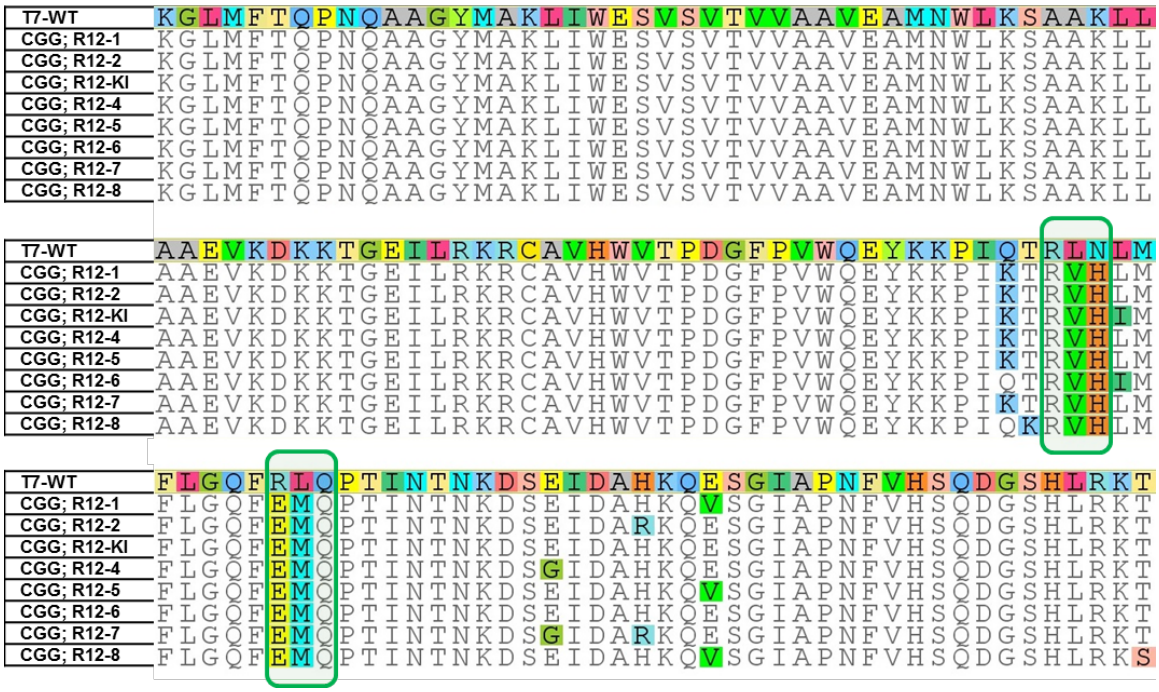


Figure 3-9. Sequence analysis of Round 12 of the T7 RNA polymerase specificity loop selection for the use of P_{CGG}.

The library from Round 12 of the selection for the use of P_{CGG} was transformed into a P_{CGG}-GFP reporter strain. Individual clones were measured for GFP fluorescence, and the eight most fluorescent variants were sequenced. The "RVH…EMQ" motif remained invariant. Clone 3 was the most active, and was renamed CGG-R12-KI. The mutations Q744K ("K"), L749I ("I"), E768G ("G"), H772R ("R"), and E775V ("V") appeared in multiple clones.

The most active T7 RNA polymerase variant from this population, CGG-R12-KI, displayed 20% the activity of the WT:P_{T7} pair (**Figure 3-10**). Additional mutations beyond those seen in CGG-R12-KI were also frequently observed in variants in the Round 12 population (**Figure 3-9**). Specifically, the mutations Q744K ("K"), L749I ("I"), E768G ("G"), H772R ("R"), and E775V ("V") appeared in multiple variants. Combinations of these frequently occurring mutations were embedded into the CGG-R7-8 sequence. The resulting mutants were assayed for activity in the P_{CGG}-GFP strain and several mutants showed even further enhancement: 40-60% activity of the WT:P_{T7} pair.

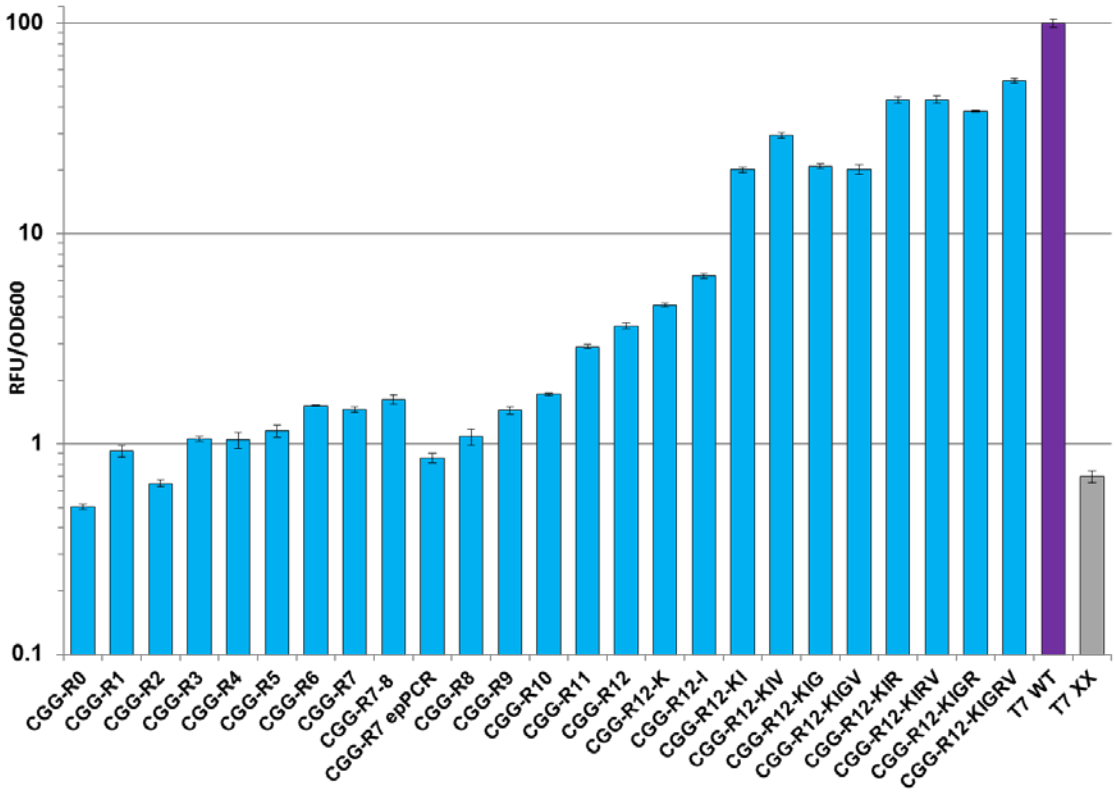


Figure 3-10. *In vivo* activity assay for the use of P_{CGG}.
The library from each round of the selection for the use of P_{CGG}, as well as various described mutants were transformed into a P_{CGG}-GFP reporter strain. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The five most active variants (CGG-R12-KIV, CGG-R12-KIR, CGG-R12-KIRV, CGG-R12-KIGR, and CGG-R12-KIGRV) were mixed in equal ratios and selected by CPR for an additional four rounds (with error-prone PCR before the first, third, and fourth round). The resulting Round 16 population closely resembled the mutant CGG-R12-KIRV, with no additional consensus mutations. The selection was therefore judged to be complete.

Characterization of CGG-R12-KIRV

In order to judge the success of the selection, CGG-R12-KIRV was tested for activity and cross-reactivity *in vivo*. A plasmid expressing the mutant enzyme was transformed into a P_{CGG}-GFP reporter strain as well as a P_{T7}–GFP reporter strain (**Figure 3-11**). The CGG-R12-KIRV:P_{CGG} pair demonstrated roughly 30% activity (GFP fluorescence/OD600) as compared to the WT:P_{T7} pair. In terms of cross-reactivity, the CGG-R12-KIRV:P_{T7} and WT:P_{CGG} pairs demonstrated only about 1% activity.

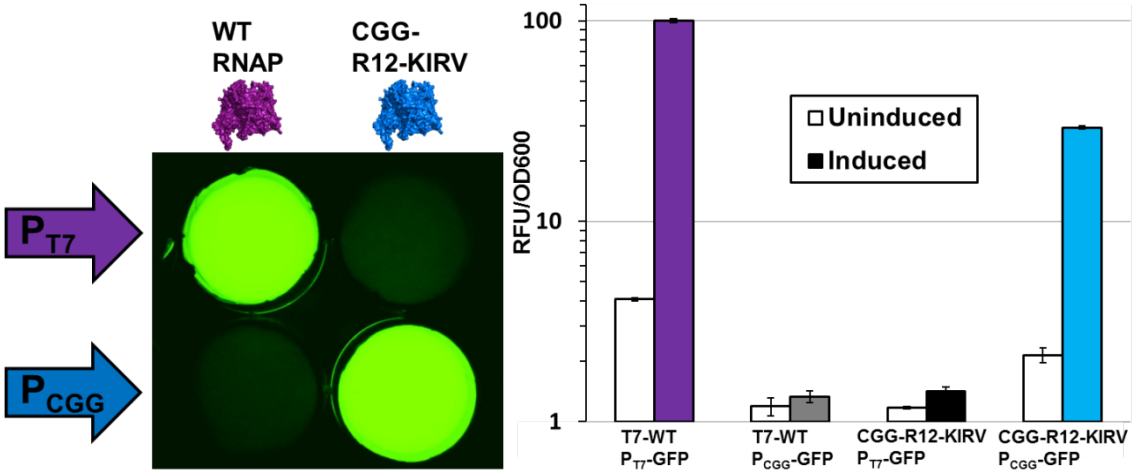


Figure 3-11. ***In vivo*** **activity and cross-reactivity assay of CGG-R12-KIRV.** Wild type T7 RNA polymerase and CGG-R12-KIRV were transformed into P_{CGG}-GFP and P_{T7}-GFP reporter lines. **Left**) Cells were grown, induced, and imaged to obtain a visual depiction of activity and cross-reactivity. **Right**) GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

Both the wild type T7 RNA polymerase and CGG-R12-KIRV were expressed and purified. The enzymes were assayed *in vitro* for transcriptional activity on both P_{T7} and P_{CGG}. For each template, the promoter was followed by the fluorescent aptamer Spinach (Paige *et al.*, 2011). Transcriptions were performed on a Tecan Safire monochromator and fluorescence readings were taken every minute. Fluorescent output is used as a real-time indication of transcriptional activity. This assay has proven to be a facile and

reliable method to judge RNA polymerase activity. The real-time measurements prevent signal saturation, which can be a limitation with traditional end-point measurements. The results of this *in vitro* assay are comparable to the *in vivo* data, with CGG-R12-KIRV showing roughly 60% on-target activity less than 0.1% cross-reactivity (**Figure 3-12**). The reduction in apparent off-target activity *in vitro* (as compared to *in vivo* cross-reactivity) suggests that the measured *in vivo* cross-reactivity is on par with the background fluorescence of *E. coli*.

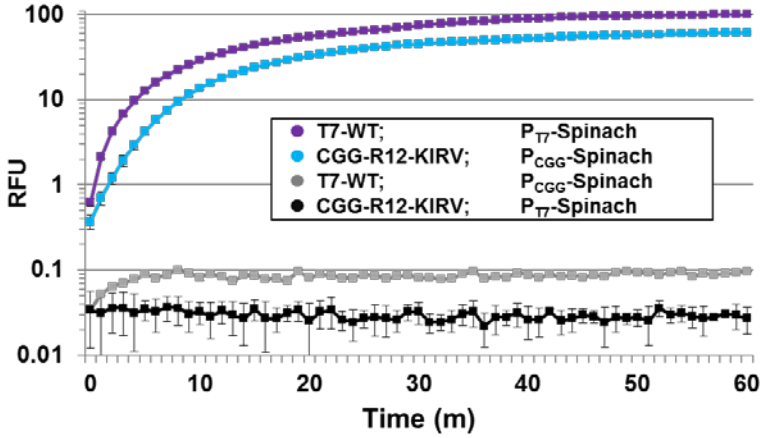


Figure 3-12. *In vitro* activity and cross-reactivity assay of CGG-R12-KIRV.
Wild type T7 RNA polymerase and CGG-R12-KIRV were used in an *in vitro* transcription assay using P_{T7} and P_{CGG}-driven expression of the Spinach aptamer as a readout. Spinach fluorescence (in Relative Fluorescence Units) is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent reactions; error bars represent standard error.

Comparison of CGG-R12-KIRV to other T7 RNA polymerase promoter specificity mutants

In order to demonstrate the quality of CGG-R12-KIRV and the utility of CPR as a method for the directed evolution of protein function, CGG-R12-KIRV was compared to the engineered polymerases from the literature reported to have the highest activity and

specificity. The comparable polymerases' method of generation, sequence, preferred promoter, and reported activity and specificity are summarized in **Table 3-1**.

Method (citation)	(Name of mutant) mutations	Mutant promoter -12 to -8 WT: CGACT	Reported activity	Reported orthogonality (vs P_{T7})
PACE (Dickinson *et al.*, 2013)	(T3-192-2-3) D130N H176R E222K V334I M439V G555C E643K N748D Q758R E775A V783I	GCCGA* (P_{final})	30	4
	(T3-192-3-14) E222K T243I G542V Q656H S686N N748D Q758R E775V	GCCGA* (P_{final})	20	4
	(SP6-192-3-8) A69T E222K A381V F646L V685A R756C Q758K A827T	GCCGA* (P_{final})	150	1.0
	(SP6-192-4-2) M54V D66N E222K Y385H N433S C510R A575V E683D V685A A724V R756C Q758R H772R	GCCGA* (P_{final})	300	1.0
PACE (Esvelt *et al.*, 2011)	(L1-108.1) I4M G175R E222K Q239R G542V N748D	ACCCT* (P_{ACCCT})	700 **125**	< 1.0 **1.5**
	(L1-192.1) I4M E63V S128R G175R E222K Q239R G542V N748D	ACCCT* (P_{ACCCT})	800 **125**	< 1.0 **1.7**
	(L2-192.1) I4M D66Y G175R E222K G280C G542V N748D E775K L864F	ACCCT* (P_{ACCCT})	600 **85**	1.1 **1.5**
Rational screen (Raskin *et al.*, 1993)	N748D	CCCCT (P_{CCCT})	**15**	**8**
Rational screen (Imburgio *et al.*, 2000; Shis and Bennett, 2013)	R756S	CACAT (P_{ACAT})	90	6
	R756K	CACAT (P_{ACAT})	130	1.1
Rational screen (Rong *et al.*, 1998)	Q758C	CGACG (P_{GACG})	**71**	**7**
	Q758S	CGACA (P_{GACA})	**53**	**7**
	Q758R	CGACC (P_{GACC})	**31**	**30**
Domain grafting (Temme *et al.*, 2012a)	(T7(T3)) T745K N748D L749M M750I	ACCCT (P_{T3})	50	8
	(T7(K1F)) Q754S R756N I761V	ACTAT (P_{K1F})	10	5
	(T7(N4)) L747I N748D L749C M750V F751I Q754T F755H L757M Q758A P759L	ACCCA (P_{N4})	150	50
CPR (Ellefson *et al.*, 2014)	(CGG-R12-KIRV) Q744K L747V N748H L749I R756E L757M H772R E775V	CCGGT (P_{CGG})	30 **61**	21 **>1000**

Table 3-1. Comparison of T7 RNA polymerase promoter specificity mutants.
Underlined sequences denote deviation from the wild type CGACT promoter motif. The * denotes that the promoter varies from P_{T7} outside of the -12 to -8 range. Specifically, "GCCGA*" is AATTAGCCGACACTAAAGAAGAA and "ACCCT*" is AATTAACCCTCACTAAAGGGAGA from -17 to +6. Mutant polymerase activity refers to the activity of the mutant polymerase on the mutant promoter (relative to the wild type pair defined as 100). Mutant orthogonality refers to the ratio of the mutant activity on its cognate promoter to its activity on the wild type promoter. Bold data are from *in vitro* assays.

Based on the reported characteristics of previously evolved or designed T7 RNA polymerase promoter specificity mutants, CGG-R12-KIRV compares well in terms of specificity but not in terms of activity. However, since the methods used to characterize the polymerases are not consistent with each other or with the above characterizations, it was necessary to provide a side by side comparison of all the listed polymerases to CGG-R12-KIRV.

First, an *in vivo* activity and cross-reactivity assay was conducted. Each polymerase was transformed into its own specific P_{cog}-GFP strain (where P_{cog}, is the cognate promoter for each polymerase) as well as the P_{T7}-GFP reporter strain. Wild type T7 RNA polymerase (WT) was also transformed into each reporter line. The activity of each polymerase with its cognate promoter, WT's activity with the same promoter, the mutant activity with P_{T7} and the activity of the WT:P_{T7} pair is visualized and quantified with GFP fluorescence (**Figure 3-13 and Figure 3-14**).

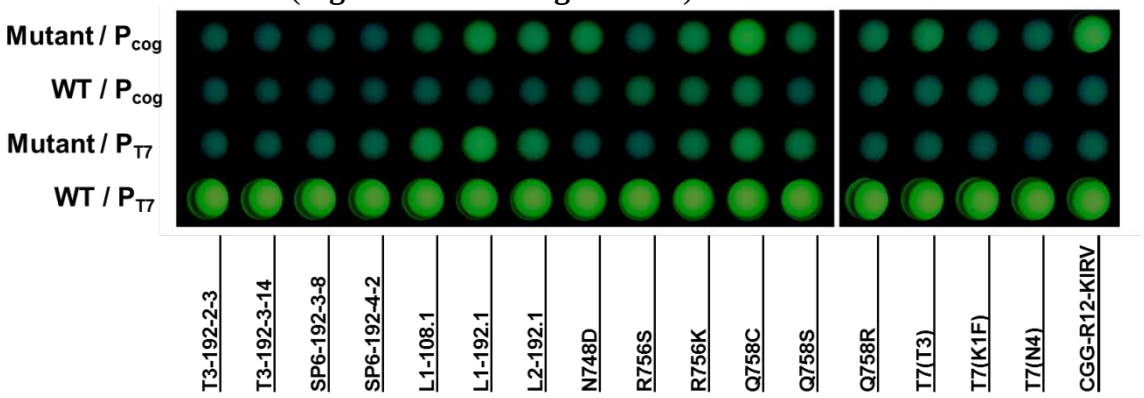


Figure 3-13. *In vivo* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants.
Wild type T7 RNA polymerase and a range of mutants were transformed into several GFP reporter lines. Cells were grown, induced, and imaged to obtain a visual depiction of activity and cross-reactivity.

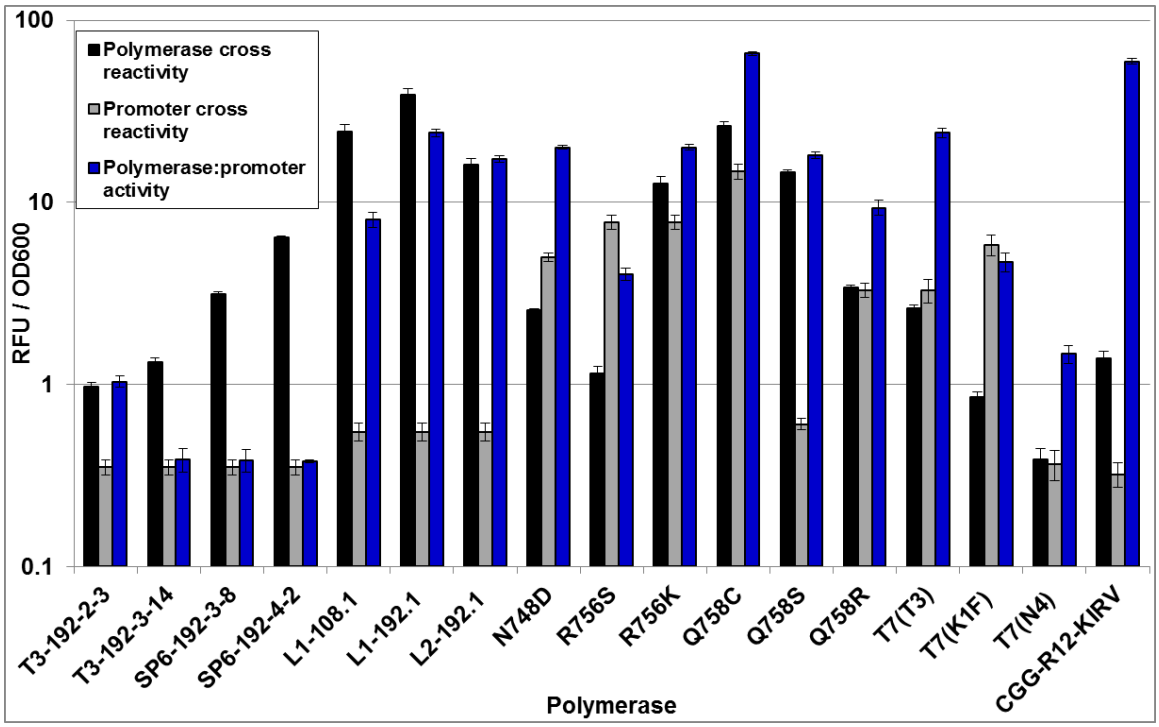


Figure 3-14. Quantitative i*n vivo* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants.

Wild type T7 RNA polymerase and a range of mutants were transformed into several GFP reporter lines to quantify activity and cross-reactivity. "Polymerase cross reactivity" (black) refers to the activity of the mutant T7 RNA polymerase on the wild type promoter. "Promoter cross reactivity" (grey) refers to the activity of wild type T7 RNA polymerase on the mutant promoter. "Polymerase:promoter activity" (blue) refers to the activity of the mutant polymerase on the mutant promoter. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

Assaying reported polymerases using consistent parameters revealed that CGG-R12-KIRV is one of the most active RNA polymerases described, with only Q758C being comparable. It should be noted that several polymerases in the literature with activities reported to be greater than wild type are either inactive (SP6-192-3-8 and SP6-192-4-2) or only weakly active (L1-108.1 and T7(N4)). It is likely that the discrepancies between reported and observed activity are due to toxicity. That is, polymerases are

assayed by being overexpressed and subsequently overexpressing a reporter. Apparent activity can be obscured due to a very active polymerase (such as the wild type pair) appearing inactive because it overtaxes its host, leading to mutation, cell sickness, or death (Shis and Bennett, 2013). In some cases, the promoter and RBS driving the polymerase was allowed to evolve and was retained during the assay. Therefore the expression of each polymerase was unequal, thus skewing the observed activity of each polymerase (Dickinson *et al*., 2013). In the case of T7(T3), T7(K1F), and T7(N4) (which had been generated by domain grafting) the RBS and active site of wild type T7 RNA polymerase were mutated before the assay, inviting a false comparison between mutant and wild type (Temme *et al*., 2012a).

To further demonstrate orthogonality, the most active polymerases from the GFP assays were transformed into a cell line in which PheS A294G is under the control of the T7 promoter. PheS A294G is a promiscuous mutant of the phenylalanine aminoacyl-tRNA synthetase (Kast and Hennecke, 1991; Thyer *et al*., 2013). Alone, PheS A294G has no observable effect on *E. coli*. However, in the presence of the unnatural amino acid 4-chloro-DL-phenylalanine (Cl-Phe), PheS A294G charges phenylalanine tRNAs with Cl-Phe, leading to random translational errors of the proteome and subsequently cell death. Cultures were grown, induced, and dilution-plated on a range of Cl-Phe concentrations (**Figure 3-15**). Cross-reactivity can be judged by the dose dependent cytotoxicity of Cl-Phe. CGG-R12-KIRV forms the same number of colonies at 20 mM Cl-Phe as with no Cl-Phe. All other mutants show a marked decrease in colony formation at 5 to 15 mM Cl-Phe. The data from this cell toxicity assay are in fine agreement with the GFP assay and indicate that CGG-R12-KIRV is much more orthogonal to P_{T7} than any other tested T7 RNA polymerase specificity mutant.

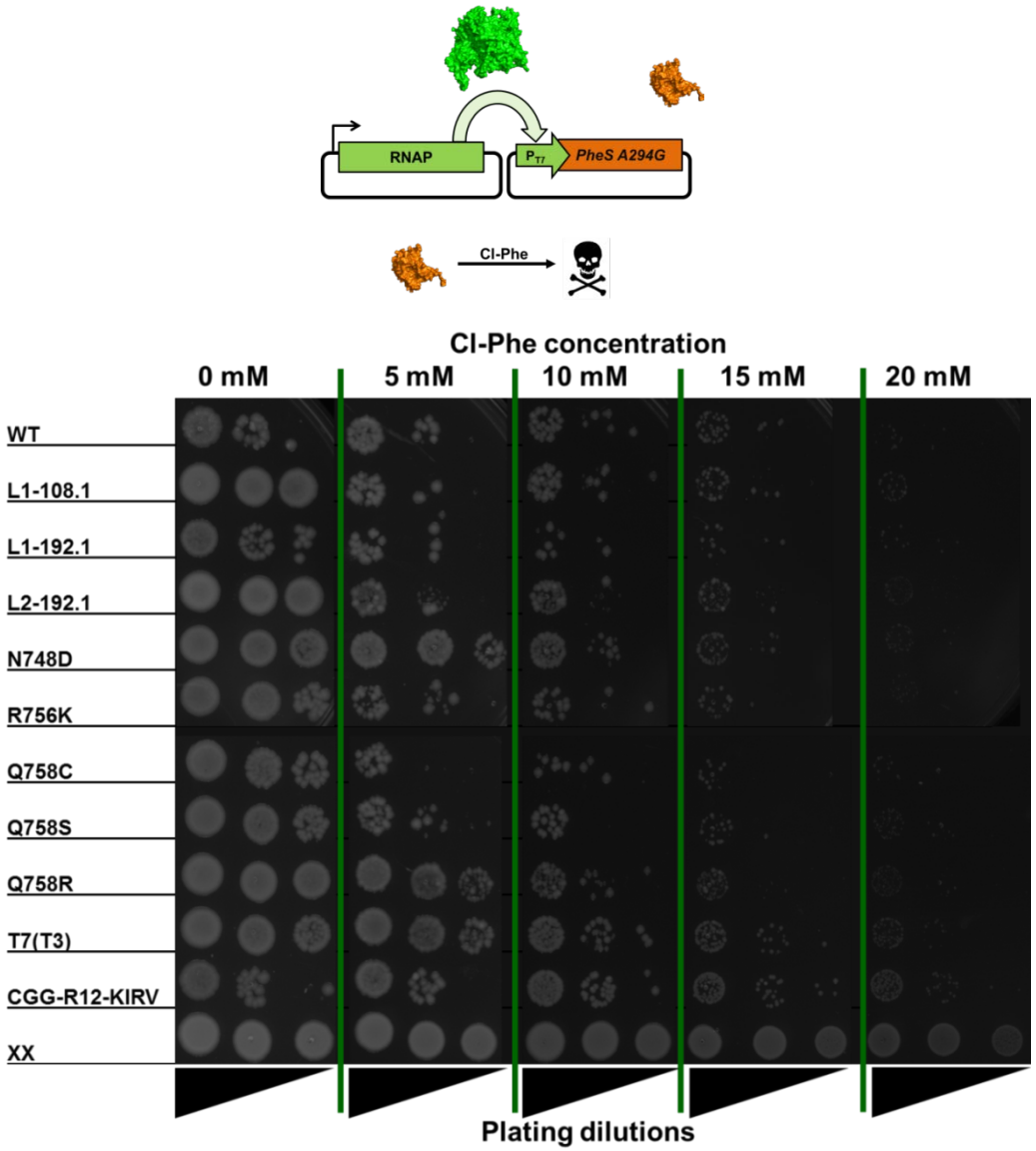


Figure 3-15. *In vivo* orthogonality of T7 RNA polymerase promoter specificity mutants.

Wild type T7 RNA polymerase and a range of mutants were transformed into a reporter line in which PheS A294G, is driven by P_{T7}. Cultures were induced with 0.05 mM IPTG before being diluted to an OD600 of 0.1, 0.01, and 0.001. Dilutions were plated on media containing 0, 5, 10, 15, or 20 mM Cl-Phe. RNA polymerase variants driving PheS A294G from P_{T7} are expected to die in the presence of Cl-Phe.

As a final measure of activity and specificity, each previously published polymerase was expressed, purified, and assayed *in vitro*. Each polymerase was allowed

to transcribe the fluorescent aptamer Spinach from its cognate promoter or P_{T7}. Wild type T7 RNA polymerase was also allowed to transcribe Spinach from each promoter. Measurements of fluorescence were taken every two minutes. The fluorescent reading for each polymerase: polymerase promoter pair was analyzed after 30 minutes (enough time to overcome the initial background fluorescence, but prior to signal saturation) (**Figure 3-16**). The trends are consistent with the *in vivo* assays, with CGG-R12-KIRV being the most specific and among the most active polymerases (**Table 3-2**). Taken together, these data suggest that CGG-R12-KIRV is the most successful attempt at engineering T7 RNA polymerase promoter specificity to date. More broadly, the successful evolution of a highly active and specific T7 RNA polymerase promoter specificity mutant highlights the usefulness of CPR as a method for the evolution of enzyme activity and specificity.

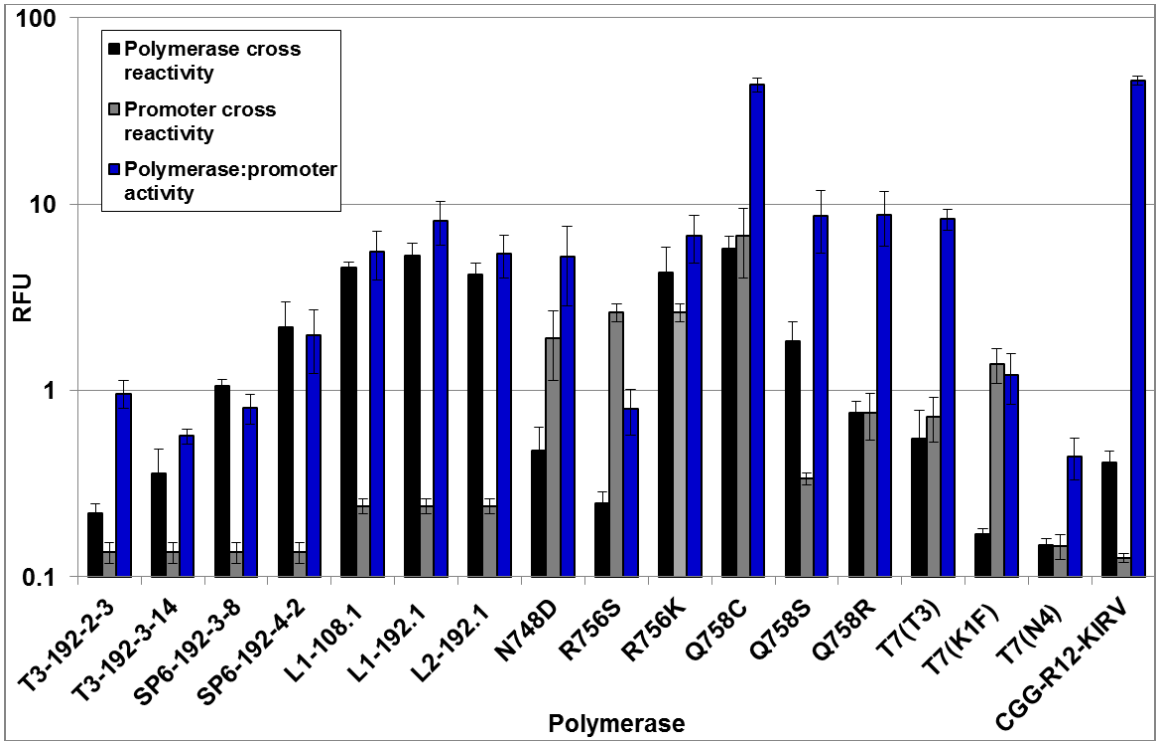


Figure 3-16. *In vitro* activity and cross-reactivity assay of T7 RNA polymerase promoter specificity mutants.

Wild type T7 RNA polymerase and a range of mutants were used in an *in vitro* transcription assay. P_{T7} or mutant promoter driven expression of the Spinach aptamer was used as a readout to quantify activity and cross-reactivity. "Polymerase cross reactivity" (black) refers to the activity of the mutant T7 RNA polymerase on the wild type promoter. "Promoter cross reactivity" (grey) refers to the activity of wild type T7 RNA polymerase on the mutant promoter. "Polymerase:promoter activity" (blue) refers to the activity of the mutant polymerase on the mutant promoter. Spinach fluorescence (in Relative Fluorescence Units) is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent reactions; error bars represent standard error.

Name of mutant	Reported activity	Measured activity	Reported orthogonality (vs P_{T7})	Measured orthogonality (vs P_{T7})
T3-192-2-3	30	1.0	4	1.1
		1.0		**4.4**
T3-192-3-14	20	<1	4	<1
		<1		**1.6**
SP6-192-3-8	150	<1	1.0	<1
		<1		**<1**
(SP6-192-4-2	300	<1	1.0	<1
		2.0		**<1**
L1-108.1	700	8.0	< 1.0	<1
	125	**5.6**	**1.5**	**1.2**
L1-192.1	800	24.1	< 1.0	<1
	125	**8.1**	**1.7**	**1.6**
L2-192.1	600	17.3	1.1	1.1
	85	**5.4**	**1.5**	**1.3**
N748D		20.0		7.8
	15	**5.2**	**8**	**11.0**
R756S	90	4.0	6	3.5
		<1		**3.2**
R756K	130	20.0	1.1	1.6
		6.8		**1.6**
Q758C		65.5		2.5
	71	**43.8**	**7**	**7.6**
Q758S		18.2		1.2
	53	**8.6**	**7**	**4.7**
Q758R		9.3		2.8
	31	**8.8**	**30**	**11.6**
T7(T3)	50	24.0	8	9.2
		8.3		**15.0**
T7(K1F)	10	4.7	5	5.5
		1.2		**7.1**
(T7(N4)	150	1.5	50	3.8
		<1		**3.0**
CGG-R12-KIRV	30	59.4	21	42
	61	**46.1**	**>100**	**>100**

Table 3-2. Comparison of T7 RNA polymerase promoter specificity mutants.
"Activity" refers to the activity of the mutant polymerase on the mutant promoter (relative to the wild type pair defined as 100). "Orthogonality" refers to the ratio of the mutant activity on its cognate promoter to its activity on the wild type promoter. Bold data are from *in vitro* assays. Reported activity and orthogonality is taken from data in the literature (or **Figures 3-11 and 3-12** for CGG-R12-KIRV). Measured activity and orthogonality are taken from data presented in **Figures 3-14 and 3-16**.

Directed evolution of a panel of orthogonal T7 RNA polymerase:promoter pairs

Selection for the use of P_{CTGA}

To increase the number of available, mutually orthogonal T7 RNA polymerase variants, an attempt was made to alter the molecular recognition of T7 RNA polymerase to drive transcription from another novel promoter, P_{CTGA}. This promoter was chosen because it is different from the wild type T7 promoter at all four of the most critical base pairs (-11 to -8). It is also different from the most orthogonal existing promoters (P_{CGG}, P_{T3}, P_{K1F}, and P_{N4}) at two or more of those positions.

A construct was generated in which P_{CTGA} drives *Taq* DNA polymerase, and a strain carrying this plasmid was transformed with the Round 0 population used previously (CRSL R0). This library began to converge on active sequences after four rounds of selection by CPR. After seven rounds the library was dominated by a single variant (termed CTGA-R7-1 (**Figure 3-17**) that had the sequence L747I, N748T, R756T, Q758K. In order to further ascertain the functionality of CTGA-R7-1 an expression construct for the polymerase was transformed into an *E. coli* strain that also carried a GFP gene coupled to the P_{CTGA} promoter (**Figure 3-17**). CTGA-R7-1 produced about 1% as much GFP from P_{CTGA} as the wild type T7 RNA polymerase produced from P_{T7}.

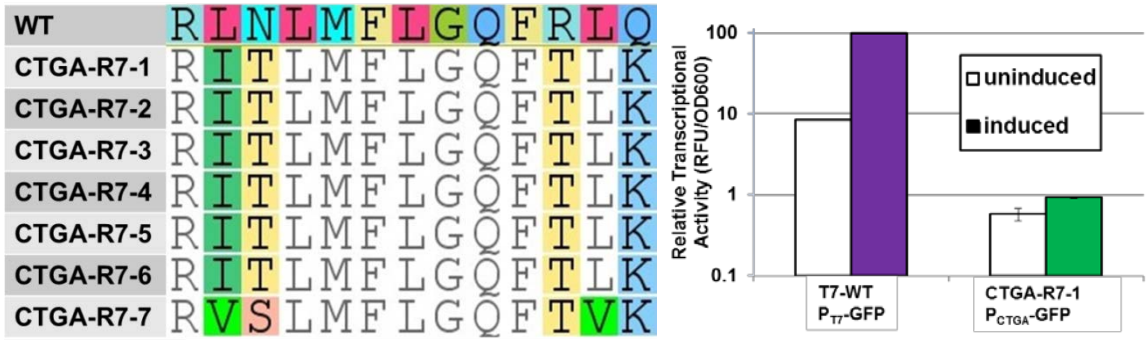


Figure 3-17. Analysis of Round 7 of the T7 RNA polymerase specificity loop selection for the use of P_{CTGA}.

Seven variants from Round 7 of the selection for the use of P_{CTGA} were sequenced. Mismatches from the wild type sequence (shown at the top) are indicated by colored blocks. CTGA-R7-1 was transformed into a P_{CTGA}-GFP reporter line. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity the polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The weak activity of CTGA-R7-1 was in line with the generally low efficiency seen by CGG-R7-8. It was therefore hypothesized that further mutations in the regions flanking the promoter specificity determinants would provide substantive improvements in activity. As before, a larger region (amino acid residues 663 to 793) was subjected to error-prone PCR and CPR optimization. After six further rounds of selection (with error-prone PCR at Rounds 8, 10, 11, and 13) some 288 clones from the library were assayed for *in vivo* GFP expression from P_{CTGA}. Active clones (judged by fluorescence readings) were sequenced, and several mutations were found to occur frequently in the population, namely, I681L, Q744K, H772R, E775V and to a lesser extent V725A, T748S, and L749I, (**Figure 3-18**).

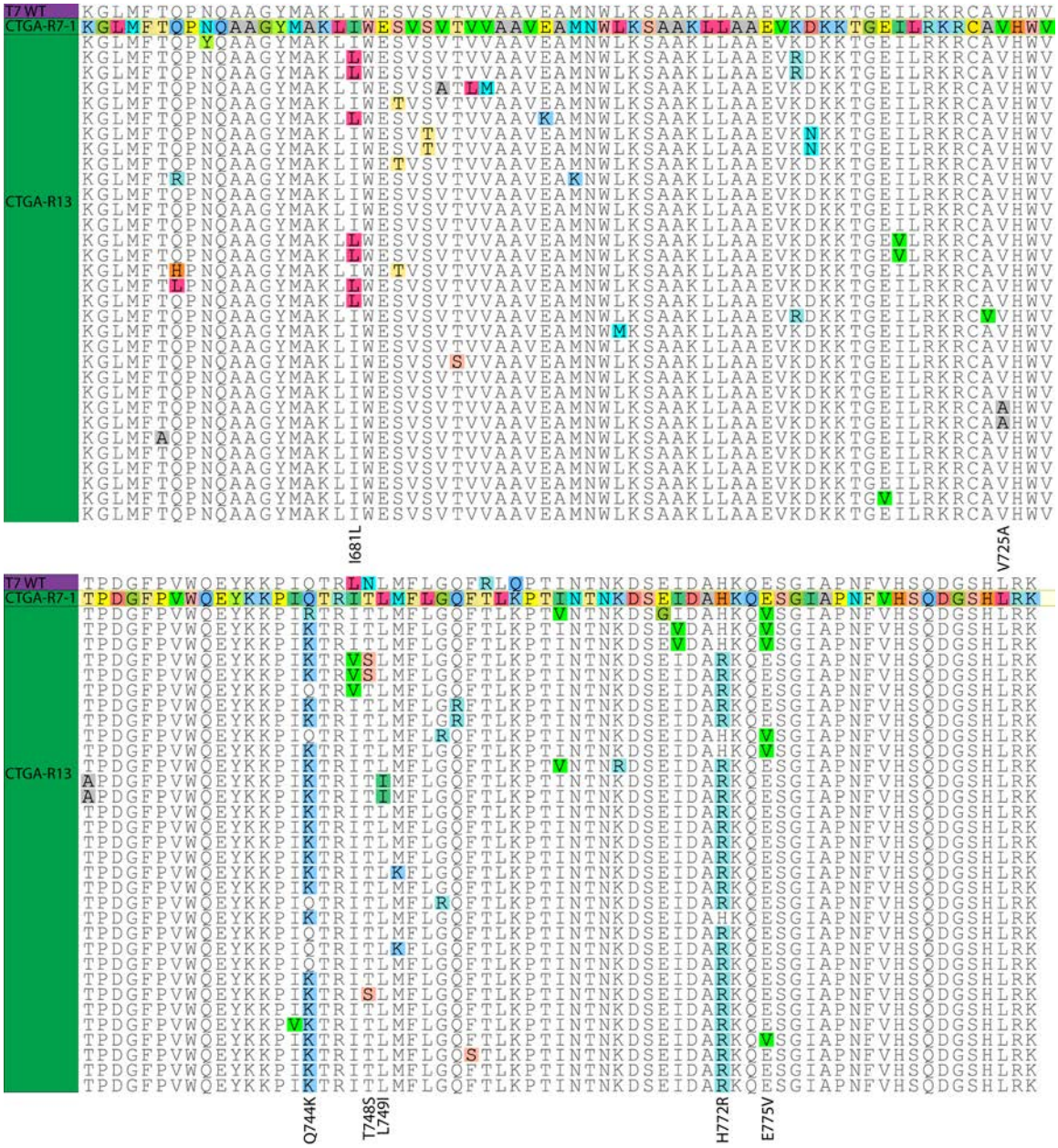


Figure 3-18. Sequence analysis of Round 13 of the T7 RNA polymerase specificity loop selection for the use of P_{CTGA}.

The library from Round 13 of the selection for the use of P_{CTGA} was transformed into a P_{CTGA}-GFP reporter strain. Individual clones were measured for GFP fluorescence, and the 32 most fluorescent variants were sequenced. Wild type T7 RNA polymerase sequence is shown at the top. Mismatches from the CTGA-R7-1 polymerase sequence (shown near the top) are indicated by colored blocks. Residues 633 to 793 are shown.

Different combinations of the selected mutations were added to the original CTGA-R7-1 variant, and then assayed for activity in the P_{CTGA}-GFP strain (**Figure 3-19**). The clone CTGA-R13-AKSRV that contained the mutations V725A, Q744K, T748S, H772R, E775V on top of the parental CTGA-R7-1 (L747I, N748T, R756T, Q758K) was found to be highly active, demonstrating 43.5% activity on P_{CTGA} relative to the wild type combination.

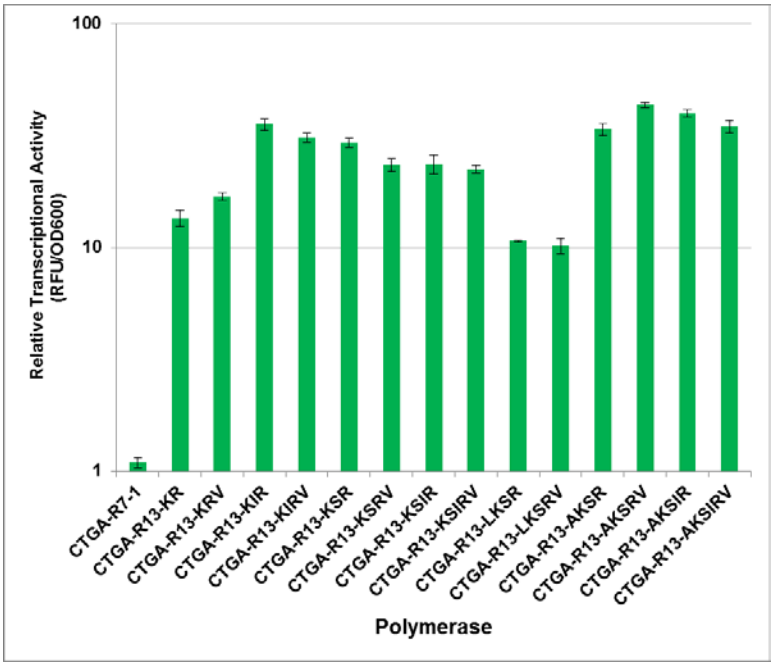


Figure 3-19. *In vivo* activity assay for the use of P_{CTGA}.
Various combinations of consensus mutations were added to CTGA-R7-1. These mutants were transformed into a P_{CTGA}-GFP reporter strain. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase with P_{CTGA} relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

Selection for the use of P_{T3}, P_{K1F}, and P_{N4}

T7 RNA polymerase variants capable of utilizing P_{T3}, P_{K1F}, and P_{N4} promoters have been described above (**Table 3-1**). T7(T3), T7(K1F), T7(N4) were made by grafting the promoter specificity loop from related bacteriophage encoded polymerases

onto the T7 RNA polymerase scaffold (Temme *et al*., 2012a). Upon initial testing it was discovered that while these polymerases were relatively specific for their respective promoters, their overall activity was low. Each polymerase displayed only about 1-10% of the *in vivo* activity of the wild type polymerase:promoter pair (**Figures 3-13, 3-14, 3-15, and Table 3-2**). This is perhaps because the grafted promoter specificity loops are not oriented to the promoter with the optimal geometry. Even though it is remarkable that such grafting was able to confer specificity, the differences in the structural context likely limited activity. It was therefore hypothesized that the activity could be improved by directed evolution.

Each polymerase variant was subject to five rounds of CPR (with error-prone PCR prior to Rounds 1, 3 and 5). As above, the final population was subjected to a GFP-based screen, consensus mutations were determined (**Figures 3-20, 3-21 and 3-22**), and combinatorial mutants were generated and assayed (**Figures 3-23, 3-24 and 3-25**). The most active variants from each selection were: T3-R5-RRVH (T745K, N748D, L749M, M750I, G753R, H772R, E775V, Q786H); K1F-R5-KIKR (Q744K, L749I, M750K, Q754S, R756N, I761V, H772R); and N4-R5-YRNRV (N671Y, L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V) (**Table 3-3**). These variants demonstrated 182.0%, 78.6%, and 22.2% activity in comparison to wild type T7 RNA polymerase with its promoter, which were respectively 20-, 119-, and 91-fold improvements over the parental enzymes.

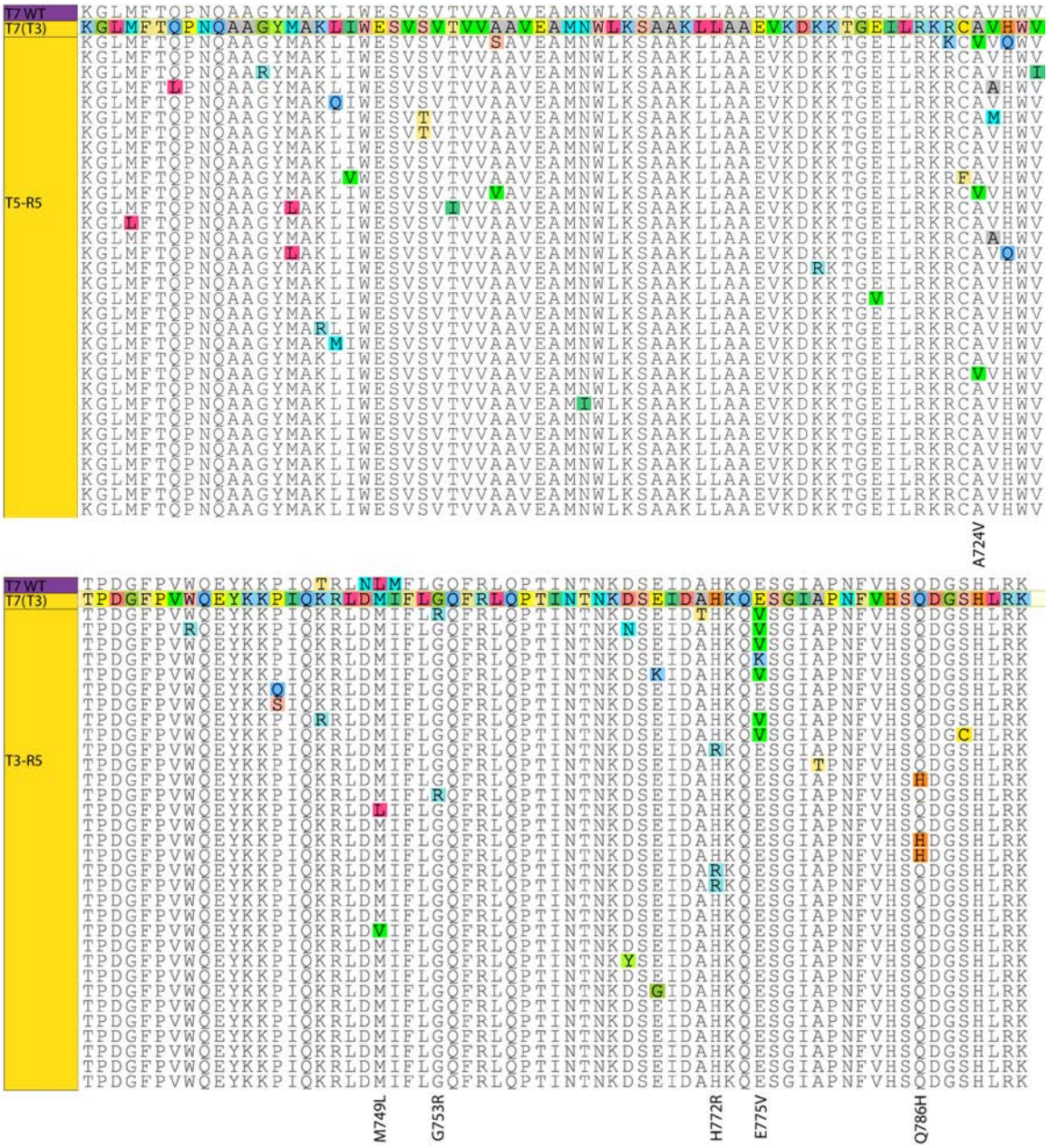


Figure 3-20. Sequence analysis of Round 5 of the selection for the use of P_{T3}.
The library from Round 5 of the selection for the use of P_{T3} was transformed into a P_{T3}-GFP reporter strain. Individual clones were measured for GFP fluorescence, and the 32 most fluorescent variants were sequenced. Wild type T7 RNA polymerase sequence is shown at the top. Mismatches from the T7(T3) polymerase sequence (shown near the top) are indicated by colored blocks. Residues 633 to 793 are shown.

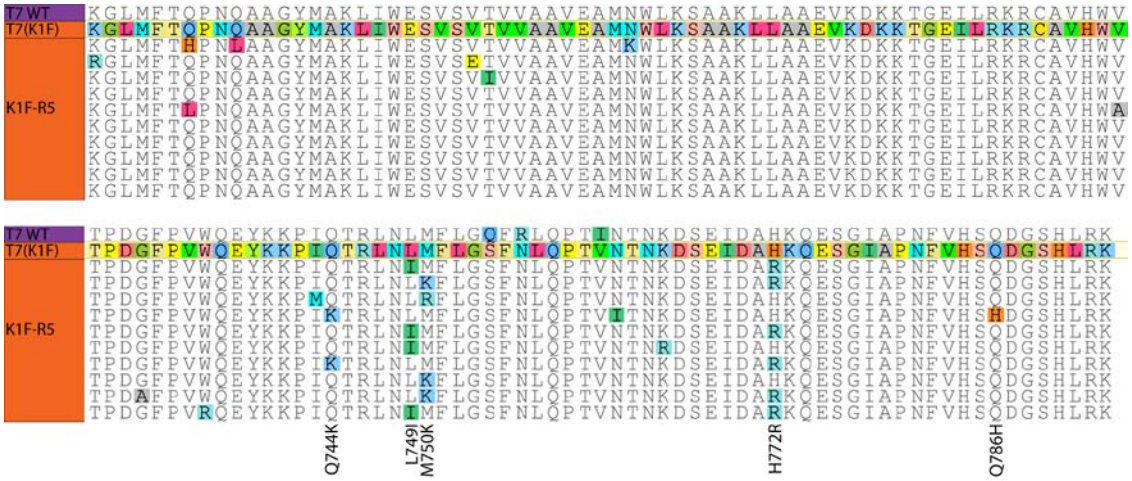


Figure 3-21. Sequence analysis of Round 5 of the selection for the use of P_{K1F}.
The library from Round 5 of the selection for the use of P_{K1F} was transformed into a P_{K1F}-GFP reporter strain. Individual clones were measured for GFP fluorescence, and the 10 most fluorescent variants were sequenced. Wild type T7 RNA polymerase sequence is shown at the top. Mismatches from the T7(K1F) polymerase sequence (shown near the top) are indicated by colored blocks. Residues 633 to 793 are shown.

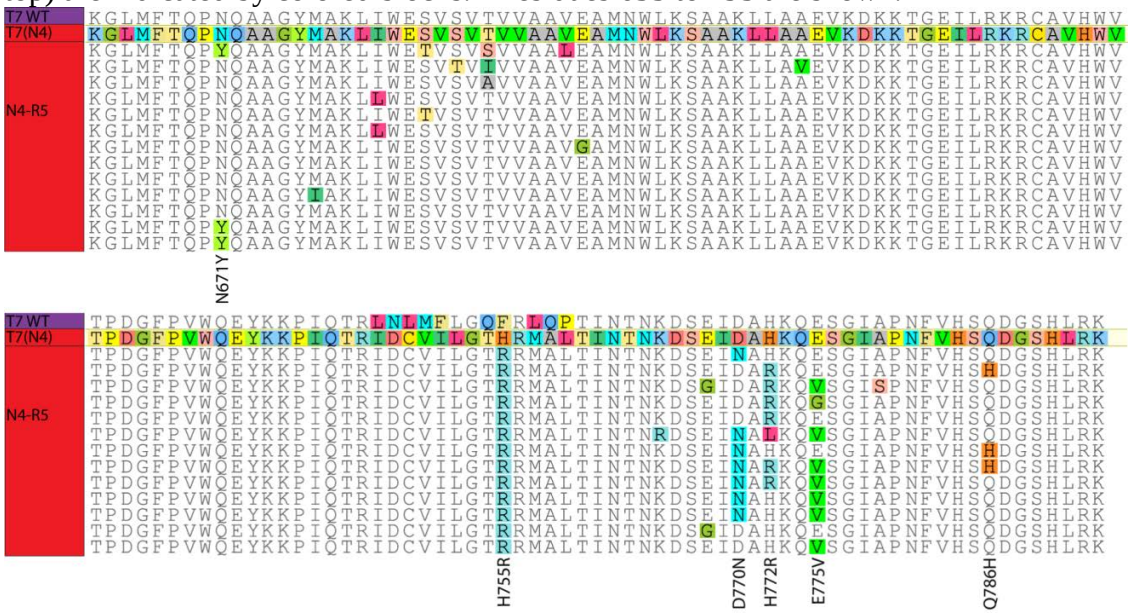


Figure 3-22. Sequence analysis of Round 5 of the selection for the use of P_{N4}.
The library from Round 5 of the selection for the use of P_{N4} was transformed into a P_{N4}-GFP reporter strain. Individual clones were measured for GFP fluorescence, and the 13 most fluorescent variants were sequenced. Wild type T7 RNA polymerase sequence is shown at the top. Mismatches from the T7(N4) polymerase sequence (shown near the top) are indicated by colored blocks. Residues 633 to 793 are shown.

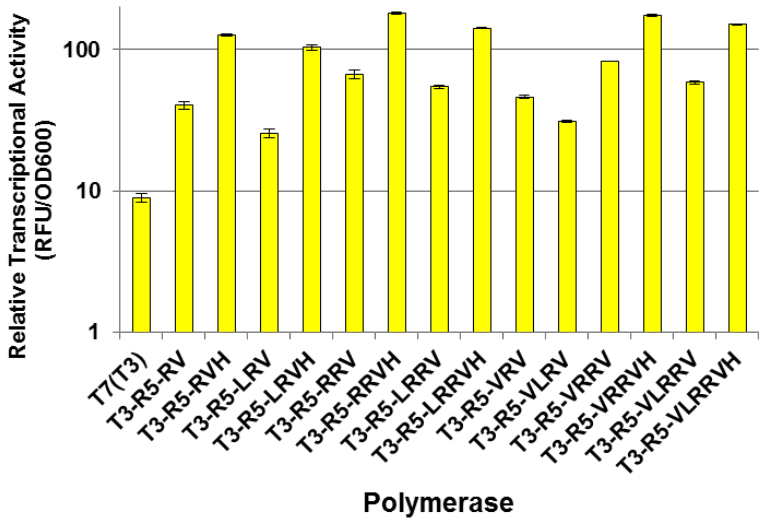


Figure 3-23. *In vivo* activity assay for the use of P_{T3}.
Various combinations of consensus mutations were added to T7(T3). These mutants were transformed into a P_{T3}-GFP reporter strain. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase with P_{T3} relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

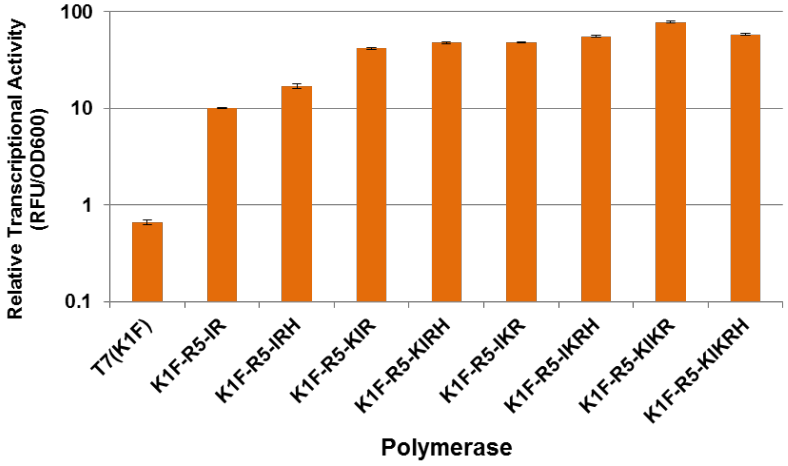


Figure 3-24. *In vivo* activity assay for the use of P_{K1F}.
Various combinations of consensus mutations were added to T7(K1F). These mutants were transformed into a P_{K1F}-GFP reporter strain. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase with P_{K1F} relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

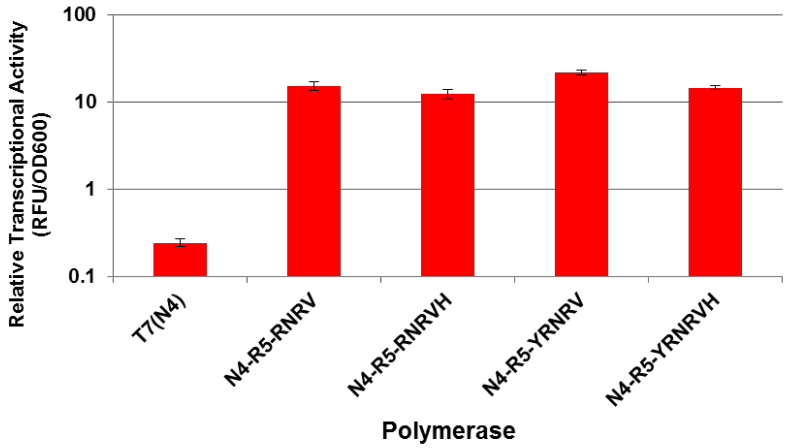


Figure 3-25. *In vivo* activity assay for the use of P_{N4}.
Various combinations of consensus mutations were added to T7(N4). These mutants were transformed into a P_{N4}-GFP reporter strain. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase with P_{N4} relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

Assay of the most active mutant from each selection

It was hypothesized that selecting for very active variants would, in turn, lead to very specific variants, based on previous studies (Matsumura and Ellington, 2001; Tokuriki *et al.*, 2012). For example, Tawfik and colleagues showed that as a hydrolase gained activity with a new substrate, it also gained specificity towards that substrate relative to the original substrate (Tokuriki *et al.*, 2012). Each of the five variant polymerases, and the wild type T7 RNA polymerase was tested against each of the five cognate promoters and P_{T7} (6 x 6 in total). Each polymerase was separately transformed into six distinct cell lines in which the six different promoters drove the production of GFP (**Figure 3-26 and 3-27, Table 3-3**). While no polymerase showed less than 3.3-fold selectivity for its cognate promoter over any other promoter, there were several cases of cross-reactivity. Specifically, T3-R5-RRVH was quite cross-reactive with P_{T7}, showing only an 8.8-fold preference for P_{T3}. Similarly, K1F-R5-KIKR prefers P_{K1F} by only 3.3-

fold over P_{T7} and 6.1-fold over P_{T3}. All other polymerase variants utilize their cognate promoters with at least 10-fold specificity, making them among the most orthogonal polymerases ever created by enzyme engineering.

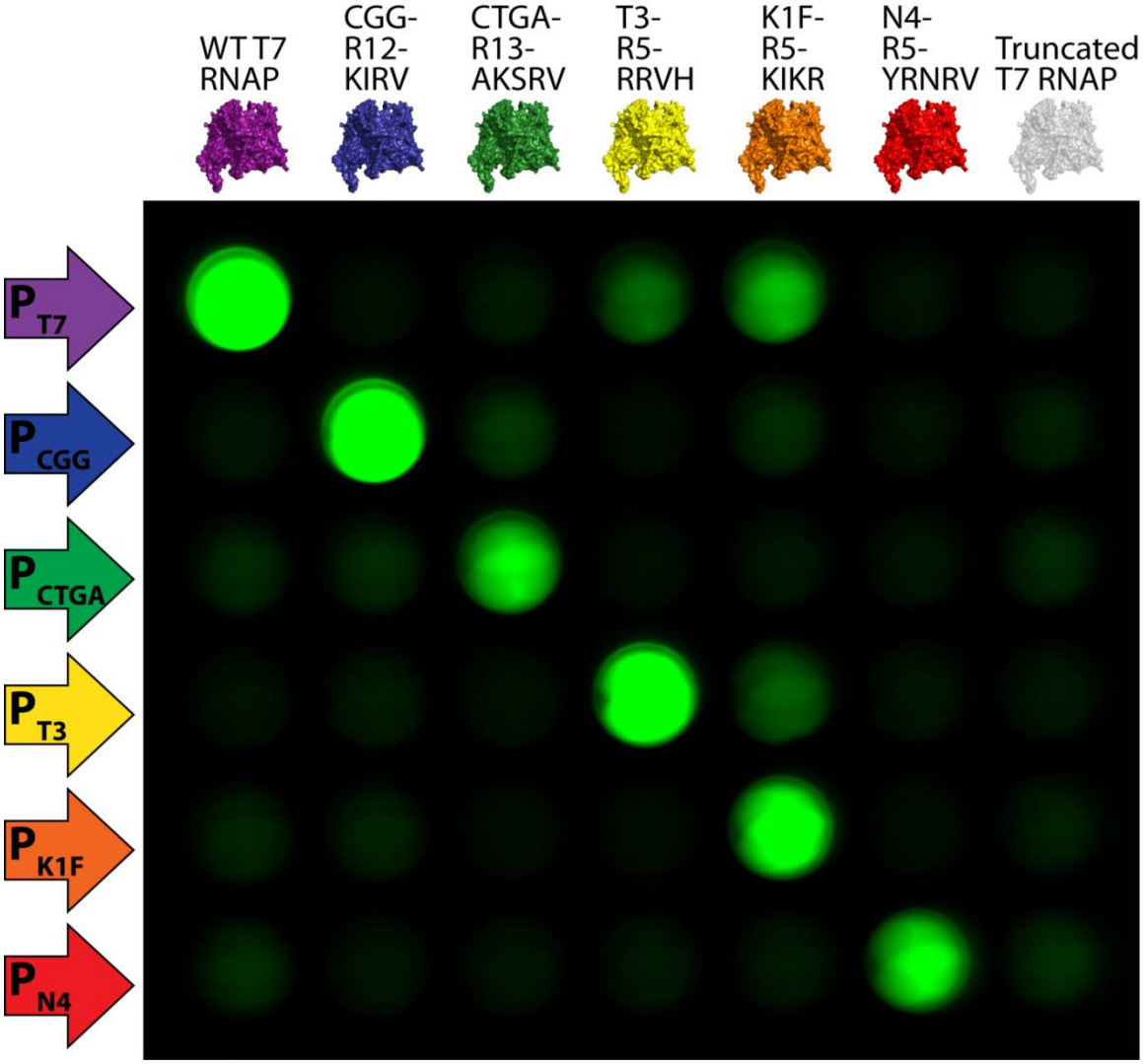


Figure 3-26. *In vivo* activity and cross-reactivity assay of the most active T7 RNA polymerase mutants from each selection.
The most active variant from each selection was transformed into each of six different *E. coli* strains, which contain the GFP gene under the control of one of the six T7 promoter variants. Cells were grown, induced, and imaged to obtain a visual depiction of activity and cross-reactivity.

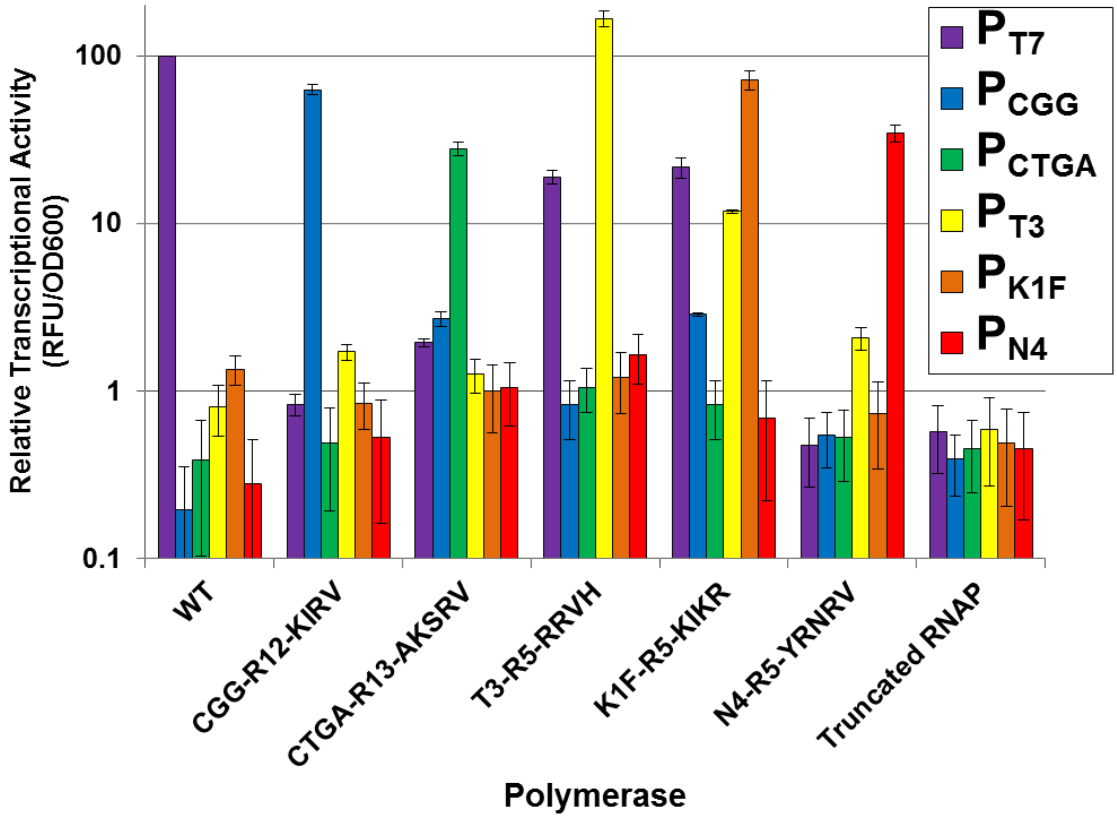


Figure 3-27. Quantitative *in vivo* activity and cross-reactivity assay of the most active T7 RNA polymerase mutants from each selection.

The most active variant from each selection was transformed into each of six different *E. coli* strains, which contain the GFP gene under the control of one of the six T7 promoter variants. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

.

Mutant	Sequence	Cognate Promoter	Activity on Cognate
WT	WT	P_{T7} TAATACGACTCACTATA	100
CGG-R12-KIRV	Q744K, L747V, N748H, L749I, R756E, L757M, H772R, E775V	P_{CGG} TAATAC**CGG**TCACTATA	50.5
CTGA-R13-AKSRV	V725A, Q744K, L747I, N748S, R756T, Q758K, H772R, E775V	P_{CTGA} TAATAC**CTGA**CACTATA	42.9
T3-R5-RRVH	T745K, N748D, L749M, M750I, G753R, H772R, E775V, Q786H	P_{T3} TAATA**ACC**CTCACTATA	160.3
K1F-R5-KIKR	Q744K, L749I, M750K, Q754S, R756N, I761V, H772R	P_{K1F} TAATA**ACTA**TCACTATA	76.2
N4-R5-YRNRV	N671Y, L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V	P_{N4} TAATA**ACC**C**A**CACTATA	25.5

Table 3-3. A list of the most active T7 RNA polymerase promoter specificity mutants from each selection.
The name, sequence, and preferred promoter of the most active T7 RNA polymerase mutant from each selection are shown. "Activity on cognate" data are taken from (*in vivo*) experiments depicted in **Figures 3-10, 3-19, 3-23, 3-24, 3-25, 3-26, 3-27, 3-28, 3-29, 3-30, 3-31, 3-32**. They are the average of three different experiments (each in triplicate) on three different days.

Cross-reactivity assays

Comparison of the data in **Figure 3-26** and **Figure 3-14** suggests that the most active polymerases from each selection are also more orthogonal than most previously described polymerases. However, while it is reasonable to believe that selecting and screening for highly active mutants can yield highly specific mutants, some mutations that improve efficiency do so via non-specific mechanisms (for example, by forming an interaction with the phosphate backbone). Therefore, the previously generated panel of variants from each selection was rescreened not only for their ability to drive GFP production from their cognate promoters but also for their cross-reactivity. In each case, several additional candidates that were very active yet less cross-reactive were obtained.

The mutants from the P_{CGG} selection were tested against P_{CGG}, as well as P_{T7} and P_{T3} (**Figure 3-28**). CGG-R12-KIR (Q744K, L747V, N748H, L749I, R756E, L757M, H772R) was 26.0% active with P_{CGG} and had at least a 36.4-fold selectivity over P_{T7} and P_{T3} (compared to 28.5-fold for CGG-R12-KIRV.)

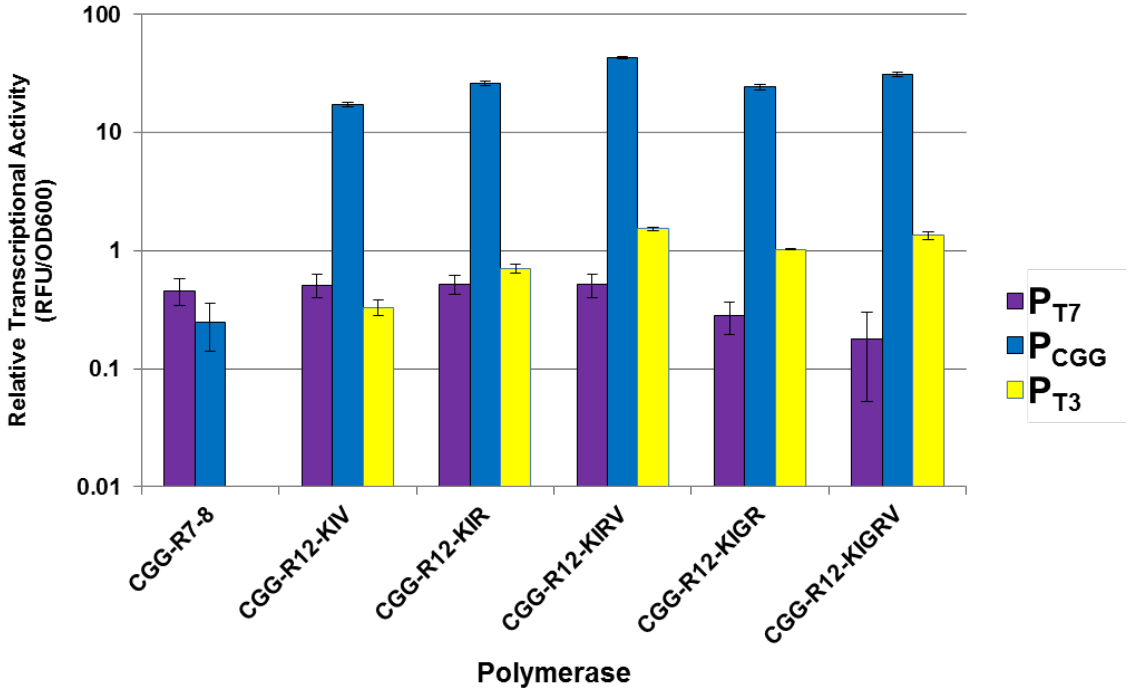


Figure 3-28. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{CGG} selection.
The mutants from the P_{CGG} selection were transformed into reporter strains in which GFP was driven by P_{T7}, P_{CGG}, or P_{T3}. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The mutants from the P_{CTGA} selection were tested against P_{CTGA}, as well as P_{T7} and P_{CGG} (**Figure 3-29**). CTGA-R13-AKSIRV (V725A, Q744K, L747I, N748S, L749I, R756T, Q758K, H772R, E775V) was 17.0% active with P_{CTGA} and had at least a 40.5-fold selectivity over P_{T7} and P_{CGG} (compared to 15.2-fold for CTGA-R13-AKSRV.)

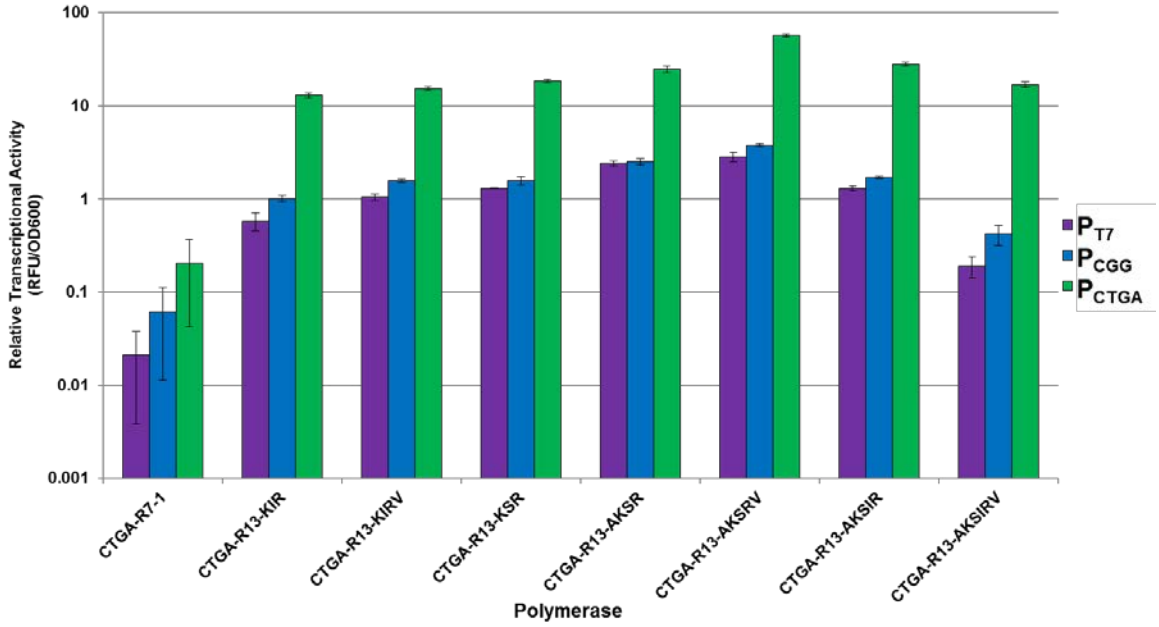


Figure 3-29. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{CTGA} selection.

The mutants from the P_{CTGA} selection were transformed into reporter strains in which GFP was driven by P_{T7}, P_{CGG}, or P_{CTGA}. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The mutants from the P_{T3} selection were tested against P_{T3}, as well as P_{T7} (**Figure 3-30**). T3-R5-RV (T745K, N748D, L749M, M750I, H772R, E775V) was 34.1% active with P_{T3} and had at least a 46.2-fold selectivity over P_{T7} (compared to 4.5-fold for T3-R5-RRVH.)

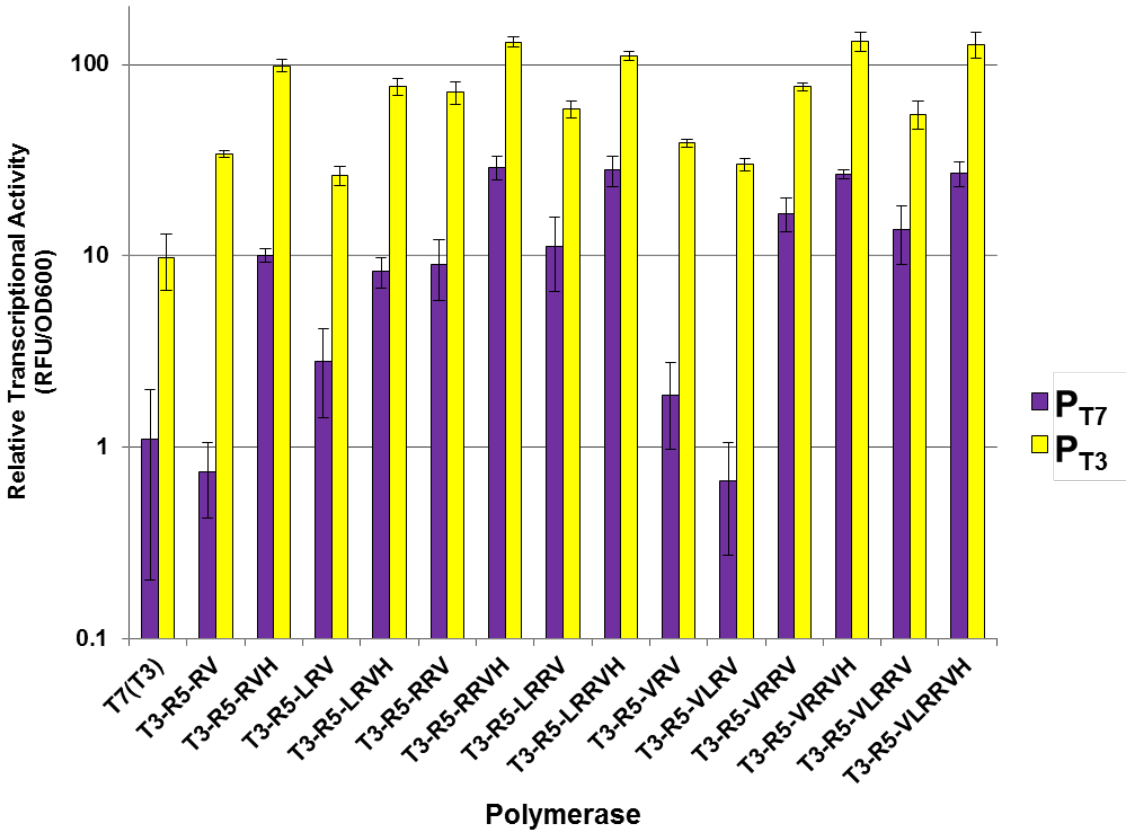


Figure 3-30. ***In vivo*** **cross-reactivity assay of T7 RNA polymerase mutants from the P_{T3} selection.**

The mutants from the P_{T3} selection were transformed into reporter strains in which GFP was driven by P_{T7} or P_{T3}. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The mutants from the P_{K1F} selection were tested against P_{K1F}, as well as P_{T7}, P_{CGG}, and P_{T3} (**Figure 3-31**). K1F-R5-IRH (L749I, Q754S, R756N, I761V, H772R, Q786H) was 19.5% active with P_{K1F} and had at least a 10.6-fold selectivity over P_{T7}, P_{CGG}, and P_{T3} (compared to 4.2-fold for K1F-R5-KIKR.)

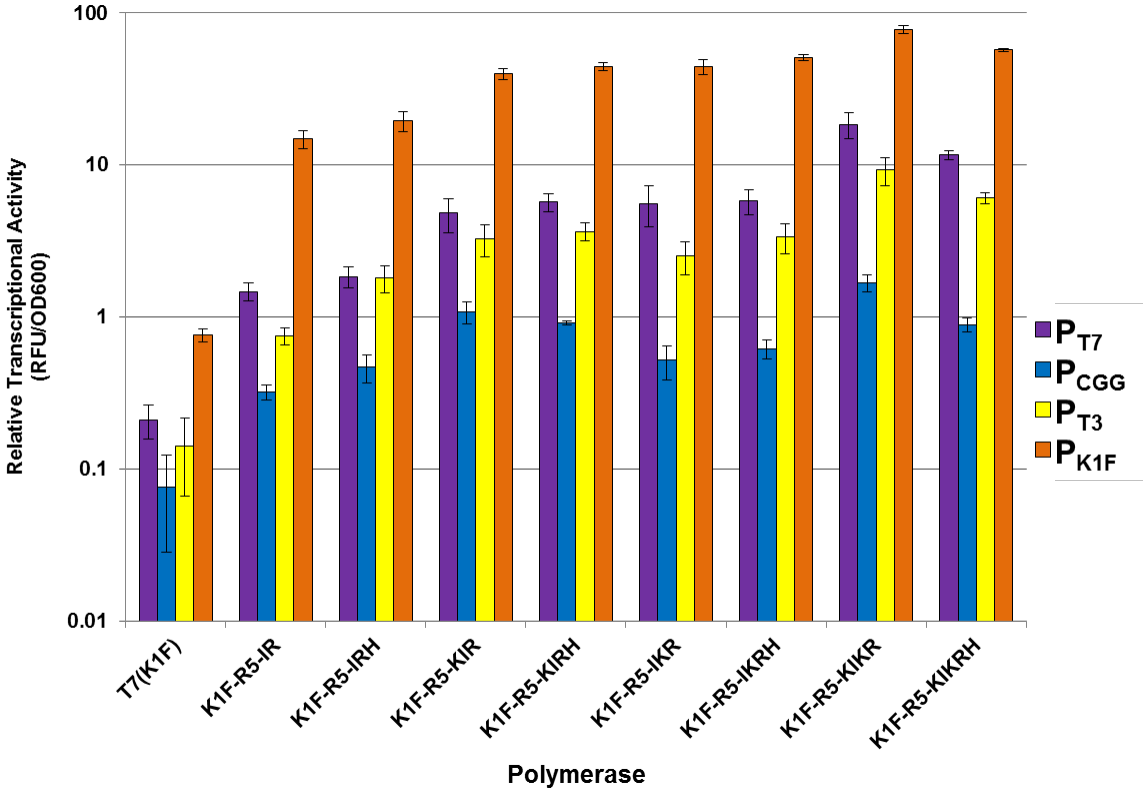


Figure 3-31. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{K1F} selection.

The mutants from the P_{K1F} selection were transformed into reporter strains in which GFP was driven by P_{T7}, P_{CGG}, P_{T3}, or P_{K1F}. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

The mutants from the P_{N4} selection were tested against P_{N4}, as well as P_{T3} (**Figure 3-32**). N4-R5-RNRV (L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V) was 13.0% active with P_{N4} and had at least an 8.5-fold selectivity over P_{T3} (compared to 8.1-fold for N4-R5-YRNRV.)

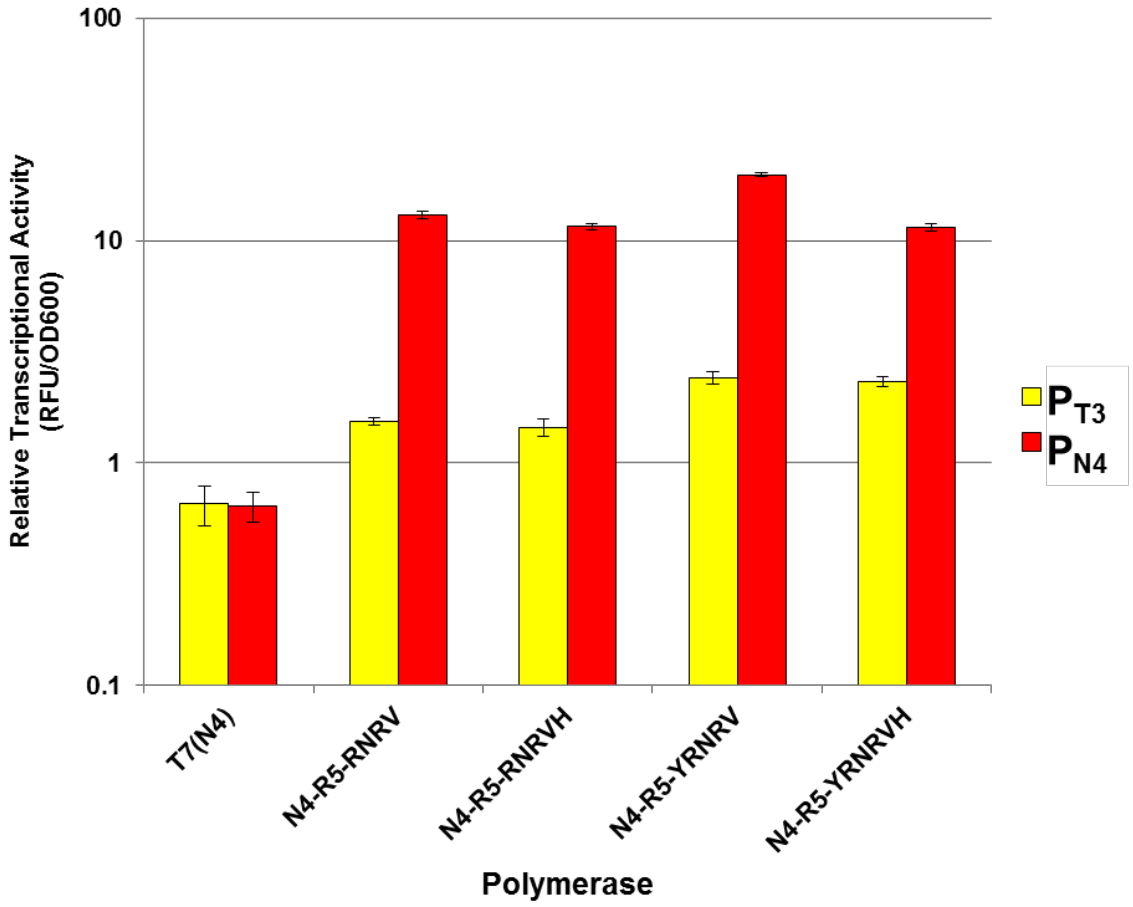


Figure 3-32. *In vivo* cross-reactivity assay of T7 RNA polymerase mutants from the P_{N4} selection.

The mutants from the P_{N4} selection were transformed into reporter strains in which GFP was driven by P_{T3} or P_{N4}. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

Assay of the most specific mutant from each selection

In order to confirm the orthogonality of these polymerase:promoter pairs *in vivo*, each polymerase was again assayed with the set of six promoters. As above, each polymerase was separately transformed into six distinct cell lines in which the six

different promoters drove the production of GFP. Visual analysis indicates that this new panel of polymerases is, in fact, truly orthogonal (**Figure 3-33**).

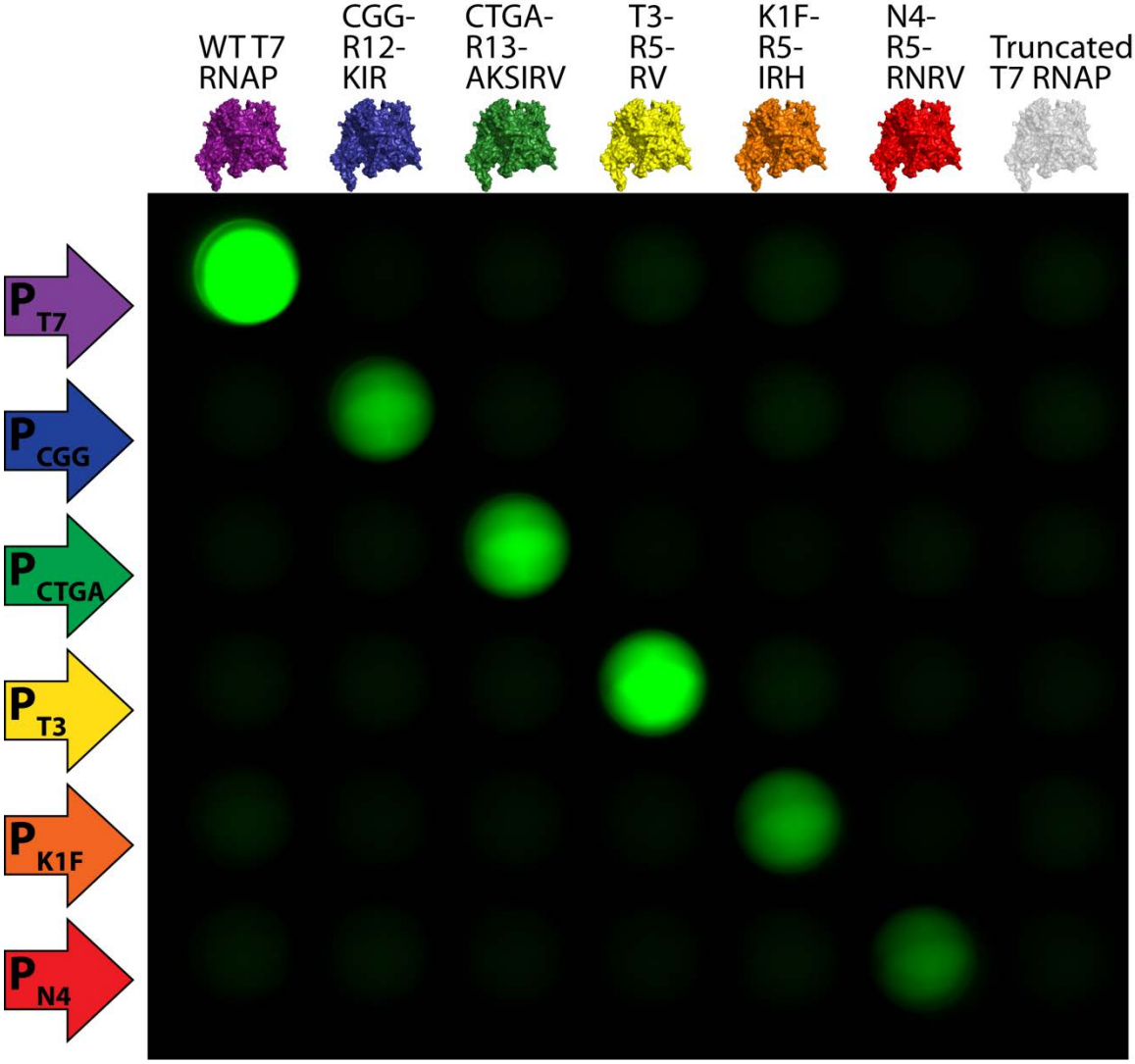


Figure 3-33. *In vivo* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection.
The most active variant from each selection/screen was transformed into each of six different E. coli strains, which contain the GFP gene under the control of one of the six T7 promoter variants. Cells were grown, induced, and imaged to obtain a visual depiction of activity and cross-reactivity.

Quantitative measurement of the activity and cross-reactivity of these polymerases showed that they ranged from 15.6 to 43.5% activity on their respective cognate promoter and no polymerase showed more than 2.5% activity on a non-cognate promoter (**Figure 3-34**).

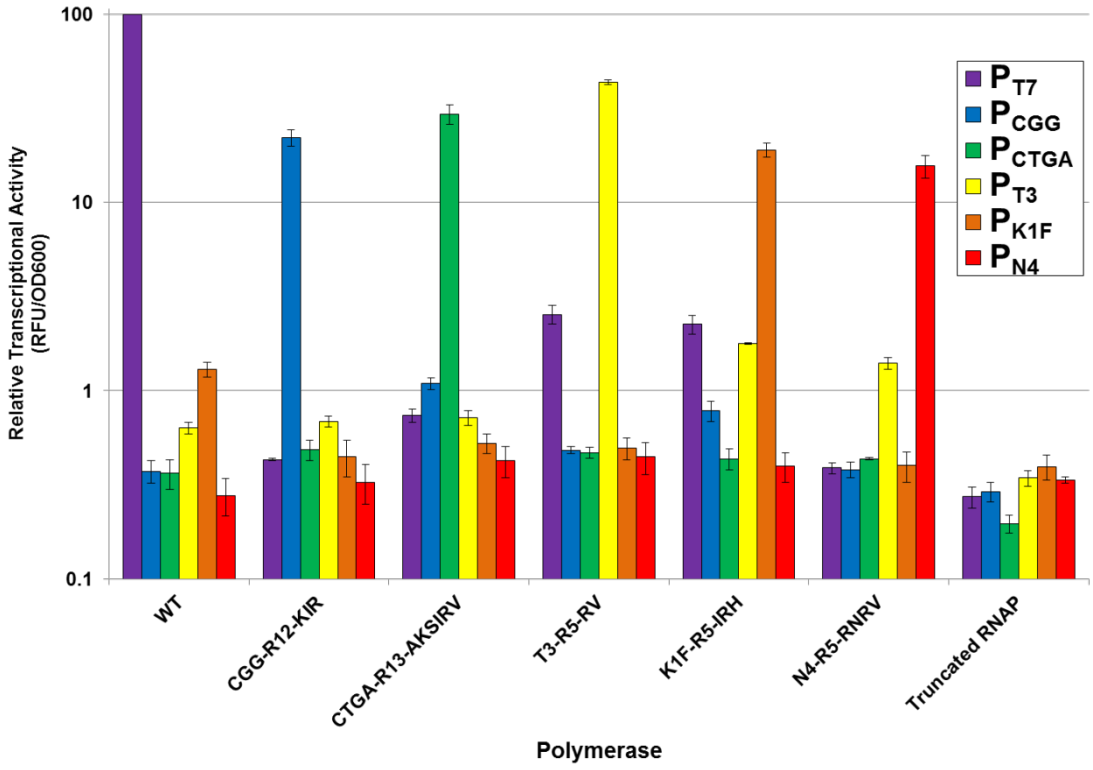


Figure 3-34. Quantitative *in vivo* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection.
The most specific variant from each selection was transformed into each of six different *E. coli* strains, which contain the GFP gene under the control of one of the six T7 promoter variants. GFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent cultures; error bars represent standard error.

It can be misleading to judge the quality of evolved proteins in a context similar to the conditions under which they evolved. To demonstrate that CPR did not just yield polymerases that were active and orthogonal *in vivo*, each polymerase was purified and

assayed for its ability to drive the transcription of the fluorescent aptamer Spinach from each promoter (**Figure 3-35**). As before, fluorescence readings were recorded as a measure of transcriptional activity, and thus as an indicator of promoter recognition *in vitro*. The evolved polymerases again showed high activity with their cognate promoters (ranging from 33% to 64% activity, relative to wild type T7 RNA polymerase with its cognate promoter) and have excellent specificity (13.5- to 45.8-fold preferences for their respective cognate promoters).

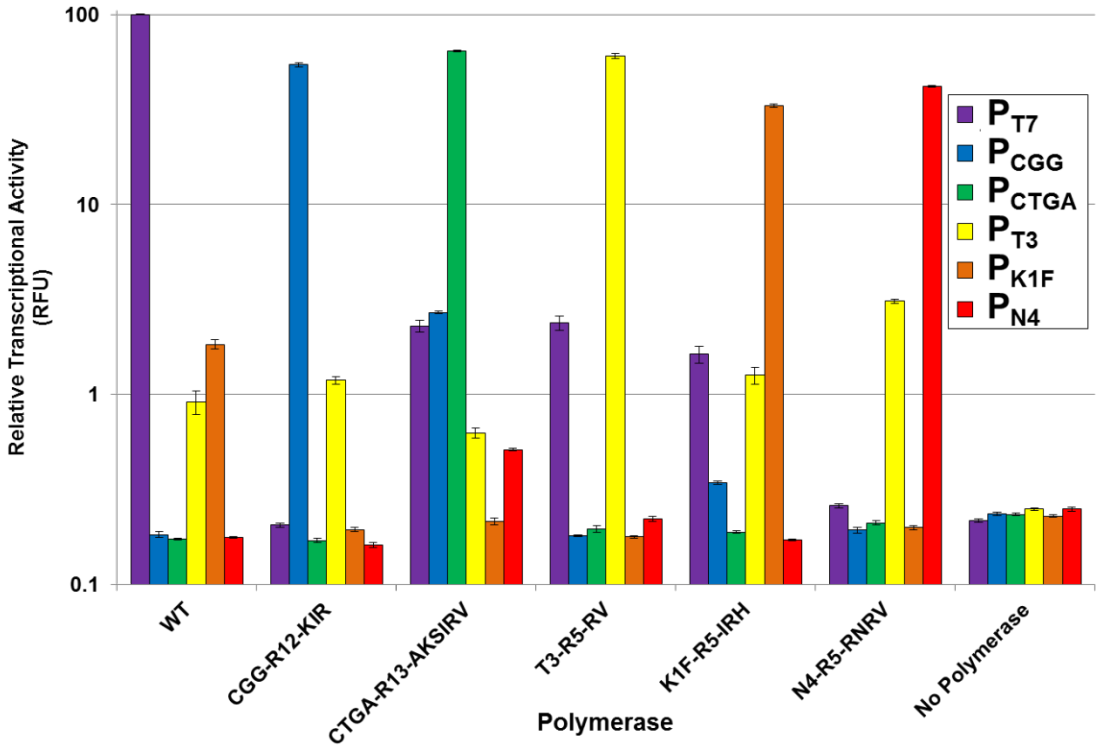


Figure 3-35. *In vitro* activity and cross-reactivity assay of the most specific T7 RNA polymerase mutants from each selection.
The most specific variant from each selection were used in an *in vitro* transcription assay. P_{T7} or mutant promoter driven expression of the Spinach aptamer was used as a readout to quantify activity and cross-reactivity. Spinach fluorescence (in Relative Fluorescence Units) is a measure of the activity of each polymerase:promoter pair relative to the activity of the WT:P_{T7} pair, defined as 100. Values are the average from three independent reactions; error bars represent standard error.

This panel of orthogonal T7 RNA polymerase promoter specificity mutants demonstrates a high degree of activity and specificity for their respective cognate promoters. Their characteristics are summarized in **Table 3-4**.

Mutant	Sequence	Cognate promoter	*In vivo* activity	*In vivo* specificity	*In vitro* activity	*In vitro* specificity
WT	WT	P_{T7} TAATACGACTCACTATA	100	75.5	100	54.6
CGG-R12-KIR	Q744K, L747V, N748H, L749I, R756E, L757M, H772R	P_{CGG} TAATAC**CGG**TCACTATA	31.0	34.3	54.4	45.8
CTGA-R13-AKSIRV	V725A, Q744K, L747I, N748S, L749I, R756T, Q758K, H772R, E775V	P_{CTGA} TAATAC**CTGA**CACTATA	27.0	33.8	64.4	23.8
T3-R5-RV	T745K, N748D, L749M, M750I, H772R, E775V	P_{T3} TAATA**ACC**CTCACTATA	39.3	31.7	60.5	25.5
K1F-R5-IRH	L749I, Q754S, R756N, I761V, H772R, Q786H	P_{K1F} TAATA**ACTA**TCACTATA	18.5	9.5	33.1	20.3
N4-R5-RNRV	L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V	P_{N4} TAATA**ACC**C**A**CACTATA	14.7	9.8	41.9	13.5

Table 3-4. A list of the most specific T7 RNA polymerase promoter specificity mutants from each selection.
The name, sequence, and preferred promoter of the most specific T7 RNA polymerase mutant from each selection are shown. "*In vivo* activity" data are taken from the experiments depicted in **Figures 3-10, 3-19, 3-23, 3-24, 3-25, 3-28, 3-29, 3-30, 3-31, 3-32, 3-33, and 3-34**. They are average of three different experiments (each in triplicate) on three different days. "*In vivo* orthogonality" data are taken from the experiments depicted in **Figures 3-28, 3-29, 3-30, 3-31, 3-32, 3-33, and 3-34**. They are average of two different experiments (each in triplicate) on two different days. "*In vitro* activity" and "*in vitro* orthogonality" data are taken from the experiment depicted in **Figure 3-35**.

CONCLUSION

The orthogonal T7 RNA polymerase specificity mutants described in this chapter are among the most active and by far the most specific ever generated by enzyme engineering. Even the least active and specific enzyme (N4-R5-RNRV) is on par with any other engineered polymerase.

The "hexacore" set of polymerases described above will be the standard for the field for either modular control of multiple pathways or parallel processing of nucleic acid computation. For applications requiring two orthogonal pairs, the WT:P_{T7} and CGG-R12-KIR:P_{CGG} set is the best option, as they are highly orthogonal to each other. If further pairs are required, the best option would be the addition of N4-R5-RNRV:P_{N4}, followed by the CTGA-R13-AKSIRV:P_{CTGA} pair .

The T3-R5-RV may be too cross-reactive with P_{T7}, and K1F-R5-IRH may be too cross-reactive with P_{T7} and P_{T3} to be useful if a high degree of orthogonality is required. However, since wild type T7 RNA polymerase is so active, this problem could be addressed by "weakening" P_{T7} and P_{T3}. For example, a T-2A mutation could be expected to reduce the activity of all polymerases on a given promoter by two-fold (Imburgio *et al.*, 2000; Temme *et al.*, 2012a, 2012b). This could bring the activity of WT:P_{T7} down to the level of the other polymerase:cognate-promoter pairs, and reduce the apparent cross-reactivity of the other polymerases with P_{T7}.

The quality of the evolved polymerase:promoter pairs underscores the value of directed evolution in general and CPR in particular. The methodology employed to create the libraries used to make polymerase that recognize P_{CGG} and P_{CTGA} took structural and mutational studies into consideration, and all residues thought to be important were randomized. Yet, the winners of the selections (CGG-R7-8 and CTGA-R7-1) were only weakly active. It took random mutagenesis to uncover mutations that

allowed robust activity to emerge. Similarly, T7(T3), T7(K1F), and T7(N4) were made with phylogenetic and structure considerations taken into account, yet they were also improved by random mutagenesis and selection. Taken together, these instances suggest that protein:DNA interactions are more complicated than simple considerations of conserved residues and hydrogen bonding. While CGG-R7-8, CTGA-R7-1, T7(T3), T7(K1F), and T7(N4) are all potentially able to form energetically favorable interactions with their respective cognate promoters, it seems likely that their specificity loops are not oriented properly to allow for such potential interaction to occur. Unpredictable mutations in regions several angstroms away were needed to unlock the latent promoter specificity.

In addition to the utility of directed evolution in general, the utility of CPR specifically is evident. As noted previously, there have been many attempts by several labs to evolve or engineer active and orthogonal T7 RNA polymerase promoter specificity variants. Their respective attempts relied on different methods, but all seemed to produce proteins with low activity and specificity. This is likely due to the problems of thresholding effects, fitness effects, and the lack of control over the parameters. CPR allows for selective enrichment over a range of activities, is largely uncoupled from host fitness, and allows the user to control the mutation rate as well as the specific regions that are allowed to evolve.

The large dynamic range of CPR is of critical importance to the success of the selections for promoter specificity. In order to make a polymerase based on wild type T7 RNA polymerase that is orthogonal to P_{T7}, one must choose a promoter that is divergent enough from P_{T7} to limit the chance of a generalist strategy being sufficient to survive the selection. This is depicted as two curves with minimal overlap in **Figure 3-36**. Several mutations are likely required to “jump” from one “fitness mountain” to the next (dashed

arrow (1) in **Figure 3-36**). This was evident in the early rounds of the selections for the use of P_{CTGA}, as the library converged on a single variant, CTGA-R7-1 (out of an initial 6.4 x 10^7) even though that variant was only about 1% active. This suggests that the large initial jump to a different "fitness mountain" is difficult, resulting in a jump from the peak of one mountain to a lower level on the other mountain (the jump from wild type T7 RNA polymerase to CTGA-R7-1 in **Figure 3-36**). The direct jump from peak to peak can only be achieved if all the mutations required for maximal activity on the new peak are present in the library. Fortunately, CPR's ability to select at the lower levels of activity allowed for a "foothold" (*i.e.* CTGA-R7-1) to be established on the new mountain. CPR's ability to select at the higher ends of activity was equally important, as CTGA-R7-1 was able to mature into CTGA-R13-AKSIRV because highly active variants were able to outcompete moderately active variants (dashed arrow (2) in **Figure 3-36**). In the end, a selection that favors the most active variant in a population is essential to the evolution of new specificity.

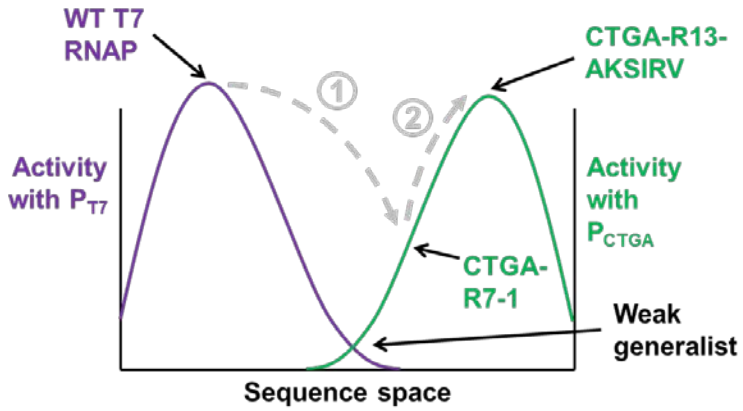


Figure 3-36. Selective pressure and the evolution of enzyme specificity.
In the evolution of a new enzyme specificity, an initial jump must be made from one specificity to the other (dashed arrow (1)). Strong selective pressure is then required to push the enzyme to the new peak (dashed arrow (2))

MATERIALS AND METHODS

Library design and selection

For the selection of an RNA polymerase mutant capable of recognizing P_{T7}, P_{CGG}, and P_{CTGA}, site saturation mutagenesis was used to randomize the residues R746, L747, N748, R756, L757, and Q758 of the T7 RNA polymerase promoter specificity loop. Diversity was introduced into the oligonucleotides during synthesis by the use of degenerate trimer phosphoramidites (Glen Research). This degenerate oligonucleotide was amplified by PCR, yielding a short double stranded library of the promoter specificity loop and short flanking regions. This short region was used in an overlap PCR with 5′ and 3′ pieces of the T7 RNA polymerase ORF. The resulting full-length T7 RNA polymerase library was digested and ligated into the pQE-RSS backbone (in which a strong T5-lac promoter drives T7 RNA polymerase).

The *Taq* DNA polymerase gene was cloned into a modified pACYC-duet (Novagen) backbone with a single T7 promoter; this construct was named pACYC-*Taq*. Variants of this plasmid with synthetic promoters drive *Taq* DNA polymerase were generated using Mega-primer PCR (Bryksin and Matsumura, 2010) or Isothermal assembly (Gibson, 2011). BL21 gold cells (Agilent) were transformed with pACYC-*Taq* (or its derivative with altered promoter) and grown overnight. 250 µl of this culture was subcultured into 20 ml 2xYT growth medium and grown at 37 °C for two hours (reaching an OD600 of approximately 0.5). The culture was then centrifuged at 4000 g at 4 °C and washed with ice cold 10% glycerol four times, with the fourth resuspension in 100 µl of 10% glycerol. This cell slurry (~200 µl total) was combined with 2 µl (~50 ng) purified ligation and electroporated using 0.2 cm cuvettes at 2.5 kV in an *E. coli* pulser (BioRad). This routinely resulted in roughly 10^7 CFUs (multiple replicates were pooled for early rounds in order to attain full coverage).

100 μl overnight transformation cultures were subcultured in 2 ml 2xYT medium, grown for one hour (OD600 ~0.6) and induced with 0.05 mM IPTG at 37 °C for four hours. 200 μl of the induced culture was centrifuged at 5,000 g for 10 minutes to pellet the cells. The supernatant was removed and cells were gently resuspended in 20 μl 10x PCR buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$) 10 μl dNTP mix (4 mM each), 4 μl each primer (20 μM), and 162 μl water. Emulsification was performed by slowly adding resuspended cells to 600 μl of spinning oil mix (438 μl Tegosoft DEC (Evonik), 42 μl AbilWE09 (Evonik), and 120 μl Mineral oil (Sigma)). The oil mixture was constantly spun in a tube (Sarstedt 13 ml 95 mm x 16.8 mm) on ice using a stirbar (Spinplus 9.5 mm x 9.5 mm Teflon, Bel-Art) on a magnetic plate (Corning) at the maximum setting (1150 rpm). The cell mixture was slowly added over a one minute interval and spun for an additional four minutes. The emulsified cells were thermal cycled (95 °C:3 min, 20 cycles [95 °C:30 sec, 55 °C:30 sec, 72 °C:2 min/kb], 72 °C:5 min) such that cells containing the most active enzymes will also contain the most *Taq* DNA polymerase and will preferentially PCR amplify. The emulsion was broken in two steps. Firstly, it was centrifuged at 10,000 g for five minutes and the oil (upper) phase was removed. Secondly, 300 μl of bead buffer (0.2 M NaCl, 10 mM Tris-HCl (pH 7.4), 1 mM EDTA, 0.1% Tween-20) and 500 μl chloroform was added and the mixture was vortexed vigorously. The mixture was transferred to a heavy-gel phase-lock tube (5 Prime) and upon centrifugation at 16,000 g for two minutes the aqueous (upper) phase was collected along with any nucleic acids present. To facilitate purification of the DNA PCR amplified by *Taq* DNA polymerase from the template plasmid DNA, primers with 5′ biotin groups were used in the emulsion PCR step. Biotinylated PCR products were purified using streptavidin coated beads (MyOne Streptavidin C1 Dynabeads, Life Technologies) and used as a template for re-amplification using nested primers. This

PCR product was gel purified and used in an assembly PCR, thus regenerating the full-length T7 RNA polymerase ORF. This was followed by digestion and ligation into the pQE-RSS backbone.

In Rounds 8 to 12 of the P_{CGG} selection, Rounds 8 to 13 of the P_{CTGA} selection, and all rounds of the P_{T3}, P_{K1F}, and P_{N4} selections, a larger region (amino acids 633 to 793) of the polymerase coding sequence was reamplified (and thus allowed to evolve). In these rounds, pQE-RSS was replaced by pLUV-RSS in which the strong T5 promoter was replaced with the moderate-strength promoter, lacUV5 promoter.

Error-prone PCR was performed on the larger region prior to P_{CGG} Rounds 8, 11, 13, 15, and 16, P_{CTGA} Rounds 8, 10, 11, and 13, as well as before P_{T3}, P_{K1F}, and P_{N4} Rounds 1, 3 and 5. Briefly the reaction mixture was composed of 50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM $MgCl_2$, 5 μg/ml BSA, 0.35 mM dATP, 0.4 mM dCTP, 0.2 mM dGTP, 1.35 mM dTTP, 0.5 mM $MnCl_2$, 0.5 μM each primer, 2 ng/μl template, and 0.8 U/μl *Taq* DNA polymerase (New England Biolabs) and was thermal cycled (95 °C:4 min, 25 cycles [95 °C:30 sec, 55 °C:30 sec, 72 °C:2 min], 72 °C:5 min). This achieved the expected one mutation per 500 base pairs.

***In vivo* GFP assay**

Purified plasmid of T7 RNA polymerase mutants (or ligations from rounds of selection) were electroporated into BL21-gold cells containing pACYC derivatives in which T7 promoter variants drive GFP production. Transformations (or single colonies) were grown at 37 °C overnight. 100 μl of the culture was grown in 2 ml 2xYT medium at 37 °C for one hour (OD600 ~0.6) and induced with 0.05 mM IPTG for four hours. This concentration of IPTG was chosen in order to limit metabolic overload on the host and prevent saturation of signal. After induction, cells were measured for OD600 on a

Synergy-HT plate reader (Bio-Tek) and GFP fluorescence (Excitation/ Emission 481 nm/507 nm) on a Safire monochromator (Tecan). Images of T7 RNA polymerase-driven GFP expression spinning down 2 ml of induced culture, decanting the supernatant, and resuspending cells in 500 µl PBS. The resuspended cells were excited with a transilluminator (475 nm Excitation) and visualized with a FluorChem Q (Cy3-filter, Protein Simple).

In vivo RNA polymerase cross-reactivity assays

A promiscuous mutant of the phenylalanine aminoacyl-tRNA synthetase, PheS A294G (Kast and Hennecke, 1991; Thyer *et al*., 2013), was cloned in place of the *Taq* DNAP open reading frame in pACYC-*Taq*. Purified plasmid of T7 RNA polymerase mutants were electroporated into BL21-gold cells containing pACYC-PheS with PheS A294G driven by the wild type T7 promoter. Transformations were grown at 37 °C overnight. 100 µl of the culture was grown in 2 ml 2xYT medium at 37 °C for one hour (OD600 ~0.6) and induced at 0.05 mM IPTG for four hours. Cells were diluted with media (containing the same antibiotics and IPTG as the growth media) to OD600s of 0.1, 0.01, and 0.001. 5 µl of each dilution was plated on 0 mM, 5 mM, 10 mM, 15 mM, or 20 mM 4-chloro-DL-phenylalanine (Cl-Phe; Sigma). The plating media also contained 0.4% glycerol, 0.5% yeast extract, 1% NaCl, 1.5% agar, 50 µg/ ml kanamycin, 24 µg/ ml chloramphenicol, and 0.05 mM IPTG. Plates were grown at 37 °C for 20 hours and visualized with ambient white light on a FluorChem Q (Protein Simple). Mutant cross-reactivity may be judged be the dose dependent cytotoxicity of Cl-Phe, which is only lethal (at the concentrations used) when PheS A294G is expressed.

T7 RNA polymerase purification

For *in vitro* transcription assays, T7 RNA polymerase variants were purified by standard Ni-NTA 6xHis (N-terminal) methods. The plasmid pQE-T7RSS (or a derivative thereof for T7 RNA polymerase mutants) was transformed in BL21-gold (Agilent). Cells were grown in 2xYT media at 37 °C until an OD600 ~0.7-0.8 was reached. Cells were induced for four hours with 1 mM IPTG, pelleted, and frozen at -80 °C. Pellets were resuspended in binding buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole). Resuspended cells were lysed via sonication on ice using 50% probe amplitude for three minutes (1 sec ON, 1 sec OFF). Cell debris was pelleted by centrifugation at 10,000 g for 30 minutes. His-tagged T7 RNA polymerase was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over 1 ml (bead volume) Ni-NTA gravity column pre-equilibrated with binding buffer. The column was washed with 10x column volumes of wash buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 20 mM imidazole). T7 RNA polymerase was eluted off the column by the addition of 4x column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 250 mM imidazole). Eluates were dialyzed twice against a storage buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DDT, 1 mM EDTA). Concentrations were adjusted to 1 mg/ml and added to an equal volume of glycerol (final concentration 0.5 mg/ml).

***In vitro* transcription assay**

Transcription templates were designed such that a T7 promoter variant was immediately upstream of the fluorescent aptamer Spinach. Templates were prepared by PCR and agarose gel purification using QIAquick Gel Extraction Kit (Qiagen). Transcription reactions were assembled by combining 40 mM Tris-HCl pH 7.0, 30 mM $MgCl_2$, 6 mM spermidine, 6 mM each NTP, 10 mM DTT, and 0.17 mg/ml DFHBI (Paige *et al*., 2011) with 0.5 μM of the appropriate T7 RNA polymerase and 0.5 μM of the

appropriate DNA template. Reactions were incubated for up to four hours at 37 °C with Spinach fluorescence (Excitation/ Emission 469 nm/501 nm) reading taken every one to three minutes in a Safire monochromator (Tecan).

REFERENCES

Bryksin, A. V, and Matsumura, I. 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* 48:463–5.

Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. 1999. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399:80–3.

Chelliserrykattil, J., Cai, G., and Ellington, A. D. 2001. A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnology* 1:13.

Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. 2012. An in vitro Autogene. *ACS Synthetic Biology* 1:190–196.

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*.

Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology* 32:97–101.

Esvelt, K. M., Carlson, J. C., and Liu, D. R. 2011. A system for the continuous directed evolution of biomolecules. *Nature* 472:499–503.

Franco, E., Friedrichs, E., Kim, J., Jungmann, R., Murray, R., Winfree, E., and Simmel, F. C. 2011. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences of the United States of America* 108:E784–93.

Gibson, D. G. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* 498:349–61.

Ikeda, R. A., Ligman, C. M., and Warshamana, S. 1992. T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Research* 20:2517–24.

Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* 39:10419–30.

Kast, P., and Hennecke, H. 1991. Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *Journal of Molecular Biology* 222:99–124.

Khersonsky, O., and Tawfik, D. S. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* 79:471–505.

Kim, J., White, K. S., and Winfree, E. 2006. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology* 2:68.

Kim, J., and Winfree, E. 2011. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology* 7:465.

Matsumura, I., and Ellington, A. D. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *Journal of Molecular Biology* 305:331–9.

Paige, J. S., Wu, K. Y., and Jaffrey, S. R. 2011. RNA Mimics of Green Fluorescent Protein. *Science* 333:642–646.

Raskin, C. A., Diaz, G. A., and McAllister, W. T. 1993. T7 RNA polymerase mutants with altered promoter specificities. *Proceedings of the National Academy of Sciences of the United States of America* 90:3147–51.

Rong, M., He, B., McAllister, W. T., and Durbin, R. K. 1998. Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 95:515–9.

Shis, D. L., and Bennett, M. R. 2013. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proceedings of the National Academy of Sciences of the United States of America*.

Subsoontorn, P., Kim, J., and Winfree, E. 2012. Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch. *ACS Synthetic Biology* 1:299–316.

Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. 2012a. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* 40:8773–81.

Temme, K., Zhao, D., and Voigt, C. A. 2012b. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America* 109:7085–90.

Thyer, R., Filipovska, A., and Rackham, O. 2013. Engineered rRNA Enhances the Efficiency of Selenocysteine Incorporation during Translation. *Journal of the American Chemical Society*.

Tokuriki, N., Jackson, C. J., Afriat-Jurnou, L., Wyganowski, K. T., Tang, R., and Tawfik, D. S. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* 3:1257.

Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. 2012. Genome-scale promoter engineering by coselection MAGE. *Nature Methods* 9:591–3.

Chapter 4: Stabilization of T7 RNA polymerase substrate specificity mutants

On average, mutations are deleterious to proteins. Mutations conferring new function to a protein often come at the expense of protein folding or stability, reducing overall activity (Wang *et al*., 2002; Tokuriki *et al*., 2008; Romero and Arnold, 2009; Levin *et al*., 2009; Soskine and Tawfik, 2010).

To date, several T7 RNA polymerase variants have been designed or evolved to accept and incorporate nucleotides with modified ribose moieties. These modified RNAs have proven useful, especially *in vivo*, but the transcriptional yields are typically lower than with native RNA, possibly due to the instability of the polymerase mutants. Here it is demonstrated that mutations previously shown to increase the thermal tolerance of T7 RNA polymerase can increase the activity of mutants with expanded substrate preferences. The resulting polymerase mutants can be employed to generate 2′-*O*-methyl modified RNA with greater yields than is achieved with enzymes currently employed.

INTRODUCTION

Background and rationale

The need for RNA polymerases capable of transcribing RNAs with 2′-modified ribose moieties

RNA is a versatile and useful macromolecule, but its chemical instability can render it unsuitable for many therapeutic and biotechnology applications. Oligonucleotides with altered chemistry, especially those with modifications at the 2′ position of the (deoxy)ribose have proven to be of great value (Wilson and Keefe, 2006). As antisense probes, 2′-*O*-methyl RNA has higher T_m values, faster binding, and greater stability than natural RNA (Majlessi *et al*., 1998). siRNA possessing 2′-F and 2′-*O*-

methyl RNA have also proven to be more stable and target-specific (Layzer, 2004; Kraynack and Baker, 2006; Jackson *et al*., 2006). Additionally, *in vitro* selections with 2′-modified NTPs have yielded aptamers and ribozymes with greater stability and enhanced chemical potential (Beaudry *et al*., 2000; Lupold *et al*., 2002; Healy *et al*., 2004; Burmeister *et al*., 2005; Keefe and Cload, 2008; Waters *et al*., 2011).

While modified RNA can be chemically synthesized it is often preferable to enzymatically produce it. This is especially important for *in vitro* selection, which relies on cycles of transcription, reverse transcription, and PCR (Ellington and Szostak, 1990; Tuerk and Gold, 1990; Knudsen *et al*., 2002). T7 RNA polymerase has long been utilized for the production of RNA *in vitro*, and has previously been the subject of engineering and evolution efforts to expand its substrate range.

T7 RNA polymerase substrate specificity mutants

The engineering and evolution of T7 RNA polymerase substrate specificity mutants is described in greater detail in **Chapter 1**. The properties of the mutants developed by those methods are now discussed. The name, sequence, substrate range, and source of these mutants are summarized in **Table 4-1**.

Enzyme	Sequence	Accepted modifications	Reference
WT	WT T7 RNAP	None	Chamberlin, *et al.*, 1970
Y639F	Y639F	2′-F, 2′-NH_2, 2′-deoxy	Huang *et al.*, 1997
FA	Y639F, H784A	2′-OMe, 2′-N_3	Padilla and Sousa, 2002
RGVG	E593G, Y639V, V685A, H784G	2′-OMe	Chelliserrykattil and Ellington, 2004
VRS	G542V, H772R, H784S	2′-F	Chelliserrykattil and Ellington, 2004
2P16	I119V, G225S, K333N, D366N, F400L, E593G, Y639V, S661G, V685A, H784G, F880Y	2′-OMe, 2′SeMe	Siegmund *et al.*, 2012
R425C	R425C	2′-OMe	Ibach *et al.*, 2013

Table 4-1. T7 RNA polymerase nucleotide specificity mutants.

The commercially important Y639F mutant allows for the polymerization of RNA transcripts containing dNTPs as well as nucleotides with 2′-fluoro and 2′-amino modified ribose (Kostyuk *et al*., 1995; Sousa and Padilla, 1995; Huang *et al*., 1997).

The Y639F, H784A double mutant ("FA") can incorporate nucleotides with bulky modifications at the 2′ position (*e.g.* 2′-*O*-methyl and 2′-azido) (Brieba and Sousa, 2000; Padilla and Sousa, 2002).

An evolved mutant, termed "RGVG," (E593G, Y639V, V685A, H784G) showed at least as much activity as FA in transcriptions containing three NTPs and either 2′-*O*-methyluridine (rVmU), 2′-*O*-methylcytidine (rDmC), or 2′-*O*-methyladenosine (rBmA). In addition, it was uniquely able to make transcripts containing two or three 2′-*O*-methyl NTPs (rRmY and rGmH), however with reduced transcription yields (Chelliserrykattil and Ellington, 2004).

Another evolved mutant, "VRS," (G542V, H772R, H784S) was reported to incorporate 2′-fluoropyrimidines (Chelliserrykattil and Ellington, 2004).

The “2P16” mutant (I119V, G225S, K333N, D366N, F400L, E593G, Y639V, S661G, V685A, H784G, F880Y) is a mutant of RGVG with seven additional mutations. It was reported to have similar activity with 2′-*O*-methyluridine (rVmU) and enhanced activity with 2′-*Se*-methyluridine (Siegmund *et al*., 2012).

The R425C mutant is reported to enable the synthesis of 1000-nucleotide long mRNA composed entirely of 2′-*O*-methylnucleotides (mN) (Ibach *et al*., 2013).

Activity stability trade-offs in enzyme mutants

While the catalytic properties of these enzymes make them useful tools, several of them suffer from low activity, even with normal ribonucleotides. It has been proposed that mutations that confer new activity to an enzyme also destabilize the protein, rendering it less active overall (Wang *et al*., 2002; Tokuriki *et al*., 2008; Romero and Arnold, 2009; Levin *et al*., 2009; Soskine and Tawfik, 2010). For example, two mutations in TEM1 β-lactamase (E104K, R164S) can alter substrate specificity to ceftazidime (CAZ) over its natural substrate 6-furylacrylpenicillanic acid (FAP) (**Figure 4-1**). However, the overall activity of this double mutant is low, and its T_m is reduced compared to wild type TEM1. Interestingly, the M182T mutation has no effect on substrate preference, but increases the T_m of wild type TEM1 as well as the E104K, R164S double mutant. This triple mutant (E104K, R164S, M182T) is approximately twice as active as the double mutant. One possible interpretation of these data is that the E104K and R164S mutations confer the new activity (use of CAZ), but these mutations also push the enzyme below a stability threshold, thus lowering its activity. As the

M182T mutant is more stable than wild type TEM1, it can tolerate the destabilizing effects of the E104K and R164S mutations.

Mutant	FAP k_{cat} (s^{-1})	CAZ k_{cat} (s^{-1})	T_m (°C)
WT (TEM1)	1210	0.0179	51.5
M182T	909	0.0159	57.7
E104K, R164S	17.7	33.9	50.9
E104K, R164S, M182T	22.0	71	57.8

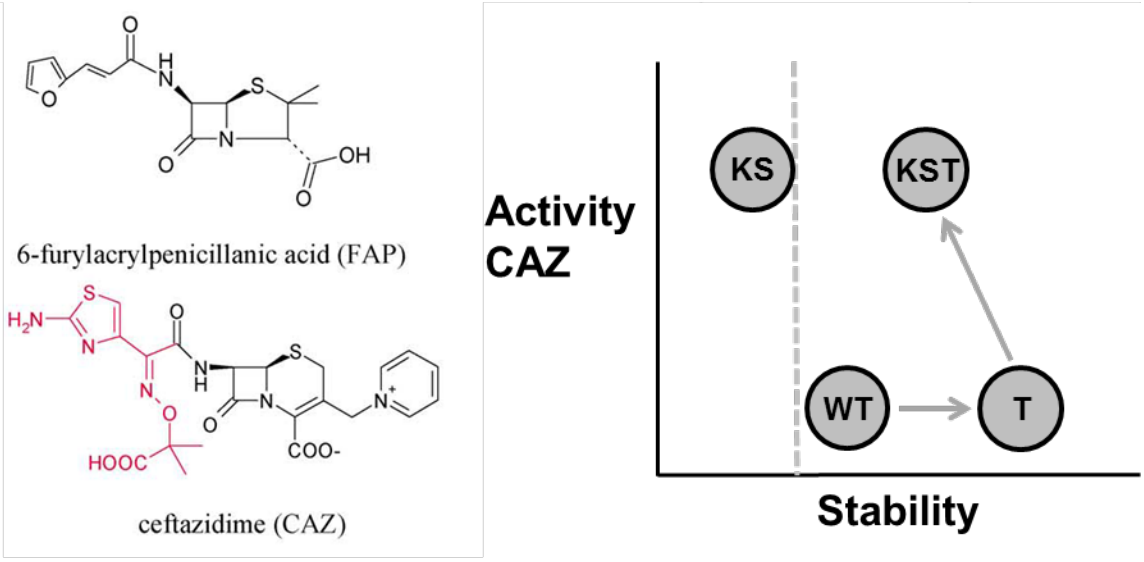


Figure 4-1. Activty stability trade-offs in TEM1 β-lactamase.
Top) The k_{cat} with two substrates and the T_m of four TEM1 mutants are shown. **Bottom left**) The structures of the two substrates are shown. **Bottom right**) The E104K, R164S (KS) mutant is below a stability threshold (dotted line). M182T (T) is more stable than wild type TEM1 (WT), so mutation from WT to M182T to E104K, R164S, M182T is a feasible pathway to a new activity (arrows). This figure is based on data from a study of TEM1 β-lactamase (Wang *et al.*, 2002), and includes the structures depicted therein.

In this chapter, this idea is put to the test, as suspected stabilizing mutations are added to the T7 RNA polymerase mutants described above. It will be seen that T7 RNA polymerase mutants with altered substrate specificity and low activity can benefit from

stabilizing mutations, which increase overall activity without affecting substrate preference.

RESULTS AND DISCUSSION

Analysis of the effects of stabilizing mutations on the activity of theT7 RNA polymerase substrate specificity mutant G542V, H784S

The Ellington Lab's previous selection for RNA polymerases with altered substrate specificity (Chelliserrykattil and Ellington, 2004) focused on the four amino acids that are proximal to the incoming nucleotide, (R425, G542, Y639, and H784), and thus likely played a role in substrate recognition (Cheetham and Steitz, 1999; Temiakov *et al*., 2004). One of the resulting mutants, designated "VRS," could incorporate 2′-F-modified pyrimidines. VRS had mutations at two of the randomized residues (*i.e.* G542V and H784S). Interestingly, the H772R mutation of VRS arose during the selection despite H772 not being one of the randomized residues. Data alluded to in the initial publication (but not shown) suggest that H772R was not important.

It was decided to investigate the G542V, H784S double mutant ("VS") as an alternative to Y639F for the generation of 2′-F-modified RNA. The enzyme was purified and tested for its ability to polymerize RNA composed of either natural NTPs (rN) or ribopurines and 2′-F-pyrimidines (rRfY). Unexpectedly, the enzyme was found to be almost completely inactive.

In order to understand the reason for the inactivity of the VS variant and the potential role of the H772R mutation, VRS was cloned, purified and tested. Real-time polymerase activity was assayed using the fluorescent aptamer Spinach (Paige *et al*., 2011) (**Figure 4-2**), which was found to fluoresce as a purely ribo-aptamer and when substituted with 2′-F-pyrimidines, although fluorescence per molecule is about 3-fold

lower in the rRfY form. It can be seen that both VRS and VS are more tolerant of 2′-F-pyrimidines than wild type T7 RNA polymerase. It is also evident that VRS is more active than VS in transcriptions using NTPs as well as 2′-F-pyrimidines.

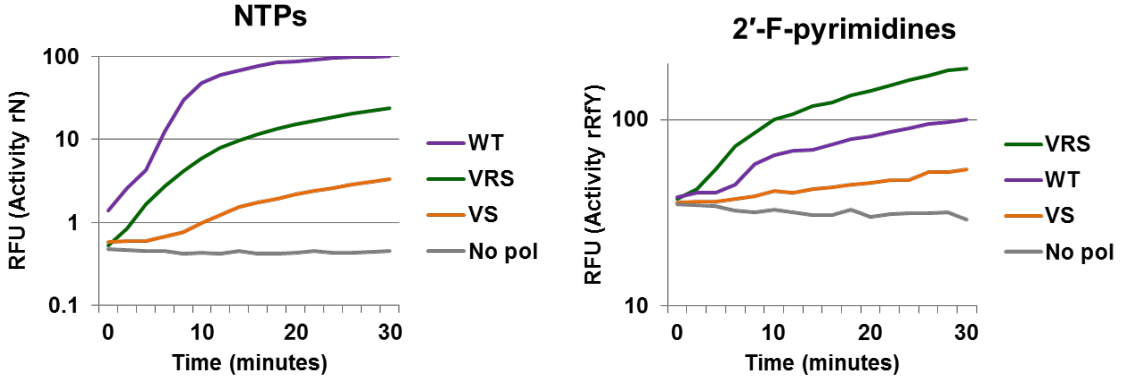


Figure 4-2. The effects of the H772R mutation on the activity of the VRS mutant. **Left**) Real time measurement of Spinach fluorescence (in Relative Fluorescence Units) from ribonucleotide (rN) transcription. **Right**) Real time measurement of Spinach fluorescence (in Relative Fluorescence Units) from 2′-F-pyrimidine (rRfY) transcription. All readings are normalized to the output of wild type T7 RNA polymerase (WT) at 30 minutes defined as 100; numbers are not comparable between graphs.

Since VS showed a decrease in activity for each substrate composition, it was hypothesized that H772R contributes to the overall activity of VRS, apart from any substrate preference considerations. H772R is not near the substrate recognition domain, but it has been observed in other published selections for T7 RNA polymerase activity (Ellefson *et al*., 2014; Dickinson *et al*., 2013) and several of the selections detailed in **Chapter 3**. This suggests that H772R may be a general stabilizer, and that it would increase the overall activity of the (possibly unstable) G542V, H784S variant. Several derivatives of VRS with suspected stabilizing or enhancing mutations were cloned, purified, and tested (**Table 4-2**).

Enzyme	Sequence
WT	WT T7 RNAP
VS	G542V, H784S
VRS	G542V, H772R, H784S
VRS-L	P266L, G542V, H772R, H784S
VLRS	G542V, V625L, H772R, H784S
VRIS	G542V, H772R, V783I, H784S
VLRIS	G542V, V625L, H772R, V783I, H784S
VRSY	G542V, H772R, H784S, F880Y
VRS-M5	S430P, N433T, G542V, S633P, H772R, H784S, F849I, F880Y
VRS-M6	P266L, S430P, N433T, G542V, S633P, H772R, H784S, F849I, F880Y

Table 4-2. Potential stabilization mutants of the T7 RNA polymerase nucleotide specificity mutant, VRS.

The list of potentially stabilizing mutations includes V625L and V783I which were suspected to confer stability to T7 RNA polymerase. Briefly, the Havranek Lab provided a list of mutations predicted by the Rosetta software to allow tighter packing of the T7 RNA polymerase hydrophobic core. Single mutants of the 10 highest scoring mutations were made and tested *in vivo*. Two mutations that abolished activity were eliminated from consideration. Several combinations of the remaining eight mutations were made and then gene shuffled together. Two rounds of selection by compartmentalized partnered replication (CPR) were then performed. Three mutations A288V, V625L, and V783I occurred in more than half of the resulting mutants. Concurrently, wild type T7 RNA polymerase was subjected to error-prone PCR and several rounds of selection by CPR. 20 total clones were sequenced after Rounds 3 and 4, revealing T7 RNA polymerase mutants with between two and 10 mutations each. Two mutants contained V625L and one contained V783I. The possibility that these two mutations may confer stability to VRS was therefore considered.

The P266L mutation was included because it has been shown to facilitate promote clearance (Guillerez *et al*., 2005). It has also anecdotally been associated with enhanced incorporation of unnatural nucleotides.

The so-called "M5" (S430P, N433T, S633P, F849I, and F880Y) sets of mutations was considered based on patent literature suggesting that it confers thermal stability to T7 RNA polymerase (Liao *et al*., 2003; Sugiyama *et al*., 2009). "M6" (P266L, S430P, N433T, S633P, F849I, and F880Y) combines the M5 mutations with P266L. F880Y was considered independently of the M5 mutations because it seemed to enhance the autogene selections (Davidson *et al*., 2012)(**Chapter 2**).

The mutants listed in **Table 4-2** were cloned, expressed, purified and used for *in vitro* transcription of the Spinach aptamer. The *in vitro* transcription reactions contained either all four NTPs (rN) or ribopurines and 2′-F-pyrimidines (rRfY) (**Figure 4-3**). The M5 and M6 mutations increase the activity of the VRS parent in each case.

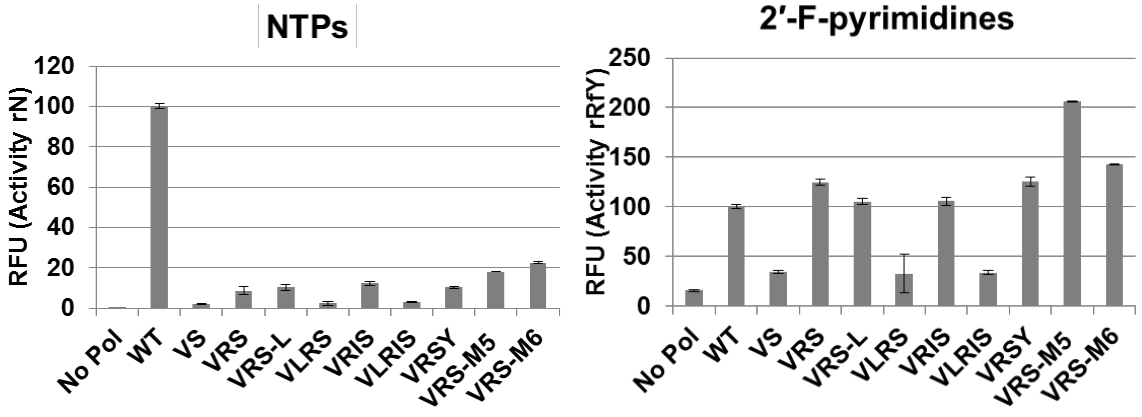


Figure 4-3. The effects of stabilizing mutations on the activity of the VRS mutant. **Left**) Measurement of ribonucleotide (rN) transcriptional output. **Right**) Measurement of 2′-F-pyrimidine (rRfY) transcriptional output. Real time readings of Spinach fluorescence (in Relative Fluorescence Units) were taken; the readings at three hours are shown. All readings are normalized to wild type T7 RNA polymerase (WT) defined as 100; numbers are not comparable between graphs. Error bars represent standard error resulting from three independently assembled reactions.

The “M5” mutations have only been described in patent literature. Patent claims suggest that the mutations arose in a T7 RNA polymerase selection for transcriptional activity at higher temperatures, described in C**hapter 1**. A random mutagenesis library of T7 RNA polymerase mutants was tasked with driving chloramphenicol acetyltransferase expression in *Geobacillus stearothermophilus* at 45 °C to 50 °C. Surviving hosts had their T7 RNA polymerase expression plasmids sequenced. The mutation S633P was described first (Liao *et al*., 2003), followed by the mutations S430P, F849I, and F880Y (Sugiyama *et al*., 2009). The quadruple mutant combining those mutations is reported to have an increased half-life at 50 °C and increased transcription at that temperature. The primers reportedly used to generate this “quadruple” mutant confer an additional N433T mutation. It is unclear if the patent holders meant to include this, as it is not discussed. Regardless, the “M5” (S430P, N433T, S633P, F849I, and F880Y) mutations increased the activity of VRS with the rN and rRfY formulations, as did the “M6” mutations, which includes the P266L mutation on top of the M5 mutations. The M5 are M6 mutations are potentially effective stabilizers and could be used to increase the activity of the substrate specificity mutants listed in **Table 4-1**.

Analysis of the effects of the M5 mutations on the activity of T7 RNA polymerase substrate specificity mutants

The Yin Lab has previously solved the crystal structure of the transcribing M5 T7 RNA polymerase mutant in the initiation complex (unpublished work). Interestingly, there are few gross morphological differences to the comparable wild type T7 RNA polymerase crystal (Cheetham and Steitz, 1999) (**Figure 4-4**). There is, however, an added hydrogen bond made by F880Y, which may stabilize the two halves of the palm domain. It should be noted that the F880Y mutation is not sufficient to increase VRS activity (see VRSY in **Figure 4-3**). It should also be noted that T7 RNA polymerase

undergoes two major conformational changes. It is therefore possible that the M5 mutations cause large structural changes in one of the other conformations.

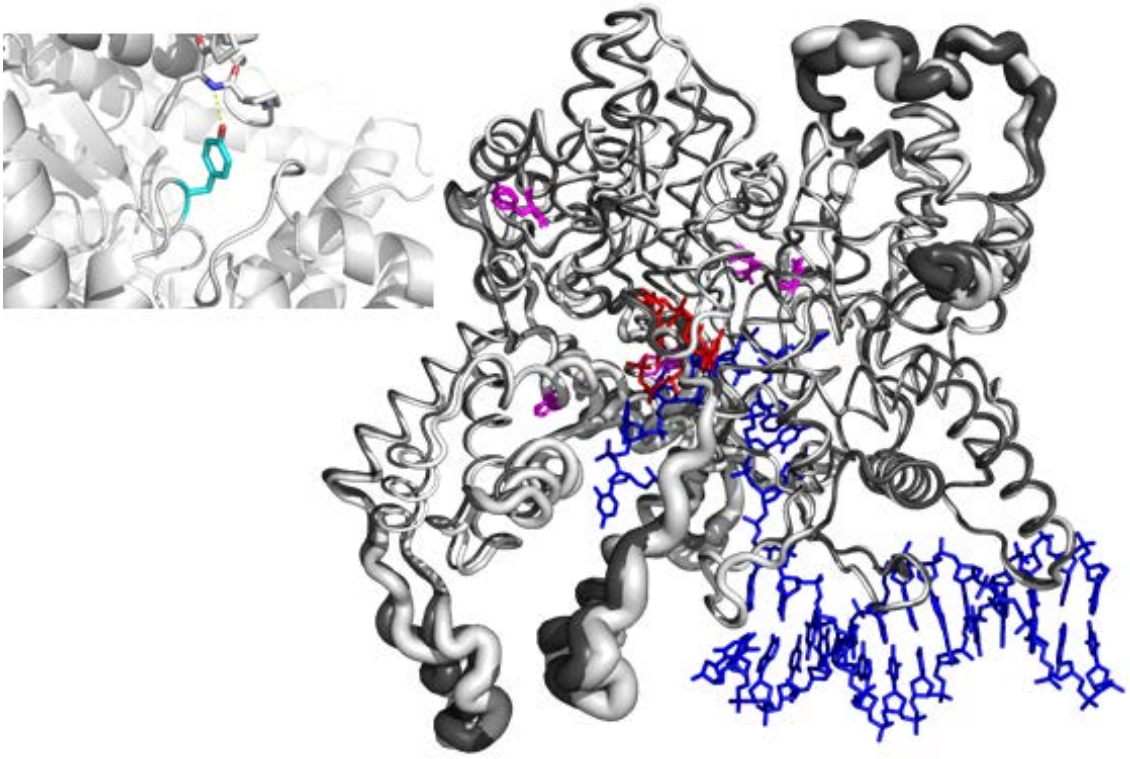


Figure 4-4. Structure of the transcribing "M5" RNA polymerase initiation complex. The M5 T7 RNA polymerase (white) overlayed with wild type T7 RNA polymerase (dark gray, PDB accession number 1QLN (Cheetham and Steitz, 1999). Amino acid changes are shown in magenta, DNA in blue, RNA in red. The B-factor, a measure of how ordered a region is, is shown. **Inset**) The added hydroxyl group resulting from the F880Y mutation (cyan) forms a hydrogen bond (yellow dashed line) with the peptide backbone between P474 and F475.

To determine whether the M5 and M6 mutations could increase the activity of other T7 RNA polymerase mutants, they were added to several "parental" mutants. Several polymerases with varying specificity for different ribose moieties, namely wild type T7 RNA polymerase (WT), Y639F, FA, RGVG, VRS, and R425C (**Table 4-1 and 4-3**) were cloned with and without the P266L mutation. Each of these 12 "parental" mutants was then mutated to contain the M5 mutations (the combination of M5 and

P266L forms M6). For comparison, a recently described mutant, 2P16, which is likely a stabilized version of RGVG, was also included. These 25 polymerases were purified and assayed for transcriptional activity with canonical NTPs *in vitro* (**Table 4-3, Figure 4-5**). The M5 mutations increase the transcriptional activity of each parental polymerase.

Enzyme	Sequence
WT	WT T7 RNAP
M5	S430P, N433T, S633P, F849I, F880Y
L	P266L
M6	P266L, S430P, N433T, S633P, F849I, F880Y
Y639F	Y639F
Y639F-M5	S430P, N433T, S633P, Y639F, F849I, F880Y
Y639F-L	P266L, Y639F
Y639F-M6	P266L, S430P, N433T, S633P, Y639F, F849I, F880Y
FA	Y639F, H784A
FA-M5	S430P, N433T, S633P, Y639F, H784A, F849I, F880Y
FA-L	P266L, Y639F, H784A
FA-M6	P266L, S430P, N433T, S633P, Y639F, H784A, F849I, F880Y
VRS	G542V, H772R, H784S
VRS-L	P266L, G542V, H772R, H784S
VRS-M5	S430P, N433T, G542V, S633P, H772R, H784S, F849I, F880Y
VRS-M6	P266L, S430P, N433T, G542V, S633P, H772R, H784S, F849I, F880Y
R425C	R425C
R425C-M5	R425C, S430P, N433T, S633P, F849I, F880Y
R425C-L	P266L, R425C
R425C-M6	P266L, R425C, S430P, N433T, S633P, F849I, F880Y
RGVG	E593G, Y639V, V685A, H784G
RGVG-M5	S430P, N433T, E593G, S633P, Y639V, V685A, H784G, F849I, F880Y
RGVG-L	P266L, E593G, Y639V, V685A, H784G
RGVG-M6	P266L, S430P, N433T, E593G, S633P, Y639V, V685A, H784G, F849I, F880Y
2P16	I119V, G225S, K333N, D366N, F400L, E593G, Y639V, S661G, V685A, H784G, F880Y

Table 4-3. T7 RNA polymerase nucleotide specificity mutants with M5 mutations.

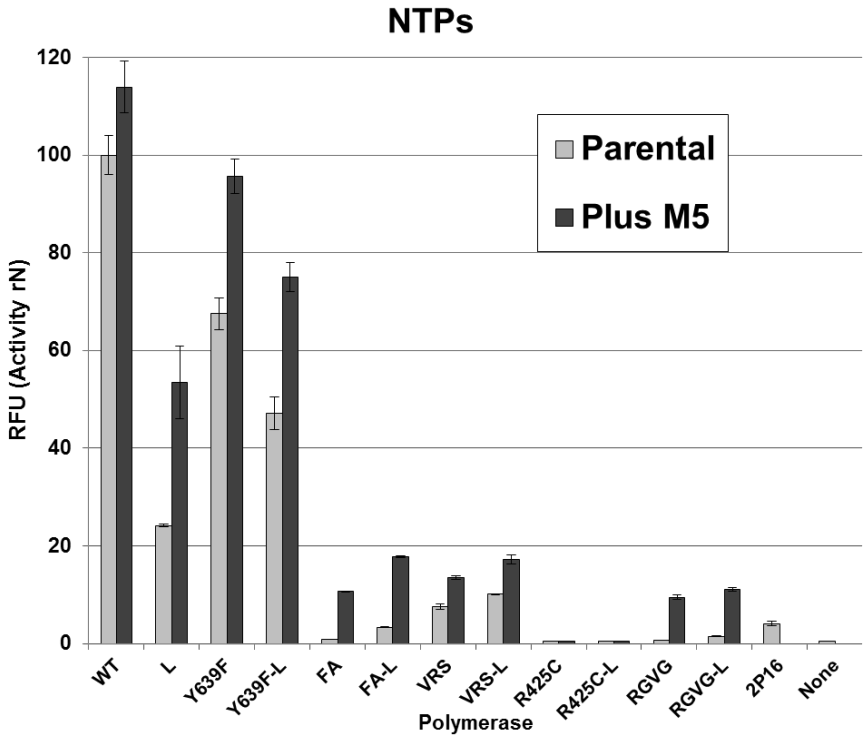


Figure 4-5. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe canonical NTPs.
Measurement of ribonucleotide (rN) transcriptional output. Real time readings of Spinach fluorescence (in Relative Fluorescence Units) were taken; the readings at one hour are shown. All readings are normalized to wild type T7 RNA polymerase (WT) defined as 100. Error bars represent standard error resulting from three independently assembled reactions.

To assess whether the addition of the M5 mutations affected the substrate recognition characteristics of the substrate activity mutants, each mutant was assayed for transcriptional activity with ribopurines and 2′-F-pyrimidines. Interestingly, wild type T7 RNA polymerase and the Y639F mutant did not benefit from M5 mutations in this context. In contrast, the FA, FA-L, VRS, RGVG, and RGVG-L variants all saw an increase in activity upon addition of the M5 mutations. These results suggest that the M5 mutations can increase activity, without altering specificity.

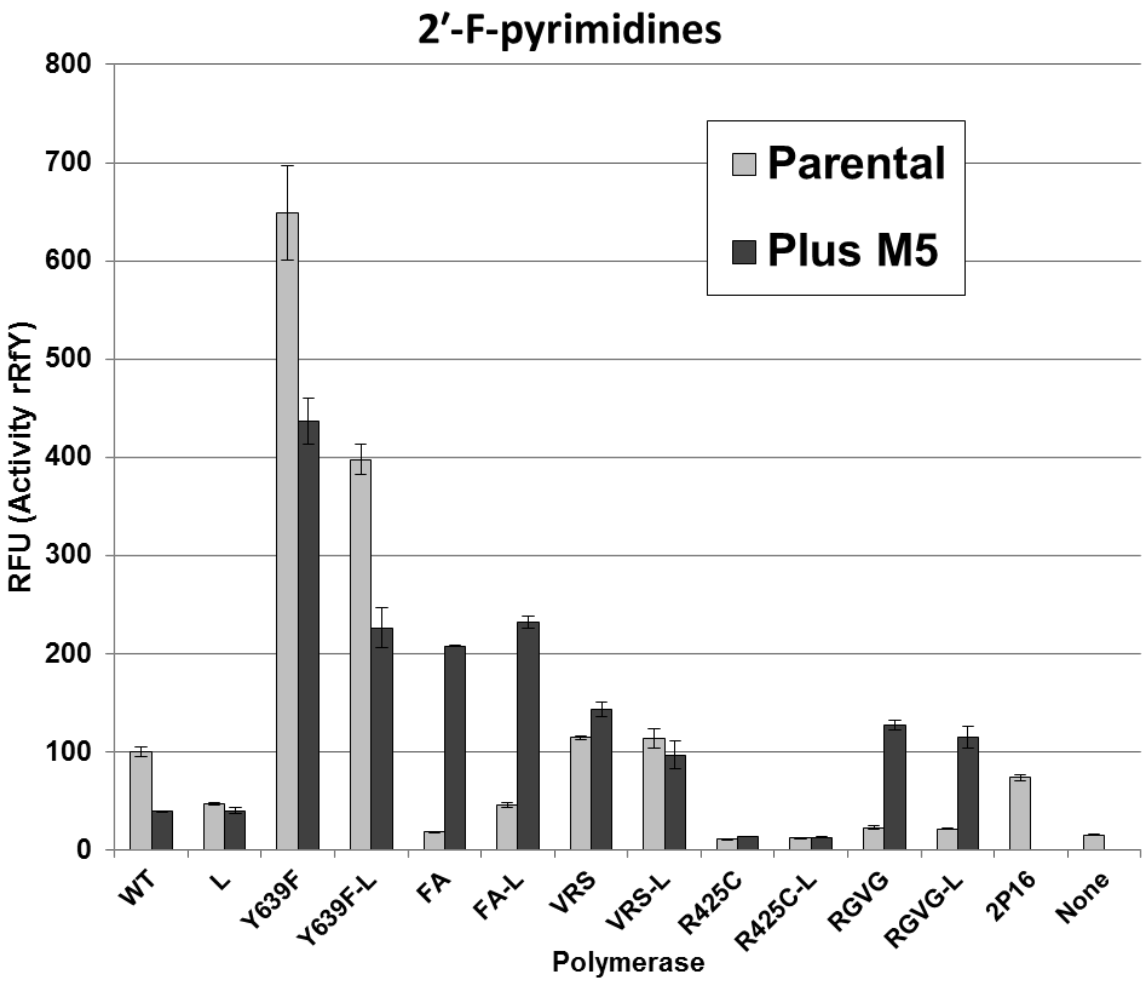


Figure 4-6. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe 2′-F-pyrimidines.

Measurement of 2′-F-pyrimidine (rRfY) transcriptional output. Real time readings of Spinach fluorescence (in Relative Fluorescence Units) were taken; the readings at two hours are shown. All readings are normalized to wild type T7 RNA polymerase (WT) defined as 100. Error bars represent standard error resulting from three independently assembled reactions.

Analysis of the effects of the M5 and M6 mutations on the incorporation of 2′-*O*-methyl NTPs

While the activities of several mutants in an rRfY context were improved by the M5 mutations, none matched the activity of Y639F, which is the standard T7 RNA

polymerase variant used to incorporate 2′-F-pyrimidines and was not improved by the M5 mutations. Additionally, the mutants most improved by the M5 mutations (FA and RGVG) are most known for their ability to incorporate 2′-*O*-methylnucleotides. Nucleotide formulations containing 2′-*O*-methylnucleotides were therefore focused upon.

Each enzyme was assayed for its ability to transcribe 2′-*O*-methyluridine along with ATP, CTP, and GTP (**Figure 4-7**). As before, FA, FA-L, RGVG, and RGVG-L all saw increases in activity when the M5 mutations were included. The VRS family of enzymes showed no activity in the rVmU context, consistent with previously reports (Chelliserrykattil and Ellington, 2004). Interestingly, transcript from the mutants based on R425C could not be detected, in marked contrast to previously published results (Ibach *et al*., 2013). 2P16 is more active than RGVG as previously claimed (Siegmund *et al*., 2012), but is not as active as either RGVG-M5 or RGVG-M6.

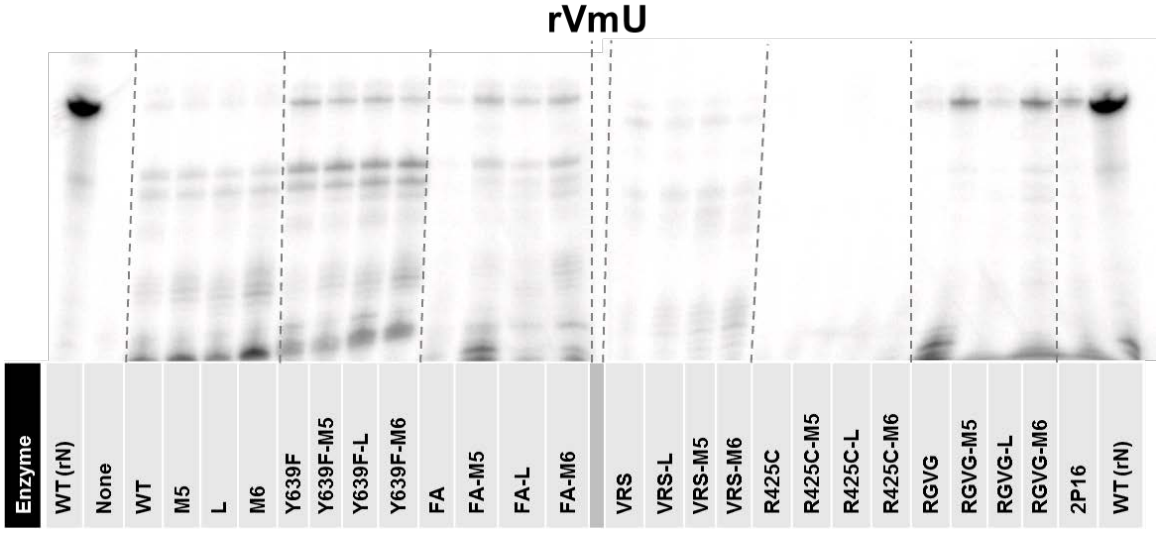


Figure 4-7. The effects of the M5 mutations on the ability of T7 RNA polymerase specificity mutants to transcribe 2′-*O*-methyluridine.
Transcription assay for the incorporation of 2′-*O*-methyluridine (rVmU). Transcripts were labelled by inclusion of [α^{32}P]ATP and analyzed by denaturing PAGE. A reaction of wild type T7 RNA polymerase (WT) with ribonucleotides (rN) is included for comparison. Transcriptions ran for four hours, two distinct gels are shown.

Based on the above results, further assays were limited to a subset of the most active polymerases and those that serve as suitable comparisons. This subset was re-assayed for the ability to incorporate 2′-*O*-methyluridine (rVmU); running samples on a single gel to allow for quantitative comparison (**Figure 4-8**). This was followed by a similar assay for the incorporation of 2′-*O*-methylpyrimidines (rRmY, **Figure 4-9**), and 2′-*O*-methyladenosine and 2′-*O*-methylpyrimidines (rGmH) (**Figure 4-10**). As before, the M5 mutations enhanced the activity of the FA and RGVG enzymes for each set of substrates. RGVG-M6 was the most active enzyme in all conditions, yielding at least 25-fold more RNA than the FA mutant, which is the most commonly used enzyme for generating 2′-*O*-methyl RNA.

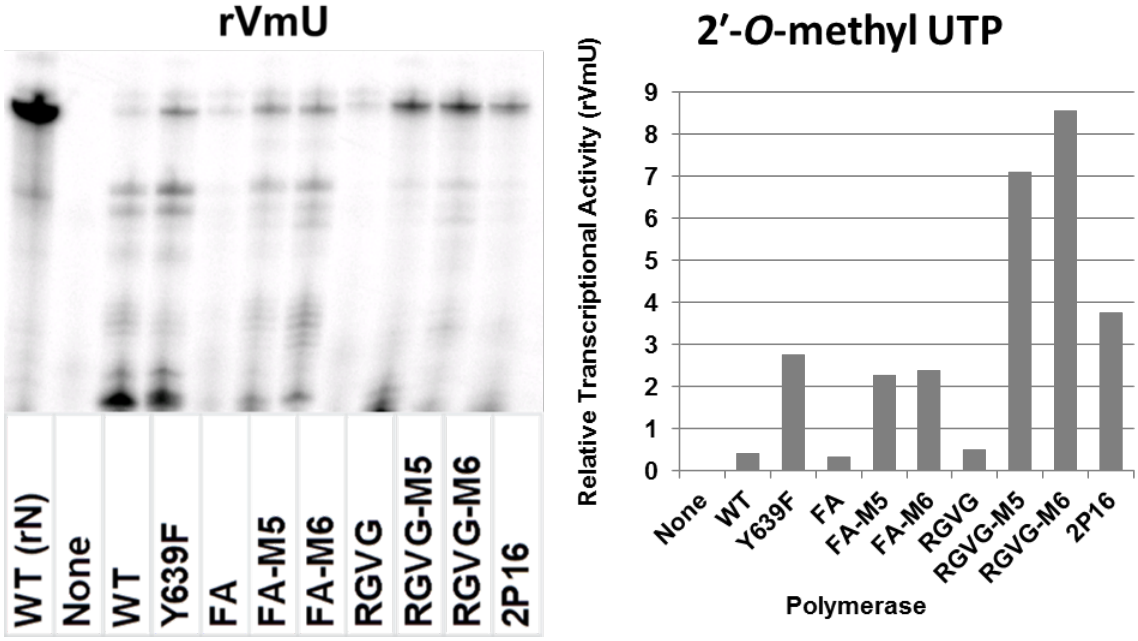


Figure 4-8. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methyluridine.

Transcription assay for the incorporation of 2′-*O*-methyluridine (rVmU). Transcripts were labelled by inclusion of [α^{32}P]ATP and analyzed by denaturing PAGE. All values are normalized to 100, representing the yield of wild type T7 RNA polymerase with ribonucleotides (rN). Transcriptions ran for four hours.

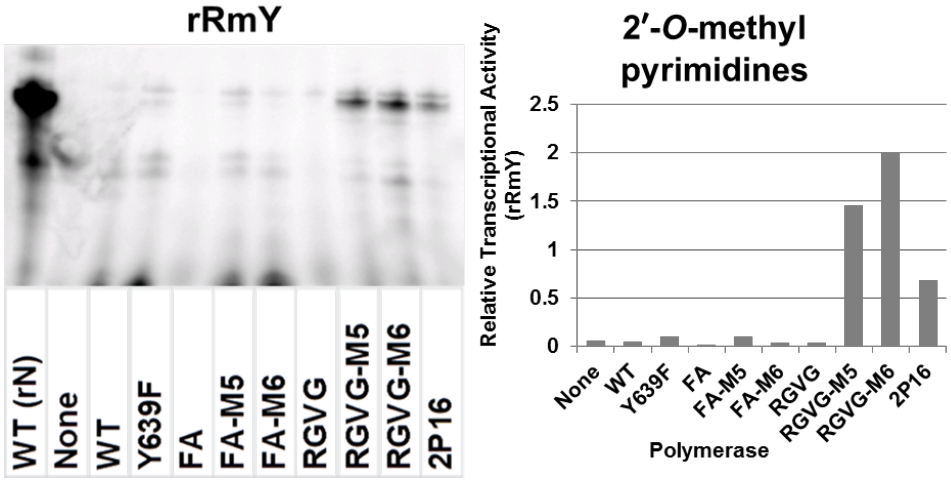


Figure 4-9. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methylpyrimidines.

Transcription assay for the incorporation of 2′-*O*-methylpyrimidines (rRmY). Transcripts were labelled by inclusion of [α^{32}P]ATP and analyzed by denaturing PAGE. All values are normalized to 100, representing the yield of wild type T7 RNA polymerase with ribonucleotides (rN). Transcriptions ran for four hours.

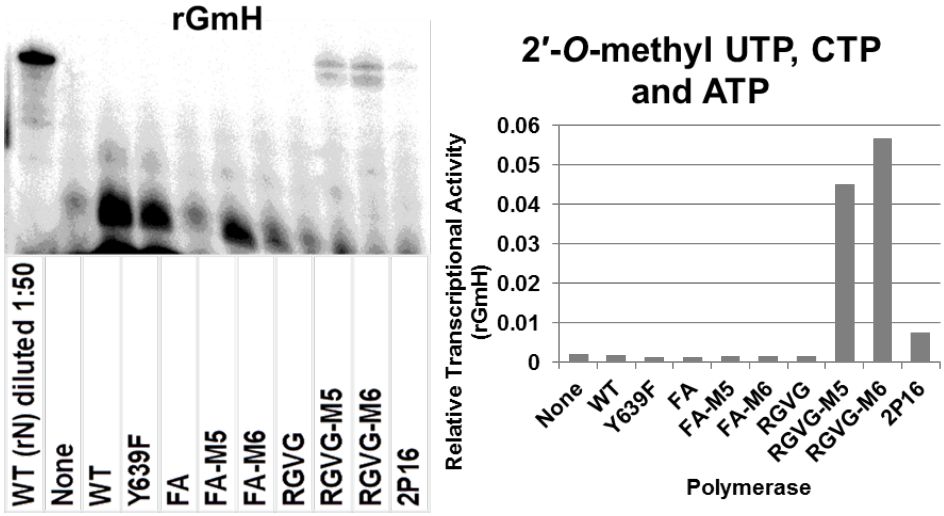


Figure 4-10. Quantitative comparison of T7 RNA polymerase specificity mutants' incorporation of 2′-*O*-methyladenosine and 2′-*O*-methylpyrimidines.

Transcription assay for the incorporation of 2′-*O*-methyladenosine and 2′-*O*-methylpyrimidines (rGmH). Transcripts were labelled by inclusion of [α^{32}P]GTP and analyzed by denaturing PAGE. All values are normalized to 100, representing the yield of wild type T7 RNA polymerase with ribonucleotides (rN). Transcriptions ran for 20 hours.

Analysis of the ability of the T7 RNA polymerase RGVG-M6 to transcribe heavily-modified RNA

After it was demonstrated that RGVG-M6 could catalyze the formation of RNA containing up to three 2′-*O*-methylnucleotides, RGVG-M6 was assayed for its ability to generate fully-modified RNA (**Figure 4-11**). RGVG-M6 was able to polymerize fully-modified RNA of two different compositions. A high yield of 2′-F-purines and 2′-*O*-methylpyrimidines (fRmY) was produced, as was a moderate amount of 2′-F-guanosine, 2′-*O*-methyladenosine, and 2′-*O*-methylpyrimidines (fGmH) RNA. Fully 2′-*O*-methyl RNA (mN) was not obtained.

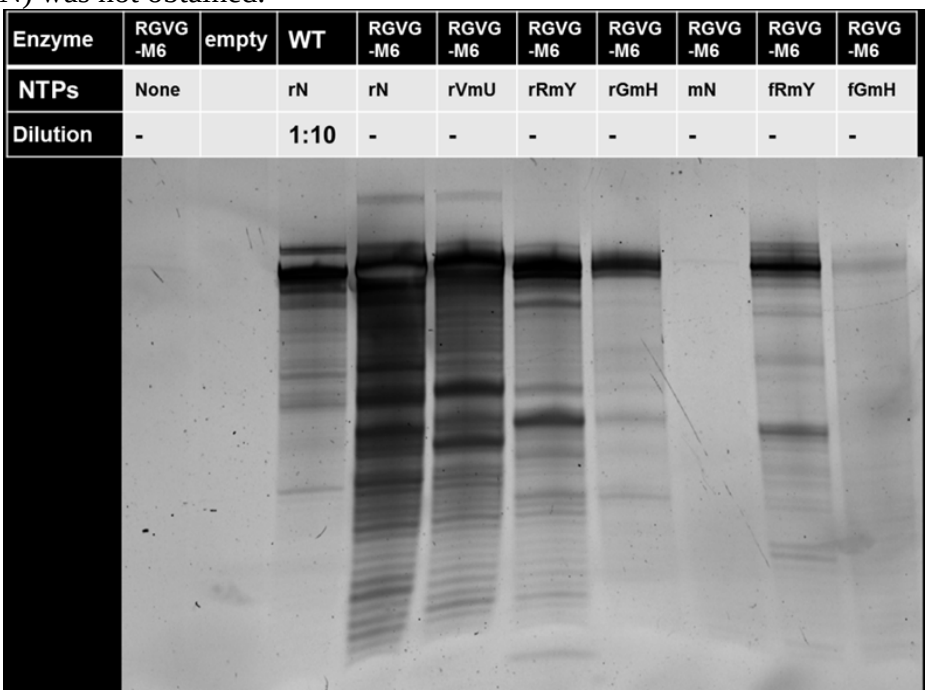
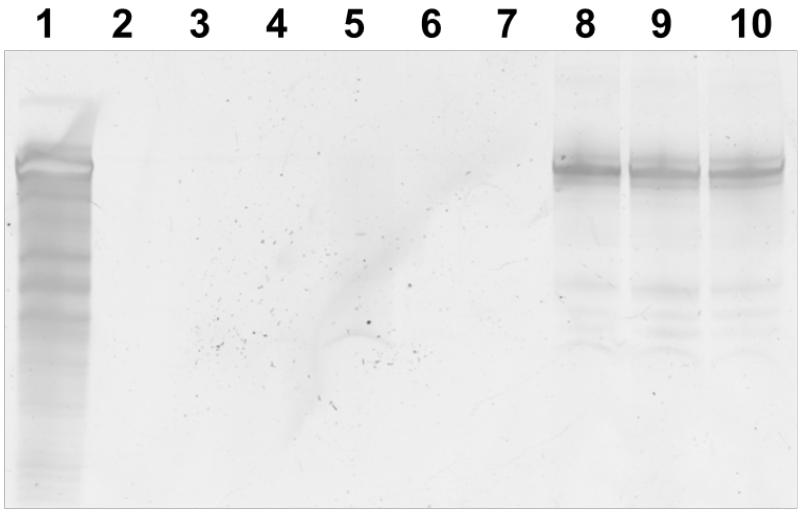


Figure 4-11. Transcription of various nucleotide compositions by RGVG-M6. Transcription assay for RGVG-M6 catalyzed incorporation of ribonucleotides (rN); 2′-*O*-methyluridine (rVmU); 2′-*O*-methylpyrimidines (rRmY); 2′-*O*-methyladenosine and 2′-*O*-methylpyrimidines (rGmH); 2′-*O*-methylnucleotides (mN); 2′-F-purines and 2′-*O*-methylpyrimidines (fRmY); and 2′-F-guanosine, 2′-*O*-methyladenosine, and 2′-*O*-methylpyrimidines (fGmH). Transcripts were analyzed by denaturing PAGE and imaged after staining in SYBR-Gold. Transcriptions ran for 20 hours. A reaction (10-fold diluted) containing wild type T7 RNA polymerase with ribonucleotides (rN) is shown for comparison.

Previous reports of mN incorporation (Burmeister *et al*., 2005; Ibach *et al*., 2013) have used more permissive buffer compositions, including manganese as well as rGMP, rGTP, or both. The ability of RGVG-M6 to polymerize mN RNA was tested in several such permissive buffers (**Figure 4-12**). High yields of RNA were produced in three similar buffers containing all four 2′-*O*-methylnucleotides, namely buffers 8, 9, and 10. Buffer 9 contained rGMP and buffer 10 contained rGMP and rGTP and thus cannot be said to allow transcription of fully 2′-*O*-methyl RNA. Buffer 8 contains manganese, yeast inorganic pyrophosphates (YIPP), Triton X-100, and polyethylene glycol (PEG). It also differs from standard transcription buffers in that it contains substantially reduced concentrations of spermidine, magnesium, 2′-*O*-methyl NTPs, template, and polymerase, but increased concentration of DTT. Buffer capacity is provided by HEPES instead of Tris.



1
500 nM DNA
40 mM Tris-HCl pH 8.0
30 mM $MgCl_2$
6 mM spermidine
6 mM each NTP
10 mM DTT
500 nM RGVG-M6

2
500 nM DNA
40 mM Tris-HCl pH 8.0
30 mM $MgCl_2$
6 mM spermidine
6 mM each 2′-*O*-methyl-NTP
10 mM DTT
No RGVG-M6

3
500 nM DNA
40 mM Tris-HCl pH 8.0
30 mM $MgCl_2$
6 mM spermidine
6 mM each 2′-*O*-methyl-NTP
10 mM DTT
500 nM RGVG-M6

4
200 nM DNA
200 mM HEPES pH 7.5
8 mM $MgCl_2$
2 mM spermidine
3.75 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
200 nM RGVG-M6

5
200 nM DNA
200 mM HEPES pH 7.5
8 mM $MgCl_2$
2 mM spermidine
3.75 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
200 nM RGVG-M6

6
200 nM DNA
200 mM HEPES pH 7.5
8 mM $MgCl_2$
2 mM spermidine
3.75 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
10% PEG 8000
200 nM RGVG-M6

7
200 nM DNA
200 mM HEPES pH 7.5
8 mM $MgCl_2$
2 mM spermidine
3.75 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
10% PEG 8000
1.5 mM $MnCl_2$
200 nM RGVG-M6

8
200 nM DNA
200 mM HEPES pH 7.5
5.5 mM $MgCl_2$
2 mM spermidine
0.5 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
10% PEG 8000
1.5 mM $MnCl_2$
10 U/ml YIPP
200 nM RGVG-M6

9
200 nM DNA
200 mM HEPES pH 7.5
5.5 mM $MgCl_2$
2 mM spermidine
0.5 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
10% PEG 8000
1.5 mM $MnCl_2$
10 U/ml YIPP
0.5 mM GMP
200 nM RGVG-M6

10
200 nM DNA
200 mM HEPES pH 7.5
5.5 mM $MgCl_2$
2 mM spermidine
0.5 mM each 2′-*O*-methyl-NTP
40 mM DTT
0.01% Triton X-100
10% PEG 8000
1.5 mM $MnCl_2$
10 U/ml YIPP
0.5 mM GMP
30 µM GTP
200 nM RGVG-M6

Figure 4-12. Transcription of 2′-*O*-methylnucleotides by RGVG-M6 in various buffers.

Transcripts were analyzed by denaturing PAGE and visualized after staining in SYBR-Gold. Transcriptions ran for 20 hours. A reaction containing RGVG-M6 with ribonucleotides (rN) is shown for comparison (1). The composition of each reaction is shown.

The panel of enzymes was tested for mN polymerization in this permissive buffer 8 (**Figure 4-13**). R425C was included in the assay as it was reported to work in a similar

buffer with the mN composition. FA-M5 and FA-M6 show an increase in activity relative to the parental FA mutant. RGVG-M5 and RGVG-M6 and 2P16 are markedly improved over the parental RGVG. RGVG-M5 and RGVG-M6 make the most transcript of all variants tested. As previously observed, R425C appears to be almost completely inactive.

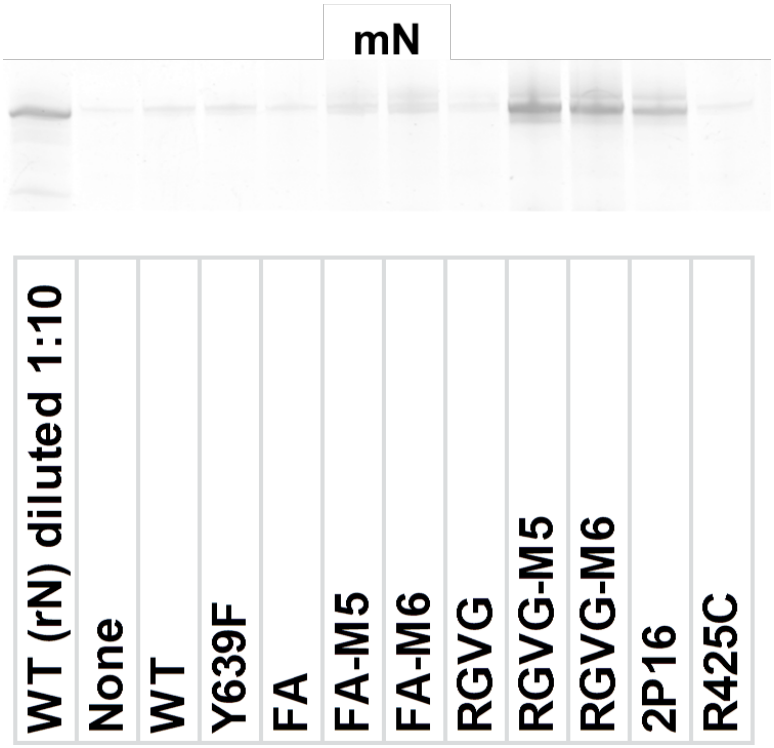


Figure 4-13. Transcription of 2′-*O*-methylnucleotides (mN) in a permissive buffer. Transcription assay for the incorporation of 2′-*O*-methylnucleotides (mN). Transcripts were analyzed by denaturing PAGE and visualized after staining in SYBR-Gold. Transcriptions ran for 20 hours. A reaction (diluted 10-fold) containing wild type T7 RNA polymerase with ribonucleotides (rN) is shown for comparison.

CONCLUSION

As an evolving (or engineered) protein explores the sequence and fitness landscapes, there is an inherent trade-off between activity and stability (Wang *et al*., 2002; Tokuriki *et al*., 2008; Romero and Arnold, 2009; Levin *et al*., 2009; Soskine and

Tawfik, 2010). As enzyme engineers endeavor to create proteins further removed from their wild type parents, they will be confronted by the limitations of the parental protein scaffold. The inclusion of global stabilizing mutations into such a scaffold makes it more accommodating to mutations that confer phenotypes of biotechnological importance. This may be used as a generalizable tactic for future protein engineering efforts.

RGVG-M6 is the most effective known enzyme for the polymerization of 2′-modified RNA. It is capable of generating transcripts containing 2′-fluoro and/or 2′-*O*-methyl NTPs in high yields and is capable of polymerizing nucleic acids consisting entirely of modified nucleotides. This enzyme should prove useful in creating modified nucleic acids for *in vitro* selection (Knudsen *et al*., 2002; Keefe and Cload, 2008) as well as therapeutic uses (Dean and Bennett, 2003; Healy *et al*., 2004; Wilson and Keefe, 2006).

The inclusion of P266L in the assayed mutants only modestly affects the yield of full length RNA. P266L tends to reduce overall RNA yield, but has been associated with enhanced promoter clearance and fewer abortive products, especially when suboptimal initially transcribed sequences (ITS) are used (Guillerez *et al*., 2005). Since premature termination, especially in the ITS, is a major problem for unnatural nucleotide polymerization, P266L may prove beneficial in certain contexts. The choice of whether to use an M5 or M6 (containing P266L) mutant depends on the DNA template, nucleotide composition, and intended use of the RNA product.

MATERIALS AND METHODS

Preparation of T7 RNA polymerase variants

The T7 RNA polymerase coding sequence was cloned into pQE-80L (Qiagen). All T7 RNA polymerase variants were derived from this plasmid either by Mega-primer

PCR (Bryksin and Matsumura, 2010) or Isothermal assembly (Gibson, 2011). Plasmids were transformed into BL21-gold *E. coli* cells (Agilent). Cells were grown in 2xYT media at 37 °C overnight. Subcultures were grown at 37 °C until reaching OD600 ~0.7-0.8 at which point 1 mM IPTG was added. Cells were grown for four hours at 37 °C, pelleted, and frozen at -80 °C. Pellets were resuspended in binding buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole). Resuspended cells were lysed via sonication on ice using 50% probe amplitude for three minutes (1 sec ON, 1 sec OFF). Cell debris was pelleted by centrifugation at 10,000 g for 30 minutes. His-tagged T7 RNA polymerase was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over 1 ml (bead volume) Ni-NTA (Fisher) gravity column pre-equilibrated with binding buffer. The column was washed with 10 column volumes of wash buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 20 mM imidazole). T7 RNA polymerase was eluted off the column by the addition of three column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 250 mM imidazole). Eluted T7 RNA polymerase was dialyzed against the final storage buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DDT, 1 mM EDTA). Dialates were adjusted to 1 mg/ml and added to an equal volume of glycerol (final concentration 0.5 mg/ml).

In vitro transcription assays

Real-time transcription reactions contained 40 mM Tris-HCl pH 8.0, 30 mM $MgCl_2$, 6 mM spermidine, 6 mM each NTP (or modified NTP), 10 mM DTT, 500 mM T7 RNA polymerase, 500 mM DNA template, and 0.17 mg/ml DFHBI (in DMSO) (Paige *et al*., 2011). Reactions were incubated for up to four hours at 37 °C with Spinach fluorescence (Excitation/ Emission 469 nm/501 nm) readings taken every one to four minutes in a Safire monochromator (Tecan).

Spinach templates were made by thermal cycling overlapping primers (5′-AATATAATACGACTCACTATAGAGGAGACTGAAATGGTGAAGGACGGGTCCAGTGCTTCG and 5′-GAAAAGACTAGTTACGGAGCTCACACTCTACTCAACA-GTGCCGAAGCACTGGACCCG) with Accuprime Pfx DNA polymerase (Life Technologies) in its standard buffer (94 °C:2 min, 12 cycles [94 °C:15 sec, 50 °C:30 sec, 68 °C:30 sec], 68 °C:1 min). Templates were purified by QIAquick Gel Extraction Kit (Qiagen).

End point transcription reactions contained 40 mM Tris-HCl pH 8.0, 30 mM $MgCl_2$, 6 mM spermidine, 6 mM each NTP (or modified NTP), 10 mM DTT, 500 mM T7 RNA polymerase, 500 mM DNA template. Reactions were incubated for four or 20 hours at 37 °C. DNA templates were made as above. rVmU reactions and rRmY reaction were run for four hours, labelled by inclusion of 0.17 µM [α^{32}P]ATP (3000 Ci/mMol,) and analyzed by denaturing PAGE. rGmH reactions were run for 20 hours, labelled by inclusion of 0.17 µM [α^{32}P]GTP (3000 Ci/mMol,) and analyzed by denaturing PAGE. RGVG-M6 reactions with varied nucleotide composition were run twenty hours, incubated for one hour at 37 °C with 0.03 U/µl Baseline-ZERO DNase in its supplied buffer, analyzed by denaturing PAGE and visualized after staining in SYBR-Gold. The buffer comparison experiment using fully 2′-*O*-methl NTPs used the buffers listed in **Figure 4-12**, was run for 20 hours, was incubated for one hour 37 °C with 0.03 U/µl Baseline-ZERO DNase in its supplied buffer, was analyzed by denaturing PAGE and was imaged after staining in SYBR-Gold.

mN in the permissive buffer contained 200 mM HEPES pH 7.5, 5.5 mM $MgCl_2$, 2 mM spermidine, 0.5 mM each 2′-*O*-methyl-NTP, 40 mM DTT, 0.01% Triton, 10% PEG8000, 1.5 mM $MnCl_2$, 10 U/ml YIPP, 200 nM RNA polymerase, and 200 nM DNA. Reactions were run for 20 hours, incubated for one hour at 37 °C with 0.03 U/µl

Baseline-ZERO DNase in its supplied buffer, analyzed by denaturing PAGE and visualized after staining in SYBR-Gold.

^{32}P gels were exposed on a storage phosphor screen (Molecular Dynamics) before imaging on a STORM 840 Phosphoimager (GE Healthcare). Autoradiographs were analyzed using ImageQuant (GE Healthcare). SYBR-Gold gels were directly visualized on a STORM 840 Phosphoimager.

REFERENCES

Beaudry, A., DeFoe, J., Zinnen, S., Burgin, A., and Beigelman, L. 2000. In vitro selection of a novel nuclease-resistant RNA phosphodiesterase. *Chemistry & Biology* 7:323–34.

Brieba, L. G., and Sousa, R. 2000. Roles of histidine 784 and tyrosine 639 in ribose discrimination by T7 RNA polymerase. *Biochemistry* 39:919–23.

Bryksin, A. V, and Matsumura, I. 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* 48:463–5.

Burmeister, P. E., Lewis, S. D., Silva, R. F., Preiss, J. R., Horwitz, L. R., Pendergrast, P. S., McCauley, T. G., Kurz, J. C., Epstein, D. M., Wilson, C., et al. 2005. Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chemistry & Biology* 12:25–33.

Cheetham, G. M., and Steitz, T. A. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286:2305–9.

Chelliserrykattil, J., and Ellington, A. D. 2004. Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nature Biotechnology* 22:1155–60.

Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. 2012. An in vitro Autogene. *ACS Synthetic Biology* 1:190–196.

Dean, N. M., and Bennett, C. F. 2003. Antisense oligonucleotide-based therapeutics for cancer. *Oncogene* 22:9087–96.

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. 2013. Experimental interrogation of the path dependence and stochasticity of protein

evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*.

Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology* 32:97–101.

Ellington, A. D., and Szostak, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.

Gibson, D. G. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* 498:349–61.

Guillerez, J., Lopez, P. J., Proux, F., Launay, H., and Dreyfus, M. 2005. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America* 102:5958–63.

Healy, J. M., Lewis, S. D., Kurz, M., Boomer, R. M., Thompson, K. M., Wilson, C., and McCauley, T. G. 2004. Pharmacokinetics and biodistribution of novel aptamer compositions. *Pharmaceutical Research* 21:2234–46.

Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. 1997. Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* 36:8231–42.

Ibach, J., Dietrich, L., Koopmans, K. R. M., Nöbel, N., Skoupi, M., and Brakmann, S. 2013. Identification of a T7 RNA polymerase variant that permits the enzymatic synthesis of fully 2'-O-methyl-modified RNA. *Journal of Biotechnology* 167:287–95.

Jackson, A. L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J. M., Lim, L., Karpilow, J., Nichols, K., et al. 2006. Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* 12:1197–205.

Keefe, A. D., and Cload, S. T. 2008. SELEX with modified nucleotides. *Current Opinion in Chemical Biology* 12:448–56.

Knudsen, S. M., Robertson, M. P., and Ellington, A. D. 2002. In vitro selection using modified or unnatural nucleotides. *Current Protocols in Nucleic Acid Chemistry* Chapter 9:Unit 9.6.

Kostyuk, D. A., Dragan, S. M., Lyakhov, D. L., Rechinsky, V. O., Tunitskaya, V. L., Chernov, B. K., and Kochetkov, S. N. 1995. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Letters* 369:165–8.

Kraynack, B., and Baker, B. 2006. Small interfering RNAs containing full 2′-O-methylribonucleotide-modified sense strands display Argonaute2/eIF2C2-dependent activity. *RNA*:163–176.

Layzer, J. M. 2004. In vivo activity of nuclease-resistant siRNAs. *RNA* 10:766–771.

Levin, K. B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C., and Tawfik, D. S. 2009. Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nature Structural & Molecular Biology* 16:1049–55.

Liao, H., van Gemen, B., and Sugiyama, A. 2003. Mutant of RNA Polymerases with Increased Stability. 1:U. S. Patent. US 6,524,828 B1.

Lupold, S., Hicke, B., Lin, Y., and Coffey, D. 2002. Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Research*:4029–4033.

Majlessi, M., Nelson, N. C., and Becker, M. M. 1998. Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Research* 26:2224–9.

Padilla, R., and Sousa, R. 2002. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Research* 30:e138.

Paige, J. S., Wu, K. Y., and Jaffrey, S. R. 2011. RNA Mimics of Green Fluorescent Protein. *Science* 333:642–646.

Romero, P. a, and Arnold, F. H. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews. Molecular Cell Biology* 10:866–76.

Siegmund, V., Santner, T., Micura, R., and Marx, A. 2012. Screening mutant libraries of T7 RNA polymerase for candidates with increased acceptance of 2'-modified nucleotides. *Chemical Communications* 48:9870–2.

Soskine, M., and Tawfik, D. S. 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews. Genetics* 11:572–82.

Sousa, R., and Padilla, R. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *The EMBO Journal* 14:4609–21.

Sugiyama, A., Nishiya, Y., and Kawakami, B. 2009. RNA Polymerase Mutants with Increased Thermostability. U. S. Patent. US 7,507,567 B2.

Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassylyev, D. G. 2004. Structural basis for substrate selection by T7 RNA polymerase. *Cell* 116:381–91.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. 2008. How protein stability and new functions trade off. *PLoS Computational Biology* 4:e1000002.

Tuerk, C., and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.

Wang, X., Minasov, G., and Shoichet, B. K. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology* 320:85–95.

Waters, E. K., Genga, R. M., Schwartz, M. C., Nelson, J. A., Schaub, R. G., Olson, K. A., Kurz, J. C., and McGinness, K. E. 2011. Aptamer ARC19499 mediates a procoagulant hemostatic effect by inhibiting tissue factor pathway inhibitor. *Blood* 117:5514–22.

Wilson, C., and Keefe, A. D. 2006. Building oligonucleotide therapeutics using non-natural chemistries. *Current Opinion in Chemical Biology* 10:607–14.

Chapter 5: Lessons, ongoing work, and future directions

The preceding chapters have discussed the evolution and engineering of T7 RNA polymerase. T7 RNA polymerase has been evolved and engineered to alter or enhance various characteristics to make it more suitable for biotechnology purposes. These efforts have been undertaken by multiple research groups, each taking different approaches. The wealth of information this has generated allows us to make several generalizations regarding the nature of molecular evolution and protein function. These can be applied to a range of research goals and provide lessons to future researchers.

LESSONS FROM THE EVOLUTION AND ENGINEERING OF T7 RNA POLYMERASE

Mutation rate and selective advantage

Most mutations are neutral or slightly deleterious, with a substantial fraction being highly deleterious, and beneficial mutations are quite rare (Soskine and Tawfik, 2010). Thus, the evolution of new protein function, which often requires the accumulation of multiple beneficial mutations, represents a major challenge.

As was seen in the selections involving the *in vitro* autogene (**Chapter 2**), a high mutation rate can be quite problematic. This leaves open the question of what constitutes a mutation rate that is too high. This depends on many factors, but the most important one is the rate at which deleterious mutations can be purged from the population. Simply put, if a selection method is capable of purging deleterious mutations from the population as fast as they accumulate, then the population can evolve as desired. There are two ways to accomplish this: (1) have a very low mutation rate, such that the few deleterious mutations that emerge can be easily purged or (2) have a very powerful selection such that mutants free of deleterious mutations can remain a substantial portion of the population.

A low mutation rate may be the easier path to achieve, but this comes at the expense of the chance to make evolutionary jumps in which two or more mutations are needed to achieve a new function. If each individual mutation confers a selective advantage, then stepwise evolution is possible, and a very low mutation rate is acceptable. However, it has been shown that the number of pathways allowing stepwise evolution can be limited (Weinreich *et al*., 2006). If pleiotropy or latent functionality is at work, then stepwise evolution may not be possible, and a low mutation rate will not suffice.

A powerful selection can overcome a high mutation rate, and also has benefits with regard to specificity (discussed in **Chapter 3**). The power of a selection is usually discussed in terms of an enrichment factor, or the extent to which a single functional mutant in a sea of non-functional mutants can increase its representation in a single round. For example, if a starting population has one functional mutant for each 1000 non-functional ones before selection, and the functional mutants represent 50% of the population after one round of selection, then a 500-fold enrichment factor can be claimed. This simple test is a good way to estimate the number of rounds required to find the rare active species in a library. This idea works well in certain contexts; however it oversimplifies the idea of a powerful selection. Enrichment factor as described does not attempt to address the dynamic range of a given selection. That is, can a weakly-active variant outcompete a non-functional variant and, more importantly, can a highly-active variant outcompete a mediocre variant? If a selection method does not sufficiently reward the best members of a population then the hordes of mediocre mutants will dominate the population. This can lead either to the evolution of mediocre enzymes or to the catastrophic collapse of the population, depending on the mutation rate.

Activity and specificity

A powerful selection method that is able to reward the most active mutants has additional benefits with respect to specificity (Khersonsky and Tawfik, 2010). An evolving enzyme faces trade-offs between its new function and its old function. There are documented cases in which the acquisition of a new specificity comes with a sharp decrease in the ancestral activity (strong negative trade-off) (Vick and Gerlt, 2007; McLoughlin and Copley, 2008). However, the vast majority of experiments point to a weak negative trade-off in which a large increase in the new activity comes with a modest reduction in the old activity (Matsumura and Ellington, 2001; Aharoni *et al*., 2005; Khersonsky and Tawfik, 2010; Tokuriki *et al*., 2012). Either model suggests that as the new function is enhanced, the old function diminishes (**Figure 5-1**). Therefore a selection that allows the most active mutants to outcompete moderately active ones will encourage the evolution of specificity.

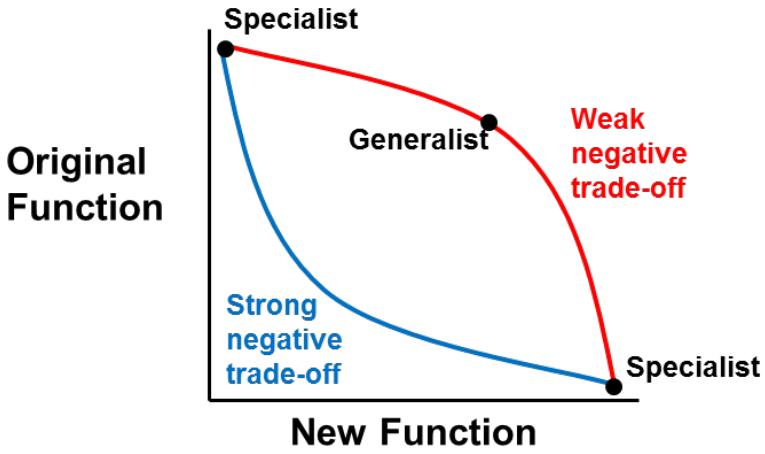


Figure 5-1. Activity and specialization.
During the evolution of a new substrate specificity, an enzyme will lose its ancestral function. In the weak negative trade-off model (red), a large increase in new function is coupled to a weak decrease in old function, resulting in a moderately active generalist. In the strong negative trade-off model (blue) a small increase in the new activity results a large decrease in ancestral activity. This figure is inspired by a review of enzyme promiscuity (Khersonsky and Tawfik, 2010).

Previous efforts in evolving T7 RNA polymerase, and the efforts outlined in **Chapter 3** illustrate this point. The RNA polymerase promoter specificity variants evolved by phage-assisted continuous evolution (PACE) were never subject to strong selective pressure at the higher ends of activity. This allowed mediocre, generalist polymerases to dominate the population (Dickinson *et al*., 2013). It should be noted the emergence of a weak generalist phenotype was accompanied by an increase in expression (caused by mutations to the RBS). This accords with previous observations that strong negative trade-offs are often accompanied by gene duplication or promoter and RBS strengthening (McLoughlin and Copley, 2008; Khersonsky and Tawfik, 2010).

In contrast, T7 RNA polymerase promoter specificity variants evolved by compartmentalized partnered replication (CPR) were subject to powerful selection pressures at the higher ends of activity, and consequently evolved to be both highly active and specific. There are, of course, limits to this model, as the most active polymerases in **Chapter 3** were similar to, but not exactly the same as, the most specific ones.

Assay conditions

As can be seen in **Table 3-2**, the reported and observed activity of T7 RNA polymerases mutants can vary widely. This is due in part to inappropriate assay conditions. In one instance the wild type reference is mutated, resulting in reduced activity, thus increasing the apparent activity of the mutants to which it is compared (Temme *et al*., 2012). A more common problem is drastically overexpressing T7 RNA polymerase and allowing it to overexpress a reporter. This causes an enormous burden on cells, resulting in lower protein yield, mutation, and cell death. This has led to reports that the wild type T7 RNA polymerase:promoter pair is among the weakest pairs tested

(Shis and Bennett, 2013) and to erroneous reports of enzymes with activities many-fold greater than wild type T7 RNA polymerase (Esvelt *et al*., 2011).

In one case, the promoter and RBS driving an evolving T7 RNA polymerase was allowed to evolve, and this evolved promoter and RBS was retained in the subsequent assays (Dickinson *et al*., 2013). This confuses the function per enzyme with the relative abundance of protein. This highlights a more fundamental issue, namely, that the verification of the activity of an evolved enzyme should not mimic the selection conditions. As the adage that "you get what you select for" implies, an evolved enzyme might not have the exact desired functionality, but may be an artifact of the selection. An assay that mimics the selection conditions may miss subtle differences in the evolved and desired function. When possible, *in vivo* evolved enzymes should be assayed *in vitro* and vice versa.

Activity and stability

An estimated 40% of mutations are highly or weakly deleterious, with the majority of their effects being due to a reduction in folding or stability (Camps *et al*., 2007; Soskine and Tawfik, 2010). Mutations that confer a new functionality tend to exhibit even stronger destabilizing effects (Wang *et al*., 2002; Tokuriki *et al*., 2008; Levin *et al*., 2009; Soskine and Tawfik, 2010). Thus, compensatory mutations are needed to allow a potentially adaptive, but destabilizing mutation to emerge. The ability of the M5 stabilizing mutations to unlock the substrate expanding potential of the FA and RGVG mutations underscores this point. Going forward, selections for new function may benefit from the addition of global stabilizing mutations (if any are known) or by starting with an intense neutral drift, discussed below.

ONGOING WORK AND FUTURE DIRECTIONS

A neutral network of T7 RNA polymerases

As natural proteins have evolved over millions of years, there has been a trade-off between robustness to mutation and evolvability. The degree to which a protein retains its character in the face of mutation is likely an evolved phenotype dependent on the mutation rate and selective pressure it has faced over evolutionary time. However, robustness and evolvability do not necessarily have to be antagonistic on a population level, as a neutral network of robust mutants can be quite evolvable (Wagner, 2008; Payne and Wagner, 2014).

About 8% of mutations ablate activity and another 30% reduce activity *(Camps et al*., 2007; Soskine and Tawfik, 2010). Most of the mutations appear neutral, with a small percentage being beneficial. This makes selecting for new functionality problematic, especially if multiple beneficial mutations are required. Mutating a library to the point where there is a reasonable chance of a single variant having multiple beneficial mutations all but assures that each variant will have many deleterious mutations. This problem can be overcome by purging deleterious mutations with iterative rounds of purifying selection (Bershtein *et al*., 2006, 2008). This method, termed "neutral drift" entails repeatedly selecting for wild-type function within a highly mutagenic regime. Each round, new mutations are introduced into the population, and the population is subject to a purifying selection. Since only active variants survive, all highly deleterious mutations are necessarily purged from the population. A selection (like CPR) that enriches for highly active variants also purges mildly deleterious mutations to a lesser extent. After many such cycles, neutral and beneficial mutations can accumulate. The resulting neutral network is a highly diverse library that (1) may contain stabilizing mutations and (2) may contain mutations with latent potential to confer a new function.

As the mutations accumulate in a neutral drift selection, mildly destabilizing mutations begin to pile up. Thus, any mutation that can broadly compensate for the destabilizing effects of other mutations will be selected for. Thus, a persistent neutral drift selection can be expected to favor the accumulation of global stabilizing mutations (Bershtein *et al*., 2006, 2008; Soskine and Tawfik, 2010; Goldsmith and Tawfik, 2009). It should be noted that while stabilizing mutations should accumulate, destabilizing mutations are much more common. It should therefore not be expected that each individual variants is more stable than wild type, although many may be (Bershtein *et al*., 2008).

The neutral drift population should not only contain mutations conferring robustness, but also mutations conferring latent functionality. Essentially, selecting for wild-type function allows for the accumulation of mutations that appear neutral in that context. When the selective pressure changes, a subset of the previously neutral mutations can confer an advantage in the new context (Amitai *et al*., 2007; Gupta and Tawfik, 2008).

This framework was applied to the creation of a neutral network of T7 RNA polymerase variants. Initially, wild type T7 RNA polymerase was amplified by error-prone PCR, adding an expected one to two mutations per gene. The resulting library was subject to selection by CPR, essentially as described in **Chapter 3** (**Figure 5-2**). However, unlike the selection for promoter utilization, the entire open reading frame was allowed to evolve. A PCR product spanning the entire open reading frame could not be efficiently captured on a bead, so a new recovery method was devised. Essentially, the emulsion PCR step used primers that appended a novel sequence tag. When the emulsion was broken, all DNA was recovered (including input plasmid) with a PCR-cleanup column. A second PCR step was performed, in which primers annealed to the recently

appended tags. The product was gel purified and used as the input for a third PCR with nested primers. This third PCR helped reduce the accumulation of shorter amplicons. This PCR product was gel purified, digested, and ligated into an expression vector, thus preparing it for further cycles.

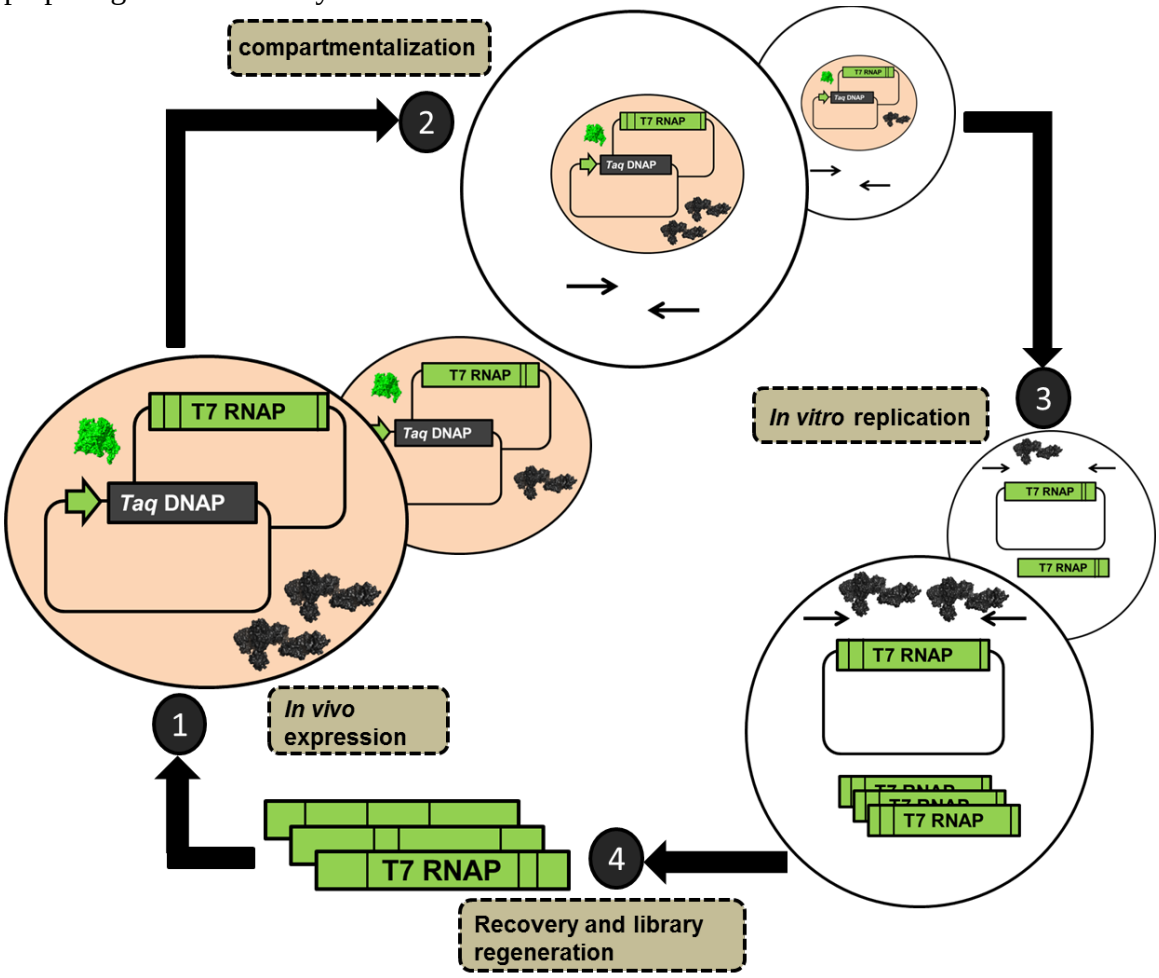


Figure 5-2. Compartmentalized partnered replication selection scheme.

Error-prone PCR was performed before Rounds 4, 7, 9 and 14, and 17 (**Table 5-1**). The error-prone PCR before Round 17 was not the traditional protocol involving *Taq* DNA polymerase and manganese (Fromant *et al.*, 1995). Instead, it was essentially a "normal" PCR using a newly evolved polymerase termed A12 DNA polymerase. A12

DNA polymerase is a shuffled chimera of KOD and Pfu DNA polymerases that was evolved by Dan Garry and screened by Tom Wall of the Ellington Lab. A12 DNA polymerase is extremely fast and synthesizes large quantities of product, but is highly error-prone. A12 was therefore found to be suitable for facile mutagenesis between rounds.

Library preparation	Round	Promoter driving T7 RNAP	[IPTG] added	Target plasmid	Emulsion conditions	DNAP used for regeneration
epPCR	1	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	2	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	3	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	4	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	5	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	6	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	7	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	8	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	9	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	10	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	11	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	12	lacUV5	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	13	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	14	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
	15	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
	16	lacUV5	0.05 mM	pACYC-PK6	Nuevo	A12
	17	lacUV5	0.05 mM	pACYC-PK6	Nuevo	Accuprime Pfx

Table 5-1. Conditions used in neutral drift experiments.

As rounds progressed, the stringency was increased and the protocol was improved. In Rounds 1 to 11, T7 RNA polymerase was driven by the strong T5-lac promoter. Rounds 1 to 8 were strongly induced with 0.5 mM IPTG while Rounds 9 to 11 were moderately induced with 0.1 mM IPTG. In Rounds 12 to 17, T7 RNA polymerase

was driven by the moderate lacUV5 promoter. Round 12 was strongly induced with 0.5 mM IPTG while Rounds 13 to 17 were moderately induced with 0.05 mM IPTG.

After Round 15, CPR was performed under the "CPR Nuevo" conditions. In CPR Nuevo, developed by Michael Ledbetter of the Ellington Lab, the water-in-oil emulsification is performed in a TissueLyser LT at 42 Hz for four minutes. Agitation is aided by the inclusion of the rubber piston from a 1 ml syringe. These conditions have been shown to create more consistent emulsions with only one bacterium per droplet.

Additionally, after Round 15, *Taq* DNA polymerase was replaced with PK6 DNA polymerase. PK6 DNA polymerase is a shuffled chimera of KOD and Pfu DNA polymerases that was evolved and screened by Dan Garry of the Ellington Lab. Similarly to A12 DNA polymerase, PK6 DNA polymerase is extremely fast and generates vast quantities of product. It allowed for greater recovery and less short-amplicon accumulation, enabling a larger library to be transformed each round.

The Round 17 library was transformed into a pCTC-Y2 reporter line (described below). Of 186 clones assayed, the 21 most active variants, judged by T7 RNA polymerase driven fluorescence output, were sequenced. The population was heavily mutated, ranging from 16 to 33 mutations each. On average each variant possessed 23.4 mutations or 8.8 mutations per kilobase. Of these, 38.5% were non-synonymous, resulting in 9.0 coding mutations per gene. Surprisingly only three mutations occurred in more than 10% of genes. R31H, E108K, and D130V occurred three, three, and four times respectively. Of the M5 mutations, S430P and S633P each occurred once.

This neutral drift library is heavily mutated and may serve as a potential starting point for several directed evolution efforts. Processivity and thermal stability are examples of phenotypes for which a diverse neutral network is an ideal starting library. The neutral drift library may also be continued further in hopes that more consensus

mutations emerge. The resulting mutations may be applied to RGVG or another substrate specificity mutant (**Chapter 4**) in an effort to increase its activity.

T7 RNA polymerase processivity

T7 RNA polymerase undergoes a large conformational change as it moves from initiation to elongation. In the initiation phase, it makes contact with the promoter upstream of the active site and feeds the template strand into the active site. In the elongation phase, it no longer make contact with upstream DNA, but instead makes contact with downstream DNA and funnels the downstream template strand into the active site. The fingers and N-terminal domain make long-range electrostatic interaction with this DNA and likely contribute to the processivity of the polymerase (Tahirov *et al*., 2002). There is little known about the specific residues involved in the processivity of T7 RNA polymerase.

Processivity is likely a distributed phenotype so a targeted mutagenesis library would not be effective for efforts to enhance to processivity of T7 RNA polymerase. Similarly, several mutations are likely required to see a noticeable increase in processivity, so a random mutagenesis library would likely not be ideal. Thus, the neutral drift library described above was chosen as a starting point for the evolution of a hyper-processive T7 RNA polymerase mutant.

Directed evolution by CPR as above was continued, starting with the Round 16 population (**Table 5-2**). However, in order to add selective pressure for processivity a long 5′ UTR was included in the mRNA transcribed by T7 RNA polymerase. Thus, in order to make PK6 DNA polymerase (which catalyzes the emulsion PCR) T7 RNA polymerase would need to make a roughly 6-kb transcript, as opposed to the 2.5-kb mRNA previously required.

Library preparation	Round	Promoter driving T7 RNAP	[IPTG] added	Target plasmid	Emulsion conditions	DNAP used for regeneration
epPCR	1	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	2	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	3	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	4	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	5	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	6	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	7	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	8	T5-lac	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	9	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	10	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	11	T5-lac	0.1 mM	pACYC-Taq	Classic	Accuprime Pfx
	12	lacUV5	0.5 mM	pACYC-Taq	Classic	Accuprime Pfx
	13	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
epPCR	14	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
	15	lacUV5	0.05 mM	pACYC-Taq	Classic	Accuprime Pfx
	16	lacUV5	0.05 mM	pACYC-PK6	Nuevo	A12
	17	lacUV5	0.05 mM	pC1RCT-PK6	Nuevo	A12
	18	lacUV5	0.05 mM	pC1RCT-PK6	Nuevo	A12
	19	lacUV5	0.05 mM	pC1RCT-PK6	Nuevo	A12
Shuffle	20	lacUV5	0.05 mM	pC1RCT-PK6	Nuevo	A12
	21	lacUV5	0.05 mM	pC1RCT-PK6	Nuevo	A12
	22	lacUV5	0.05 mM	pCTC-PK6	Nuevo	Accuprime Pfx
	23	lacUV5	0.05 mM	pCTC-PK6	Nuevo	Accuprime Pfx
Shuffle	24	lacUV5	0.05 mM	pCTC-PK6	Nuevo	Accuprime Pfx
	25	lacUV5	0.05 mM	pCTC-PK6	Nuevo	Accuprime Pfx

Table 5-2. Conditions used in the selection for processivity.

In order to determine the effectiveness of such a spacer, a construct (pC1Y2) was made in which the T7 promoter was followed by eCFP immediately followed by seYFP (**Figure 5-3**). The two fluorescent proteins used different sets of codons to minimize the possibility of recombination. eCFP used "version 1" codons and is shorthanded as C1, seYFP used "version 2" codons and is shorthanded as Y2. A similar construct (pC1-

RCT-Y2) was made in which C1 and Y2 were separated by a roughly 2.5-kb spacer comprising the reverse complement of a null-mutant of *Taq* DNA polymerase (RCT).

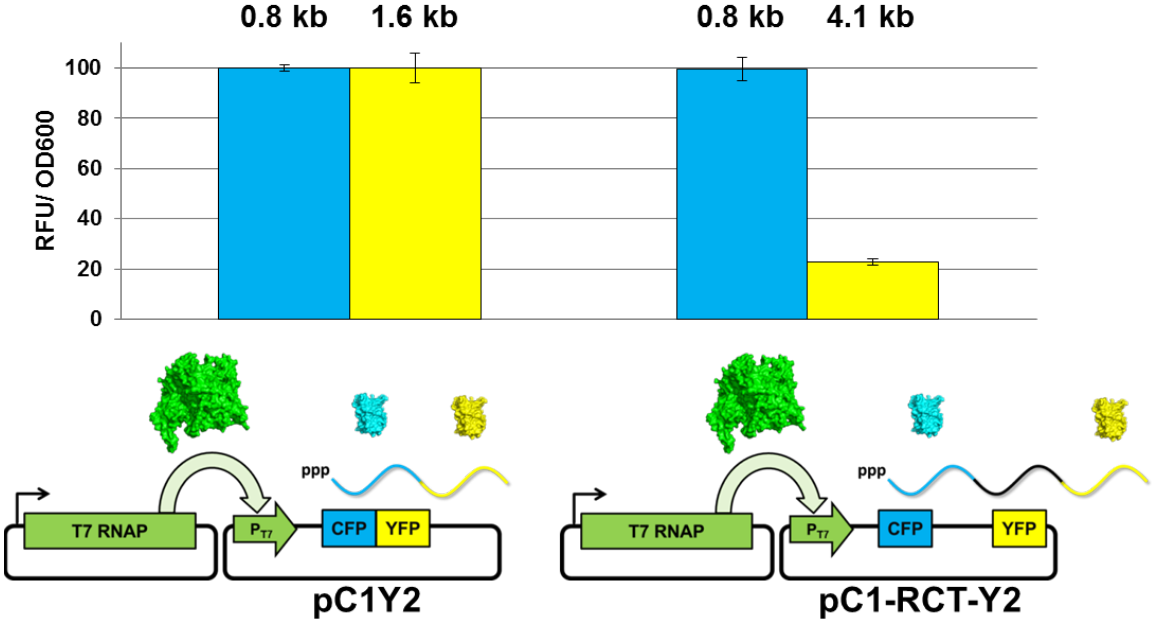


Figure 5-3. 5′ UTR length and protein yield.
Wild type T7 RNA polymerase was transformed into two reporter lines. In the first (pC1Y2), a T7 promoter is followed by eCFP and seYFP. In the second, eCFP and seYFP are separated by a 2.5-kb spacer. The eCFP and seYFP fluorescence (in Relative Fluorescence Units) / OD600 from the first construct were defined as 100 and compared to reading from the second construct. Values are the average from three independent cultures; error bars represent standard error.

Wild type T7 RNA polymerase and an inactive T7 RNA polymerase were transformed into cell lines containing pC1Y2 and pC1-RCT-Y2. The inactive control indicated that there were no cryptic promoters in the spacer sequence and that seYFP was only being driven from the T7 promoter. The eCFP and seYFP outputs from wild type T7 RNA polymerase with pC1Y2 were each defined as 100. These outputs were compared to the eCFP and seYFP outputs of wild type T7 RNA polymerase with pC1-RCT-Y2. As expected the eCFP output was similar for each construct, and the addition

of the 2.5-kb spacer reduced seYFP yield about five-fold. Thus the spacer was effective and could serve as the basis for selection.

A construct (pC1-RCT-PK6) was designed in which the T7 promoter was followed by a roughly 3.5-kb spacer, consisting of C1 and RCT, followed by PK6 DNA polymerase. The Round 16 neutral drift library was transformed into a cell line containing this construct and subjected to two rounds of selection by CPR, with A12 DNA polymerase-based error-prone PCR between rounds (**Table 5-2**).

The Round 18 population was transformed into the pC1-RCT-Y2 reporter line and 93 clones were assayed for eCFP and seYFP. 11 clones showed activity comparable to wild type T7 RNA polymerase and were sequenced. These clones ranged from 20 to 32 mutations, averaging 26.5 mutations per gene (10.0 mutations per kilobase). 43.0% of mutations were non-synonymous, resulting in an average of 11.4 coding mutations per gene. Of the 11 mutants, seven contained the mutation D130V, while the H161Y mutation was observed in three mutants.

The preponderance of mutations and the emergence possible consensus mutations suggested that gene shuffling was in order. The Round 19 population was shuffled amongst itself (Stemmer, 1994; Meyer *et al*., 2014), and two additional rounds of selection were performed.

The Round 21 population was assayed as before, except a new reporter construct (pCTC-Y2) was used. pCTC-Y2 contains the T7 promoter followed by C1, RCT, the reverse complement of the Cas9 gene, and Y2. In this case, seYFP was only made if a greater than 8-kb transcript was made. 93 clones from Round 21 were assayed, seven showed levels comparable to wild type and were sequenced. These ranged from 26 to 39 mutations each and averaged 33.4 mutations each (12.6 mutations per kilobase). 40.5% of the mutations were non-synonymous, resulting in 13.6 non-synonymous mutations per

gene. Three mutants contained the D130V mutation. One mutant from Round 21 was able to produce about 60% more seYFP than wild type T7 RNA polymerase (requiring an 8-kb transcript) despite having only been selected for the ability to generate 6-kb transcripts.

To increase the selective pressure for processivity, a new construct (pCTC-PK6) was made in which the seYFP from the pCTC-Y2 reporter was replaced with PK6 DNA polymerase. Four additional rounds were performed with this new construct, which required more than 10-kb of transcription to yield PK6 DNA polymerase. The PCR steps between rounds were performed with Accuprime Pfx DNA polymerase instead of the error-prone A12 DNA polymerase and gene shuffling was performed prior to Round 24. These actions were performed in order to allow deleterious mutations to be purged and beneficial mutations to be combined. 186 clones from the Round 25 population were assayed using the pCTC-Y2 line as described previously. The eCFP and seYFP output (normalized to OD600) for the 39 most active clones are shown in (**Figure 5-4**). Several mutant are able to make considerably more seYFP (requiring an 8-kb transcript) than wild type T7 RNA polymerase.

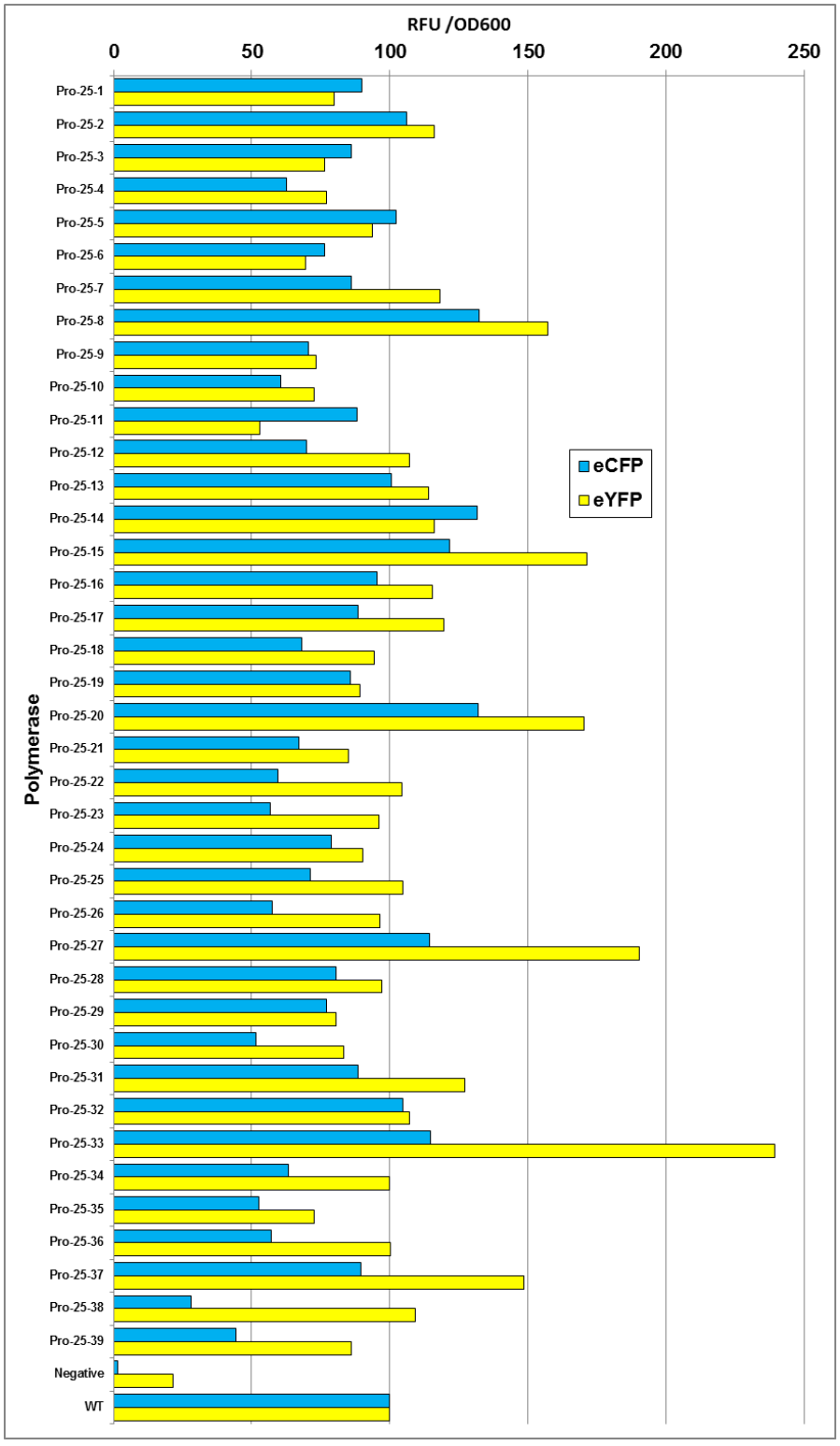

RFU /OD600
0
50
100
150
200
250
Polymerase
Pro-25-1
Pro-25-2
Pro-25-3
Pro-25-4
Pro-25-5
Pro-25-6
Pro-25-7
Pro-25-8
Pro-25-9
Pro-25-10
Pro-25-11
Pro-25-12
Pro-25-13
Pro-25-14
Pro-25-15
Pro-25-16
Pro-25-17
Pro-25-18
Pro-25-19
Pro-25-20
Pro-25-21
Pro-25-22
Pro-25-23
Pro-25-24
Pro-25-25
Pro-25-26
Pro-25-27
Pro-25-28
Pro-25-29
Pro-25-30
Pro-25-31
Pro-25-32
Pro-25-33
Pro-25-34
Pro-25-35
Pro-25-36
Pro-25-37
Pro-25-38
Pro-25-39
Negative
WT
eCFP
eYFP

Figure 5-4

Figure 5-4. *In vivo* activity assay of Round 25 of the T7 RNA polymerase selection for processivity.

The library from Round 25 of the selection for processivity was transformed into the pCTC-Y2 reporter plasmid. The eCFP and seYFP fluorescence (in Relative Fluorescence Units) / OD600 is a measure of 1-kb and 8-kb mRNA output, respectively. The fluorescence outputs from eCFP and seYFP driven by wild type T7 RNA polymerase are each defined as 100. The 39 most active clones out of 186 tested are shown.

37 clones were sequenced and they contain 24 to 49 mutations. They averaged 36.6 mutations per gene, 13.8 per kilobase. 40.1% of mutations were non-synonymous, resulting in 14.7 coding mutations per gene. R52H, E63V, V134I, and Q169R each occurred five times, H772R occurred seven times, and D130V was present in 26 of 37 mutants. It is not immediately apparent whether these mutations directly increase processivity, or merely improve stability. The most common mutation, D130V, was seen frequently in Round 17, before the selection for processivity began. The next most common mutation, H772R, has been shown to confer stability to the G542V H784S mutant (**Chapter 4**). It is possible that the selection for processivity is merely a very stringent selection for activity. More selection, more precise analysis of the selected mutants, and isolation of the effects of each consensus mutation is required.

Thermal stability

Pooling the sequence data from Rounds 17, 18, 21, and 25 described above, it is easier to see the emergence of several consensus mutations. Three mutations occur in at least 10% of the 77 mutants from these rounds. Specifically, D130V occurs in 40 mutants, H161Y occurs in 11 mutants, and H772R occurs in eight mutants. These mutations all arose before selection for processivity, so they may be conferring robustness to mutation. They can be added to the substrate specificity mutants described in **Chapter 4** to enhance their activity, or serve as the basis for a selection for thermal stability.

There may already be enough information gathered to design a thermostable T7 RNA polymerase mutant. The M5 mutations (S430P, N433T, S633P, F849I, F880Y) and the mutations described in this chapter may be added in some combination to generate a more thermostable mutant.

Alternatively, a direct selection for thermal stability may be required. The CPR selection as described is not compatible with a direct selection for thermal stability. This is because T7 RNA polymerase performs its function in *E. coli*, whose temperature range is limited. This problem may be overcome by either adapting the selection to a thermostable organism or by altering the selection method.

In order to select for a function not compatible with *E. coli* function (*i.e.* thermal stability or unnatural nucleotide incorporation), T7 RNA polymerase must perform its function outside of the cell. In such a selection, the host cell would function only to produce T7 RNA polymerase. The cell would then be emulsified and lysed, allowing T7 RNA polymerase to access the emulsion compartment and perform a function that would allow its genotype to succeed. This could take the form of an autogene like selection in which T7 RNA polymerase transcribes its own genetic information. Alternatively, T7 RNA polymerase could transcribe an oligonucleotide that can be used to capture the plasmid encoding it. This would be analogous to, but quite distinct from, compartmentalized self-tagging (CST) (Pinheiro *et al*., 2012). Other schemes can be envisioned in which T7 RNA polymerase makes or activates primers used by a DNA polymerase to amplify the T7 RNA polymerase gene.

CONCLUDING REMARKS

T7 RNA polymerase remains a fascinating and useful enzyme. It has been engineered and evolved by several labs over the past few decades. In this work, these previous engineering efforts have been built upon, and some of the most active and potentially useful mutants to date have been generated. The methods, challenges, and successes presented herein can hopefully provide valuable instruction to future protein engineers.

MATERIALS AND METHODS

Compartmentalized partnered replication

Libraries (~50 ng purified ligation) were transformed into BL21 gold cells (Agilent) containing the target selection plasmid. Library DNA was introduced by electroporation using 0.2 cm cuvettes at 2.5 kV in an *E. coli* pulser (BioRad). This routinely resulted in roughly 10^7 CFUs (multiple replicates were often pooled).

100 µl overnight cultures from the transformation were subcultured in 2 ml 2xYT medium, grown for one hour (OD600 ~0.6) and induced with IPTG at 37 °C for four hours. 200 µl of the induced culture was centrifuged at 5,000 g for 10 minutes to pellet the cells. The supernatant was removed and cells were gently resuspended in 20 µl 10x PCR buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$) 10 µl dNTP mix (4 mM each), 4 µl each primer (20 µM), and 162 µl water. Emulsification was performed by slowly adding resuspended cells to 600 µl of spinning oil mix (438 µl Tegosoft DEC (Evonik), 42 µl AbilWE09 (Evonik), and 120 µl Mineral oil (Sigma)).

In "CPR classic" the oil mixture was constantly spun in a tube (Sarstedt 13 ml 95 mm x 16.8 mm) on ice using a stirbar (Spinplus 9.5 mm x 9.5 mm Teflon, Bel-Art) on a magnetic plate (Corning) at the maximum setting (1150 rpm). The cell mixture was slowly added over a one minute interval and spun for an additional four minutes.

In "CPR nuevo" the oil and cell mixtures were combined in 2 ml centrifuge tube, along with the rubber piston from a 1 ml syringe. The tube was then agitated at 42 Hz for four minutes in a TissueLyser LT (Qiagen).

The emulsified cells were thermal cycled (95 °C:3 min, 20 cycles [95 °C:30 sec, 55 °C:30 sec, 72 °C:5 min], 72 °C:5 min) such that cells containing the most active enzymes will also contain the most DNA polymerase and will preferentially amplify.

The emulsion was broken in two steps. Firstly, it was centrifuged at 10,000 g for five minutes and the oil (upper) phase was removed. Secondly, 300 µl of bead buffer (0.2 M NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.1% Tween-20) and 500 µl chloroform was added and the mixture was vortexed vigorously. The mixture was transferred to a heavy-gel phase-lock tube (5 Prime) and upon centrifugation at 16,000 g for two minutes the aqueous (upper) phase was collected along with any nucleic acids present. This was purified using SV Wizard PCR clean-up (Promega).

To facilitate purification of the DNA amplified by in the emulsion PCR step, primers that append sequence tags to the ends of each amplicon were used. Purified DNA was used as the template for re-amplification using primers that anneal to the appended tags. This PCR product was gel purified and used as the template in another PCR step using nested primers. This was purified and followed by digestion and ligation into the appropriate plasmid backbone.

Error-prone PCR

The error-prone PCR reaction mixture was composed of 50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM $MgCl_2$, 5 µg/ml BSA, 0.35 mM dATP, 0.4 mM dCTP, 0.2 mM dGTP, 1.35 mM dTTP, 0.5 mM $MnCl_2$, 0.5 µM each primer, 2 ng/µl template, and 0.8 U/µl *Taq* DNA polymerase (New England Biolabs) and was thermal cycled (95 °C:4 min, 25 cycles [95 °C:30 sec, 55 °C:30 sec, 72 °C:5 min], 72 °C:5 min). This achieved the expected one mutation per 500 base pairs.

Gene shuffling

To shuffle CPR libraries, DNA was digested, thermal cycled, and amplified. First, 4 µg DNA was mixed with 5 µl 1M Tris-HCl pH 7.4, 5 µl 200 mM $MnCl_2$, and water up to 100 µl. This mixture was incubated at 15 °C for five minutes. One unit of

DNAse (New England Biolabs) was added and the mixture was incubated at 15 °C for three minutes before heat inactivating at 80 °C for five minutes. Digested DNA was purified using SV Wizard PCR clean-up (Promega).

20% of the recovered fragments were added to Platinum *Taq* DNA polymerase High Fidelity (Life Technologies) in its normal reaction buffer with no primers (100 µl total). This was thermal cycled as follows: 94˚C:2 min; 35 cycles of (94˚C:30 sec, 65˚C:90 sec, 62˚C:90 sec, 59˚C:90 sec, 56˚C:90 sec, 53˚C:90 sec, 50˚C:90 sec, 47˚C:90 sec, 44˚C:90 sec, 41˚C:90 sec, 68˚C:4 min); 68˚C:4 min. The reassembled fragments were purified using SV Wizard PCR clean-up (Promega). This was used as a template for PCR and prepared normally for the next round of CPR.

In vivo activity assay

T7 RNA polymerase plasmids or libraries were electroporated into BL21-gold cells containing pC1Y2, pC1RCT-Y2, or pCTC-Y2 in which the T7 promoter variants controlled expression of eCFP and seYFP. Single colonies were grown at 37 °C overnight. 100 µl of the culture was grown in 2 ml 2xYT medium at 37 °C for one hour (OD600 ~0.6) and induced with 0.05 mM IPTG for four hours. This concentration of IPTG was chosen in order to limit metabolic overload on the host and prevent saturation of signal. After induction, cells were centrifuged and resuspended in 600 µl PBS. 200 µl of resuspended cells were measured for OD600 on a Synergy-HT plate reader (Bio-Tek) and eCFP fluorescence (Excitation/ Emission 434 nm /474 nm) and seYFP fluorescence (Excitation/ Emission 485 nm/535 nm) on a Safire monochromator (Tecan).

REFERENCES

Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., and Tawfik, D. S. 2005. The “evolvability” of promiscuous protein functions. *Nature Genetics* 37:73–6.

Amitai, G., Gupta, R. D., and Tawfik, D. S. 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal* 1:67.

Bershtein, S., Goldin, K., and Tawfik, D. S. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of Molecular Biology* 379:1029–44.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–32.

Camps, M., Herman, A., Loh, E., and Loeb, L. A. 2007. Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology* 42:313–26.

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*.

Esvelt, K. M., Carlson, J. C., and Liu, D. R. 2011. A system for the continuous directed evolution of biomolecules. *Nature* 472:499–503.

Fromant, M., Blanquet, S., and Plateau, P. 1995. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Analytical Biochemistry* 224:347–353.

Goldsmith, M., and Tawfik, D. S. 2009. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proceedings of the National Academy of Sciences of the United States of America* 106:6197–202.

Gupta, R. D., and Tawfik, D. S. 2008. Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods* 5:939–942.

Khersonsky, O., and Tawfik, D. S. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* 79:471–505.

Levin, K. B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C., and Tawfik, D. S. 2009. Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nature Structural & Molecular Biology* 16:1049–55.

Matsumura, I., and Ellington, A. D. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *Journal of Molecular Biology* 305:331–9.

McLoughlin, S. Y., and Copley, S. D. 2008. A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proceedings of the National Academy of Sciences of the United States of America* 105:13497–502.

Meyer, A. J., Ellefson, J. W., and Ellington, A. D. 2014. Library generation by gene shuffling. *Current Protocols in Molecular Biology* 105:Unit 15.12.

Payne, J. L., and Wagner, A. 2014. The robustness and evolvability of transcription factor binding sites. *Science* 343:875–7.

Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., et al. 2012. Synthetic genetic polymers capable of heredity and evolution. *Science* 336:341–4.

Shis, D. L., and Bennett, M. R. 2013. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proceedings of the National Academy of Sciences of the United States of America*.

Soskine, M., and Tawfik, D. S. 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews. Genetics* 11:572–82.

Stemmer, W. P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370:389–91.

Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., and Yokoyama, S. 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* 420:43–50.

Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. 2012. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* 40:8773–81.

Tokuriki, N., Jackson, C. J., Afriat-Jurnou, L., Wyganowski, K. T., Tang, R., and Tawfik, D. S. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* 3:1257.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. 2008. How protein stability and new functions trade off. *PLoS Computational Biology* 4:e1000002.

Vick, J. E., and Gerlt, J. A. 2007. Evolutionary potential of (beta/alpha)8-barrels: stepwise evolution of a “new” reaction in the enolase superfamily. *Biochemistry* 46:14589–97.

Wagner, A. 2008. Robustness and evolvability: a paradox resolved. *Proceedings. Biological sciences / The Royal Society* 275:91–100.

Wang, X., Minasov, G., and Shoichet, B. K. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology* 320:85–95.

Weinreich, D. M., Delaney, N. F., Depristo, M. A., and Hartl, D. L. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–4.

References

Agresti, J. J., Kelly, B. T., Jäschke, A., and Griffiths, A. D. 2005. Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America* 102:16170–5.

Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., and Tawfik, D. S. 2005. The "evolvability" of promiscuous protein functions. *Nature Genetics* 37:73–6.

Amitai, G., Gupta, R. D., and Tawfik, D. S. 2007. Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal* 1:67.

Bartel, D., and Szostak, J. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418.

Beaudry, A., DeFoe, J., Zinnen, S., Burgin, A., and Beigelman, L. 2000. In vitro selection of a novel nuclease-resistant RNA phosphodiesterase. *Chemistry & Biology* 7:323–34.

Bershtein, S., Goldin, K., and Tawfik, D. S. 2008. Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of Molecular Biology* 379:1029–44.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–32.

Brakmann, S., and Grzeszik, S. 2001. An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochem* 2:212–9.

Brieba, L. G., and Sousa, R. 2000. Roles of histidine 784 and tyrosine 639 in ribose discrimination by T7 RNA polymerase. *Biochemistry* 39:919–23.

Bryksin, A. V, and Matsumura, I. 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* 48:463–5.

Bull, J. J., and Pease, C. M. 1995. Why Is the Polymerase Chain Reaction Resistant to In Vitro Evolution? *Journal of Molecular Evolution*:1160–1164.

Bull, J. J., Sanjuán, R., and Wilke, C. O. 2007a. Theory of Lethal Mutagenesis for Viruses. *Journal of Virology* 81:2930–2939.

Bull, J. J., Springman, R., and Molineux, I. J. 2007b. Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. *Molecular Biology and Evolution* 24:900–8.

Burmeister, P. E., Lewis, S. D., Silva, R. F., Preiss, J. R., Horwitz, L. R., Pendergrast, P. S., McCauley, T. G., Kurz, J. C., Epstein, D. M., Wilson, C., et al. 2005. Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chemistry & Biology* 12:25–33.

Camps, M., Herman, A., Loh, E., and Loeb, L. A. 2007. Genetic constraints on protein evolution. *Critical Reviews in Biochemistry and Molecular Biology* 42:313–26.

Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. 2014. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* 10:216–22.

Carroll, S. S., Cowart, M., and Benkovic, S. J. 1991. A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* 30:804–813.

Chamberlin, M., McGrath, J., and Waskell, L. 1970. New RNA polymerase from Escherichia coli infected with bacteriophage T7. *Nature* 228:227–231.

Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. 1999. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399:80–3.

Cheetham, G. M., and Steitz, T. A. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286:2305–9.

Chelliserrykattil, J., Cai, G., and Ellington, A. D. 2001. A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnology* 1:13.

Chelliserrykattil, J., and Ellington, A. D. 2004. Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nature Biotechnology* 22:1155–60.

Compton, J. 1991. Nucleic acid sequence-based amplification. *Nature* 350:91–92.

Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. 1996. A T7 promoter-specific, inducible protein expression system for Bacillus subtilis. *Molecular & General Genetics* 250:230–236.

Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. 2012. An in vitro Autogene. *ACS Synthetic Biology* 1:190–196.

Dean, N. M., and Bennett, C. F. 2003. Antisense oligonucleotide-based therapeutics for cancer. *Oncogene* 22:9087–96.

Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M., and Liu, D. R. 2013. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America*.

Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N., and Yanagawa, H. 2004. In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Research* 32:e95.

Dubendorff, J. W., and Studier, F. W. 1991. Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. *Journal of Molecular Biology* 219:61–8.

Dunn, J. J., and Studier, F. W. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology* 166:477–535.

Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523.

Eigen, M., and Schuster, P. 1977. The Hypercycle. *Naturwissenschaften* 64:541–565.

Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology* 32:97–101.

Ellington, A. D., and Szostak, J. W. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.

Elroy-Stein, O., and Moss, B. 1990. Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 87:6743–7.

Esvelt, K. M., Carlson, J. C., and Liu, D. R. 2011. A system for the continuous directed evolution of biomolecules. *Nature* 472:499–503.

Finn, J., Lee, A. C. H., MacLachlan, I., and Cullis, P. 2004. An enhanced autogene-based dual-promoter cytoplasmic expression system yields increased gene expression. *Gene Therapy* 11:276–83.

Franco, E., Friedrichs, E., Kim, J., Jungmann, R., Murray, R., Winfree, E., and Simmel, F. C. 2011. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences of the United States of America* 108:E784–93.

Fromant, M., Blanquet, S., and Plateau, P. 1995. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Analytical Biochemistry* 224:347–353.

Fuerst, T. R., Niles, E. G., Studier, F. W., and Moss, B. 1986. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 83:8122–6.

Gamer, M., Fröde, D., Biedendieck, R., Stammen, S., and Jahn, D. 2009. A T7 RNA polymerase-dependent gene expression system for Bacillus megaterium. *Applied Microbiology and Biotechnology* 82:1195–203.

Gao, X. 2003. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research* 31:143e–143.

Ghadessy, F. J., and Holliger, P. 2004. A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Engineering, Design & Selection* 17:201–4.

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

Gibson, D. G. 2011. Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology* 498:349–61.

Goldsmith, M., and Tawfik, D. S. 2009. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proceedings of the National Academy of Sciences of the United States of America* 106:6197–202.

Griffiths, A. D., and Tawfik, D. S. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *The EMBO Journal* 22:24–35.

Guatelli, J. C., Whitfield, K. M., Kwoh, D. Y., Barringer, K. J., Richman, D. D., and Gingeras, T. R. 1990. Isothermal, in vitro amplification of nucleic acids by a

multienzyme reaction modeled after retroviral replication. *Proceedings of the National Academy of Sciences of the United States of America* 87:7797.

Guillerez, J., Lopez, P. J., Proux, F., Launay, H., and Dreyfus, M. 2005. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America* 102:5958–63.

Gupta, R. D., and Tawfik, D. S. 2008. Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods* 5:939–942.

Haruna, I., and Spiegelman, S. 1965. Autocatalytic Synthesis of a Viral RNA in vitro. *Science* 150:884–886.

Healy, J. M., Lewis, S. D., Kurz, M., Boomer, R. M., Thompson, K. M., Wilson, C., and McCauley, T. G. 2004. Pharmacokinetics and biodistribution of novel aptamer compositions. *Pharmaceutical Research* 21:2234–46.

Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. 1997. Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* 36:8231–42.

Ibach, J., Dietrich, L., Koopmans, K. R. M., Nöbel, N., Skoupi, M., and Brakmann, S. 2013. Identification of a T7 RNA polymerase variant that permits the enzymatic synthesis of fully 2'-O-methyl-modified RNA. *Journal of Biotechnology* 167:287–95.

Ikeda, R. A., Ligman, C. M., and Warshamana, S. 1992. T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Research* 20:2517–24.

Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. 2000. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* 39:10419–30.

Jackson, A. L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J. M., Lim, L., Karpilow, J., Nichols, K., et al. 2006. Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* 12:1197–205.

Jeruzalmi, D., and Steitz, T. A. 1998. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal* 17:4101–13.

Kang, S. T., Leu, J. H., Wang, H. C., Chen, L. L., Kou, G. H., and Lo, C. F. 2009. Polycistronic mRNAs and internal ribosome entry site elements (IRES) are widely used by white spot syndrome virus (WSSV) structural protein genes. *Virology* 387:353–63.

Kast, P., and Hennecke, H. 1991. Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *Journal of Molecular Biology* 222:99–124.

Keefe, A. D., and Cload, S. T. 2008. SELEX with modified nucleotides. *Current Opinion in Chemical Biology* 12:448–56.

Kennedy, W. P., Momand, J. R., and Yin, Y. W. 2007. Mechanism for de novo RNA synthesis and initiating nucleotide specificity by T7 RNA polymerase. *Journal of Molecular Biology* 370:256–68.

Khersonsky, O., and Tawfik, D. S. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* 79:471–505.

Von Kiedrowski, G. 1986. A Self-Replicating Hexadeoxynucleotide. *Angewandte Chemie International Edition in English* 25:932–935.

Kim, J., White, K. S., and Winfree, E. 2006. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Molecular Systems Biology* 2:68.

Kim, J., and Winfree, E. 2011. Synthetic in vitro transcriptional oscillators. *Molecular Systems Biology* 7:465.

Knudsen, S. M., Robertson, M. P., and Ellington, A. D. 2002. In vitro selection using modified or unnatural nucleotides. *Current Protocols in Nucleic Acid Chemistry* Chapter 9:Unit 9.6.

Kochetkov, S. N., Rusakova, E. E., and Tunitskaya, V. L. 1998. Recent studies of T7 RNA polymerase mechanism. *FEBS Letters* 440:264–7.

Kostyuk, D. A., Dragan, S. M., Lyakhov, D. L., Rechinsky, V. O., Tunitskaya, V. L., Chernov, B. K., and Kochetkov, S. N. 1995. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS Letters* 369:165–8.

Kraynack, B., and Baker, B. 2006. Small interfering RNAs containing full 2′-O-methylribonucleotide-modified sense strands display Argonaute2/eIF2C2-dependent activity. *RNA*:163–176.

Krüger, D. H., and Schroeder, C. 1981. Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. *Microbiological Reviews* 45:9–51.

Kwoh, D. Y., Davis, G. R., Whitfield, K. M., Chappelle, H. L., DiMichele, L. J., and Gingeras, T. R. 1989. Transcription-based amplification system and detection of

amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proceedings of the National Academy of Sciences of the United States of America* 86:1173–7.

Lau, M. W. L., and Ferré-D'Amaré, A. R. 2013. An in vitro evolved glmS ribozyme has the wild-type fold but loses coenzyme dependence. *Nature Chemical Biology* 9:805–10.

Layzer, J. M. 2004. In vivo activity of nuclease-resistant siRNAs. *RNA* 10:766–771.

Levin, K. B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C., and Tawfik, D. S. 2009. Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nature Structural & Molecular Biology* 16:1049–55.

Levy, M., and Ellington, A. D. 2008. Directed evolution of streptavidin variants using in vitro compartmentalization. *Chemistry & Biology* 15:979–89.

Levy, M., Griswold, K. E., and Ellington, A. D. 2005. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* 11:1555–62.

Liao, H., van Gemen, B., and Sugiyama, A. 2003. Mutant of RNA Polymerases with Increased Stability. 1:U. S. Patent. US 6,524,828 B1.

Lincoln, T. A., and Joyce, G. F. 2009. Self-sustained replication of an RNA enzyme. *Science* 323:1229–1232.

Lu, W. C., Levy, M., Kincaid, R., and Ellington, A. D. 2014. Directed evolution of the substrate specificity of biotin ligase. *Biotechnology and Bioengineering* 111:1071–81.

Lupold, S., Hicke, B., Lin, Y., and Coffey, D. 2002. Identification and characterization of nuclease-stabilized RNA molecules that bind human prostate cancer cells via the prostate-specific membrane antigen. *Cancer Research*:4029–4033.

Majlessi, M., Nelson, N. C., and Becker, M. M. 1998. Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Research* 26:2224–9.

Makarova, O. 1995. Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proceedings of the National Academy of Sciences of the United States of America* 92:12250–12254.

Mastrobattista, E., Taly, V., Chanudet, E., Treacy, P., Kelly, B. T., and Griffiths, A. D. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chemistry & Biology* 12:1291–300.

Matsumura, I., and Ellington, A. D. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *Journal of Molecular Biology* 305:331–9.

McAllister, W. T., Morris, C., Rosenberg, A. H., and Studier, F. W. 1981. Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *Journal of Molecular Biology* 153:527–44.

McBride, K. E., Schaaf, D. J., Daley, M., and Stalker, D. M. 1994. Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 91:7301–5.

McLoughlin, S. Y., and Copley, S. D. 2008. A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proceedings of the National Academy of Sciences of the United States of America* 105:13497–502.

Melton, D., and Krieg, P. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research* 18:38–44.

Meyer, A. J., Ellefson, J. W., and Ellington, A. D. 2012. Abiotic Self-Replication. *Accounts of Chemical Research* 45:2097–2105.

Meyer, A. J., Ellefson, J. W., and Ellington, A. D. 2014. Library generation by gene shuffling. *Current Protocols in Molecular Biology* 105:Unit 15.12.

Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research* 15:8783–98.

Mokrejs, M., Masek, T., Vopálensky, V., Hlubucek, P., Delbos, P., and Pospísek, M. 2010. IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Research* 38:D131–6.

Mureev, S., Kovtun, O., Nguyen, U. T. T., and Alexandrov, K. 2009. Species-independent translational leaders facilitate cell-free expression. *Nature Biotechnology* 27:747–52.

Padilla, R., and Sousa, R. 2002. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Research* 30:e138.

Padilla, R., and Sousa, R. 1999. Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutant T7 RNA polymerase (RNAP). *Nucleic Acids Research* 27:1561–1563.

Paige, J. S., Wu, K. Y., and Jaffrey, S. R. 2011. RNA Mimics of Green Fluorescent Protein. *Science* 333:642–646.

Payne, J. L., and Wagner, A. 2014. The robustness and evolvability of transcription factor binding sites. *Science* 343:875–7.

Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., et al. 2012. Synthetic genetic polymers capable of heredity and evolution. *Science* 336:341–4.

Polesky, A. H., Steitz, T. A., Grindley, N. D., and Joyce, C. M. 1990. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from Escherichia coli. *The Journal of Biological Chemistry* 265:14579–91.

Raskin, C. A., Diaz, G. A., and McAllister, W. T. 1993. T7 RNA polymerase mutants with altered promoter specificities. *Proceedings of the National Academy of Sciences of the United States of America* 90:3147–51.

Raskin, C. A., Diaz, G., Joho, K., and McAllister, W. T. 1992. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *Journal of Molecular Biology* 228:506–15.

Romero, P. a, and Arnold, F. H. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews. Molecular Cell Biology* 10:866–76.

Rong, M., He, B., McAllister, W. T., and Durbin, R. K. 1998. Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 95:515–9.

Schenborn, E. T., and Mierendorf, R. C. 1985. A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research* 13:6223–36.

Shis, D. L., and Bennett, M. R. 2013. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proceedings of the National Academy of Sciences of the United States of America*.

Siegmund, V., Santner, T., Micura, R., and Marx, A. 2012. Screening mutant libraries of T7 RNA polymerase for candidates with increased acceptance of 2'-modified nucleotides. *Chemical Communications* 48:9870–2.

Soskine, M., and Tawfik, D. S. 2010. Mutational effects and the evolution of new protein functions. *Nature Reviews. Genetics* 11:572–82.

Sousa, R. 1996. Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences* 21:186–90.

Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B. C. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature* 364:593–9.

Sousa, R., and Padilla, R. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *The EMBO Journal* 14:4609–21.

Steitz, T. A., Smerdon, S. J., Jäger, J., and Joyce, C. M. 1994. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* 266:2022–5.

Stemmer, W. P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370:389–91.

Stoflet, E., Koeberl, D., Sarkar, G., and Sommer, S. 1988. Genomic amplification with transcript sequencing. *Science* 239:491–494.

Studier, F. W. 1972. Bacteriophage T7. *Science* 176:367–76.

Studier, F. W., and Moffatt, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189:113–30.

Studier, F. W., and Rosenberg, A. H. 1981. Genetic and physical mapping of the late region of bacteriophage T7 DNA by use of cloned fragments of T7 DNA. *Journal of Molecular Biology* 153:503–25.

Subsoontorn, P., Kim, J., and Winfree, E. 2012. Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch. *ACS Synthetic Biology* 1:299–316.

Sugiyama, A., Nishiya, Y., and Kawakami, B. 2009. RNA Polymerase Mutants with Increased Thermostability. U. S. Patent. US 7,507,567 B2.

Tabor, S., and Richardson, C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America* 82:1074–1078.

Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., and Yokoyama, S. 2002. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* 420:43–50.

Tawfik, D. S., and Griffiths, A. D. 1998. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* 16:652–656.

Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassylyev, D. G. 2004. Structural basis for substrate selection by T7 RNA polymerase. *Cell* 116:381–91.

Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. 2012a. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research* 40:8773–81.

Temme, K., Zhao, D., and Voigt, C. A. 2012b. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences of the United States of America* 109:7085–90.

Thyer, R., Filipovska, A., and Rackham, O. 2013. Engineered rRNA Enhances the Efficiency of Selenocysteine Incorporation during Translation. *Journal of the American Chemical Society*.

Tokmakov, A. A., Matsumoto, E., Shirouzu, M., and Yokoyama, S. 2006. Coupled cytoplasmic transcription-and-translation--a method of choice for heterologous gene expression in Xenopus oocytes. *Journal of Biotechnology* 122:5–15.

Tokuriki, N., Jackson, C. J., Afriat-Jurnou, L., Wyganowski, K. T., Tang, R., and Tawfik, D. S. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* 3:1257.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. 2008. How protein stability and new functions trade off. *PLoS Computational Biology* 4:e1000002.

Tuerk, C., and Gold, L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.

Urabe, H., Ichihashi, N., Matsuura, T., Hosoda, K., Kazuta, Y., Kita, H., and Yomo, T. 2010. Compartmentalization in a water-in-oil emulsion repressed the spontaneous amplification of RNA by Q beta replicase. *Biochemistry* 49:1809–1813.

Vallejos, M., Ramdohr, P., Valiente-Echeverría, F., Tapia, K., Rodriguez, F. E., Lowy, F., Huidobro-Toro, J. P., Dangerfield, J. A., and López-Lastra, M. 2010. The 5'-untranslated region of the mouse mammary tumor virus mRNA exhibits cap-independent translation initiation. *Nucleic Acids Research* 38:618–32.

Vick, J. E., and Gerlt, J. A. 2007. Evolutionary potential of (beta/alpha)8-barrels: stepwise evolution of a "new" reaction in the enolase superfamily. *Biochemistry* 46:14589–97.

Wagner, A. 2008. Robustness and evolvability: a paradox resolved. *Proceedings. Biological sciences / The Royal Society* 275:91–100.

Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. 2012. Genome-scale promoter engineering by coselection MAGE. *Nature Methods* 9:591–3.

Wang, X., Minasov, G., and Shoichet, B. K. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology* 320:85–95.

Waters, E. K., Genga, R. M., Schwartz, M. C., Nelson, J. A., Schaub, R. G., Olson, K. A., Kurz, J. C., and McGinness, K. E. 2011. Aptamer ARC19499 mediates a procoagulant hemostatic effect by inhibiting tissue factor pathway inhibitor. *Blood* 117:5514–22.

Weinreich, D. M., Delaney, N. F., Depristo, M. A., and Hartl, D. L. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–4.

Wilson, C., and Keefe, A. D. 2006. Building oligonucleotide therapeutics using non-natural chemistries. *Current Opinion in Chemical Biology* 10:607–14.

Wilson, C., and Szostak, J. 1995. In vitro evolution of a self-alkylating ribozyme. *Nature*.

Wirtz, E., Hoek, M., and Cross, G. A. 1998. Regulated processive transcription of chromatin by T7 RNA polymerase in Trypanosoma brucei. *Nucleic Acids Research* 26:4626–34.

Wochner, A., Attwater, J., Coulson, A., and Holliger, P. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332:209–212.

Yin, Y. W., and Steitz, T. A. 2002. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* 298:1387–95.

Zhang, X., and Studier, F. W. 1995. Isolation of transcriptionally active mutants of T7 RNA polymerase that do not support phage growth. *Journal of Molecular Biology* 250:156–68.

Zielinski, W. S., and Orgel, L. E. 1987. Autocatalytic synthesis of a tetranucleotide analogue. *Nature* 327:346–347.