ACS SyntheticBiology

Letter

pubs.acs.org/synthbio

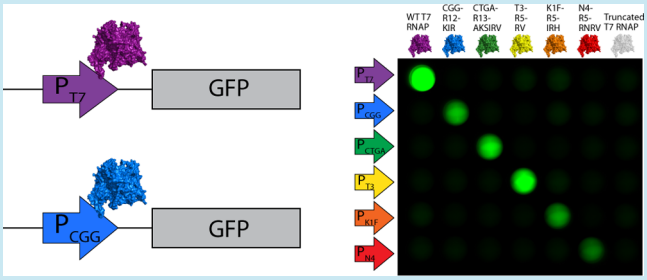
# Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for *in Vivo* or *in Vitro* Synthetic Circuitry

Adam J. Meyer,[†] Jared W. Ellefson,[†] and Andrew D. Ellington*[,†,‡,§]

[†]Institute for Cellular and Molecular Biology, [‡]Department of Chemistry and Biochemistry, and [§]Center for Systems & Synthetic Biology, University of Texas at Austin, Austin, Texas 78712, United States

Supporting Information

**ABSTRACT:** T7 RNA polymerase is the foundation of synthetic biological circuitry both *in vivo* and *in vitro* due to its robust and specific control of transcription from its cognate promoter. Here we present the directed evolution of a panel of orthogonal T7 RNA polymerase:promoter pairs that each specifically recognizes a synthetic promoter. These newly described pairs can be used to independently control up to six circuits in parallel.




**KEYWORDS:** *directed evolution, transcription, genetic circuits*

The field of synthetic biology relies on designed, synthetic circuits made from genetic parts that function robustly and orthogonally to the cell's operating system. All too often, synthetic parts are evolved, engineered, or mined to perform a particular function, but are ultimately unusable due to their weak activity or promiscuity. As parts are combined to form complex genetic circuits, each suboptimal part undermines the whole. To generate parts that behave as expected, it may be advisible to start with a well-studied and well-understood part, and tailor its function to the needs of the community.[1−4]

In particular, T7 RNA polymerase is already serving as a workhorse for synthetic biology; developing additional polymerase variants based on this scaffold is thus highly desirable. The polymerase has served as the core of synthetic transcriptional circuits such as the control of the nitrogen fixation cluster[5] and pigment production *in vivo*[6,7] and nucleic acid computation *in vitro*.[8] The ever-increasing complexity of such genetic circuits necessitates the precise control of synthetic transcriptional networks that function orthogonally. This can be achieved by engineering a repertoire of T7 RNAP variants that each specifically recognizes a synthetic promoter. Each orthogonal polymerase:promoter pair would then act as a transcriptional controller, enabling the precise and flexible temporal, spatial, or signal-dependent control of RNA expression.

T7 RNA polymerase (RNAP) has proven especially tractable for engineering and synthetic biology applications because it is a single polypeptide capable of transcribing RNA in abundance upon highly specific recognition of a short 17-bp promoter.[9,10] T7 RNAP has found wide use as a molecular tool, as it functions robustly in prokaryotic,[10,11] eukaryotic,[12−16] and *in vitro*[17] contexts, and is orthogonal to host machinery and promoters.

T7 RNAP specifically recognizes its promoter primarily through the specificity loop (residues 739 to 770), especially by interactions made between residues R746, N748, R756, and Q758 with the major groove from −7 to −11.[18] Mutations to this critical part of the promoter usually results in substantial loss of promoter recognition.[19] Promoters with single substitution that are weakly recognized by T7 RNAP can be recognized by T7 RNAPs with single mutations to N748,[20] R756,[19] or Q758.[21]

Previous attempts to create T7 RNAP mutants capable of strong and specific recognition of more divergent promoters have relied on directed evolution[22,23] or phylogenetic part mining and domain grafting.[6] These approaches have met with some success but have resulted in polymerases with very low activity or high cross-reactivity with other promoters. These limitations have restricted the utility of the polymerases for use in complex transcriptional circuitry.

In this work we utilize a directed evolution strategy, compartmentalized partnered replication[24] (CPR, Figure 1), to select for T7 RNAP mutants capable of recognizing novel promoters. Selection by CPR was followed by low throughput screening, resulting in a panel of highly active and specific T7 RNAP mutants. These mutants display comparable activity to the wild-type enzyme with limited cross-reactivity in both *in vivo* and *in vitro* transcription assays.

The basis of CPR is identifying partner-library variants that can drive the *in vivo* production of *Taq* DNA polymerase (DNAP), and then carrying out emulsion PCR with primers that allow *Taq*-mediated amplification of the genes of interest. In the current instance, in order to generate specificity variants

**Received:** August 13, 2014
**Published:** October 3, 2014

   dx.doi.org/10.1021/sb500299c | *ACS Synth. Biol.* 2015, 4, 1070−1076

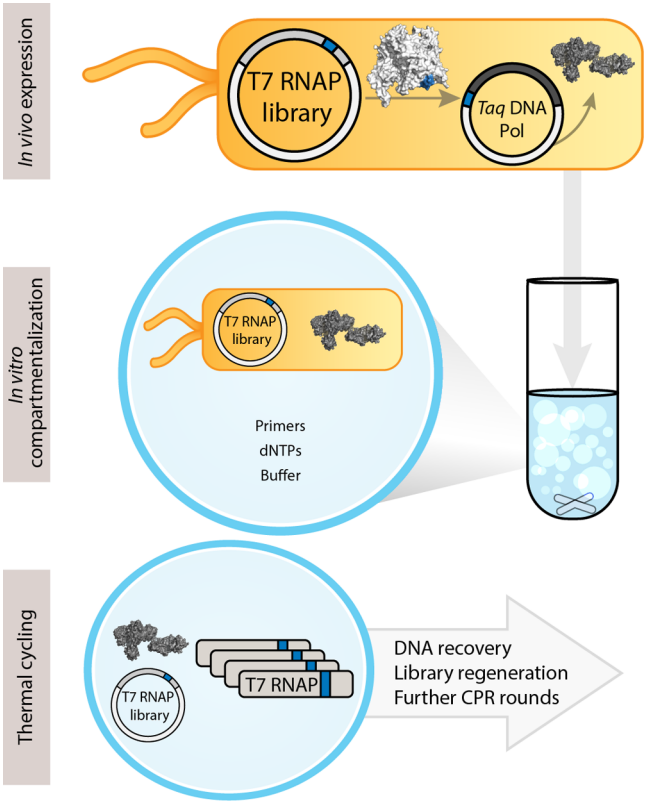



**Figure 1.** Compartmentalized partnered replication selection scheme. (Top) *E. coli* cells containing the *Taq* DNA polymerase gene under the control of a synthetic T7 promoter variant are transformed with a saturation (or error-prone) mutagenesis library of T7 RNAP mutants. Variants capable of recognizing the promoter produce *Taq* DNA polymerase protein. (Middle) Whole *E. coli* cells are compartmentalized in a water-in-oil emulsion. The aqueous droplets also contain primers, dNTPs, and *Taq* DNA polymerase buffer. (Bottom) Emulsions are thermal cycled, leading to *E. coli* cell lysis and *in vitro* PCR amplification of the T7 RNAP genes that led to the production of *Taq* DNA polymerase during the *in vivo* expression step. PCR product is recovered and prepared for the next round of selection.

of T7 RNAP we generated a strain of *E. coli* in which the production of *Taq* DNAP was dependent on the use of a synthetic T7 promoter. Individual cells were compartmentalized in a water-in-oil emulsion with 5′-biotinylated primers for PCR amplification of the T7 RNAP specificity determining region. T7 RNAP variants capable of recognizing the synthetic promoter should therefore selectively produce *Taq* DNAP protein and, in turn, be amplified by the *Taq* DNAP during PCR.

We first sought to alter the molecular recognition of T7 RNAP to drive transcription from the 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) and also different from other known orthogonal promoters ($P_{CGG}$, $P_{T3}$, $P_{K1F}$, and $P_{N4}$) at two or more of those positions.

We generated a library of T7 RNAP variants in which the six amino acids most responsible for promoter recognition (R746, L747, N748, R756, L757, and Q758)[18−21] were completely randomized by using oligonucleotides synthesized using a mix of trimer phosphoramidites.[24] The resulting library (theoretically $20^6$ or $6.4 \times 10^7$ in complexity) was transformed into *E. coli* at about 2-fold coverage. Transformed cells were then grown, induced, and emulsified. Upon thermal cycling of the emulsified reaction, cells were lysed and emulsion PCR led to selective amplification of functional specificity determining regions. The emulsion was broken with chloroform, and biotinylated amplicons were recovered by streptavidin bead-based purification. Full-length T7 RNAP genes were assembled by overlap PCR, agarose gel-purified, digested with restriction endonucleases, and ligated into an expression plasmid for further rounds of selection. The library began to converge on active sequences after four rounds of selection, and after seven rounds the library was dominated by a single variant (termed CTGA-R7-1, Supplementary Table 2) that had the sequence L747I, N748T, R756T, and Q758K (Supplementary Figure 1).

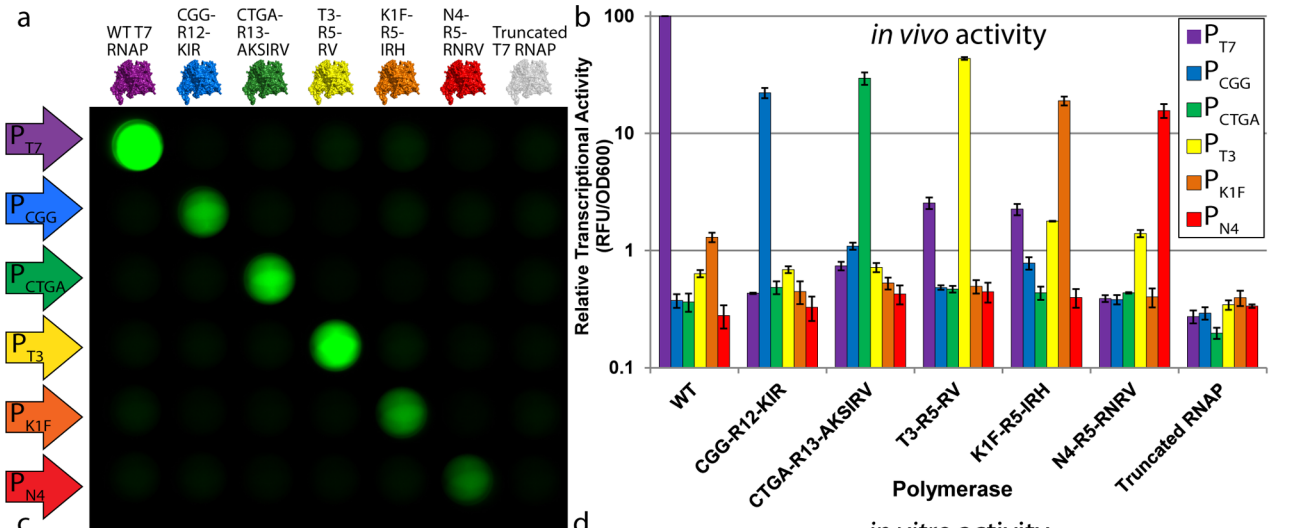
In order to further ascertain the functionality of CTGA-R7-1, we transformed an expression construct for the polymerase into an *E. coli* strain that also carried a GFP gene coupled to the $P_{CTGA}$ promoter (Supplementary Figure 1). CTGA-R7-1 produced about 1% as much GFP from $P_{CTGA}$ as the wild-type T7 RNAP produced from $P_{T7}$. This weak activity was in line with the generally low efficiencies seen by others for specificity variants.[24] We hypothesized, however, that further mutations in the regions flanking the promoter specificity determinants would provide substantive improvements in activity. We therefore subjected a larger region encompassing most of the "fingers" domain[18] (amino acid residues 663−793) to error-prone PCR (at a rate of two mutations per kb) and to CPR optimization. After six further rounds of selection (with

**Table 1. Summary of the Most Active T7 RNAP Mutants**[a]

| mutant | sequence | cognate promoter | activity on cognate |
|---|---|---|---|
| WT | WT | $P_{T7}$ TAATACGACTCACTATA | 100.0 |
| 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 |

[a]A summary of the most active variant for each promoter. The "activity on cognate" value is the average of three different experiments on three different days (each in triplicate). All values are normalized to wild-type (WT) T7 RNAP activity as 100.


dx.doi.org/10.1021/sb500299c | *ACS Synth. Biol.* 2015, 4, 1070−1076




c

| Mutant | Sequence | Cognate Promoter | *In vivo* Specificity | *In vitro* Specificity |
|---|---|---|---|---|
| WT | WT | $P_{T7}$ TAATACGACTCACTATA | 75.5 | 54.6 |
| CGG-R12-KIR | Q744K, L747V, N748H, L749I, R756E, L757M, H772R | $P_{CGG}$ TAATACCGGTCACTATA | 34.3 | 45.8 |
| CTGA-R13-AKSIRV | V725A, Q744K, L747I, N748S, L749I, R756T, Q758K, H772R, E775V | $P_{CTGA}$ TAATACCTGACACTATA | 33.8 | 23.8 |
| T3-R5-RV | T745K, N748D, L749M, M750I, H772R, E775V | $P_{T3}$ TAATAACCCTCACTATA | 31.7 | 25.5 |
| K1F-R5-IRH | L749I, Q754S, R756N, I761V, H772R, Q786H | $P_{K1F}$ TAATAACTATCACTATA | 9.5 | 20.3 |
| N4-R5-RNRV | L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V | $P_{N4}$ TAATAACCCACACTATA | 9.8 | 13.5 |

**Figure 2.** Assays of the most specific T7 RNAP mutants. The most specific variant from each selection was assayed for use of each of the six described promoters. (a,b) Six different *E. coli* strains, each of which contained the GFP gene under the control of one of the six T7 promoter variants, are transformed with each polymerase mutant. (a) Cells were grown, induced, and imaged. (b) Cells were grown, induced, and measured for fluorescence (excitation 481 nm/emission 507 nm) and $OD_{600}$. Values shown are the average of three independent cultures. Error bars represent the standard error. All values are normalized to wild-type T7 RNAP activity as 100. (c) A summary of the most specific variant for each promoter. "Specificity" refers to the activity of a mutant with its cognate promoter relative to the highest activity of the mutant with any noncognate promoter tested. The *in vivo* value is the average of two different experiments on two different days (each in triplicate). (d) Each polymerase:promoter combination was allowed to drive the *in vitro* transcription of the Spinach aptamer.[25] Spinach fluorescence (excitation 469 nm/emission 501 nm) was recorded after a 30 min incubation as a measure of relative activity. Values shown are the average of three independent transcription reactions. Error bars represent the standard error. All values are normalized to wild type T7 RNAP activity as 100.

error-prone PCR at rounds 8, 10, 11, and 13) 288 clones from the library were assayed for *in vivo* GFP expression from $P_{CTGA}$. Active clones (judged by fluorescence readings, not shown) were sequenced, and several mutations were found to occur frequently in the population (I681L, Q744K, H772R, E775V, and to a lesser extent V725A, T748S, and L749I, Supplementary Figure 2). 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 (Supplementary Figure 3). The clone CTGA-R13-AKSRV (Table 1 and Supplementary Table 2) that contained the mutations V725A, Q744K, T748S, H772R, and E775V on top of the parental CTGA-R7-1 (L747I, N748T, R756T, and Q758K) was found to be highly active (40% activity on $P_{CTGA}$ relative to the wild-type combination) and was subjected to further analyses; see below.

Previously, T7 RNAP variants capable of utilizing $P_{T3}$, $P_{K1F}$, and $P_{N4}$ promoters have been described[6] (T7(T3): T745K, N748D, L749M, M750I. T7(K1F): Q754S, R756N, I761V. T7(N4): L747I, N748D, L749C, M750V, F751I, Q754T, F755H, L757M, Q758A, P759L; Supplementary Table 2). However, upon initial assay we discovered that, while these polymerases were specific for their respective promoters, their overall activity was low, only about 1−10% of the activity of the wild-type polymerase:promoter pair.[24] These polymerases were created by grafting the promoter specificity loop of related bacteriophage RNA polymerases (with different promoter specificities) onto a T7 RNAP scaffold. Even though it is remarkable that such grafting was able to confer specificity, the differences in the structural context likely limited activity.

To increase activity, 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, and combinatorial mutants were generated and assayed (Supplementary Figures 4−9 and Table 2). 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 1). These variants demonstrated 160.3%, 76.2%, and 25.2% activity in comparison to the wild-type T7 RNAP with its promoter, which was respectively a 20-, 119-, and 91-fold improvement over the parental enzymes.

We hypothesized that selecting for very active variants would, in turn, lead to very specific variants, based on previous studies.[26,27] For example, Tawfik and colleagues showed that as a hydrolase gained activity with a new substrate, it also gained specificity toward that substrate relative to the original substrate.[26] We therefore tested each of the four variant polymerases, the previously described CGG-R12-KIRV[24] and the wild-type T7 RNAP against each of their cognate promoters (6 × 6). Each polymerase was separately transformed into six distinct cell lines in which the six different promoters drove the production of GFP (Supplementary Figures 10 and 11). Each polymerase showed at least a 3.8-fold selectivity for its cognate promoter over any other promoter (as judged by the relative protein yield driven from each promoter); however, there were several cases of cross-reactivity. Specifically, T3-R5-RRVH was quite cross reactive with $P_{T7}$, showing only a 6.7-fold preference for $P_{T3}$, while K1F-R5-KIKR prefers $P_{K1F}$ by only 3.8-fold over $P_{T7}$ and 7.3-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.[24]

While we have seen that selecting and screening for highly active mutants can yield highly specific mutants, it was also possible that our original hypothesis was incorrect and that mutations that improve efficiency do so via nonspecific mechanisms (for example, by forming an interaction with the phosphate backbone). Therefore, we rescreened our previous panel of variants from each selection not only for their ability to drive GFP from their cognate promoters but also for their cross-reactivities (Supplementary Figure 12−16). Several additional candidates that were very active yet less cross-reactive were obtained. The polymerase CGG-R12-KIR (Q744K, L747V, N748H, L749I, R756E, L757M, H772R) was 31% active on its cognate promoter (compared to wild-type) and had at least a 34.3-fold selectivity. CTGA-R13-AKSIRV (V725A, Q744K, L747I, N748S, L749I, R756T, Q758K, H772R, E775V) was 27% active with $P_{CTGA}$ and had a 33.8-fold preference for its cognate promoter. T3-R5-RV (T745K, N748D, L749M, M750I, H772R, E775V) was 39.3% active with $P_{T3}$ with a 46-fold preference; K1F-R5-IRH (L749I, Q754S, R756N, I761V, H772R, Q786H) was 18.5% active with $P_{K1F}$, with at least 9.5-fold preference; and N4-R5-RNRV (L747I, N748D, L749C, M750V, F751I, Q754T, F755R, L757M, Q758A, P759L, D770N, H772R, E775V was is 14.7% active on $P_{N4}$ and preferred it by 9.8-fold.

In order to confirm the orthogonality of these polymerase:promoter pairs *in vivo*, we again assayed each of them on the set of six promoters. This new panel of polymerases proved to be truly orthogonal, as no polymerase showed more than 2.5% cross-reactivity *in vivo* (Figure 2a−c). Further, to demonstrate that CPR did not just yield polymerases that were orthogonal *in vivo*, each polymerase was purified and assayed for its ability to drive the transcription of the fluorescent aptamer Spinach.[25] Fluorescence readings were recorded as a measure of transcriptional activity and thus as an indicator of promoter recognition *in vitro*. The evolved polymerases again show high activity with their cognate promoters (ranging from 33% to 64% activity, relative to WT with its cognate promoter) and have excellent specificity (a 13.5- to 45.8-fold preference for their respective cognate promoters; Figure 2c,d).

These results demonstrate the unequivocal role of directed evolution in the development of synthetic biology.[28] The relatively low activity of the initially evolved or engineered polymerases and their subsequent improvement by directed evolution offers important lessons with regards to the semirational alteration of molecular function. Structural and mutation studies suggest the importance of R746, N748, R756, and Q758,[18−21] yet mutations outside of these residues were required for CGG-R7-8 and CTGA-R7-1 to achieve activity comparable to WT T7 RNAP. CPR yielded more than a 50-fold increase in activity over the originally selected mutants. Similarly T7(T3), T7(K1F), and T7(N4) were designed with structural and phylogenetic considerations,[6] but they could also be improved by directed evolution as evidenced by their 2- to 40-fold improvement in activity and at minimum 2-fold improvement in specificity. It is interesting to note that some mutations outside of the immediate specificity loop were able to improve the activities of several mutants, most notably Q744K, H772R, and E775V. It is unclear whether these mutations facilitated nonspecific polymerase:promoter interactions, enhanced polymerase expression or folding, or otherwise increased polymerase activity by an unknown mechanism.

These results also demonstrate the utility of CPR as a directed evolution method. While there are many possible methods for the evolution or engineering of such systems,[6,23,29] these have by and large not produced polymerases of nearly the same activity nor specificity. The effectiveness of CPR stems from its decoupling of organismal and molecular fitness, and its sensitivity to small differences in *Taq* DNAP levels over a wide range of concentrations.[24] This allows even weakly active variants to survive the early rounds of evolution, but also leads to highly active variants outcompeting moderately active ones.

Modular control of synthetic transcriptional circuitry is critical for engineering complex gene expression *in vivo* and can contribute to the development of transcription-based nucleic acid circuits *in vitro*. Starting with the well-characterized T7 RNA polymerase, we have generated a panel of six highly active and orthogonal polymerase:promoter pairs. This "hexacore" set of transcriptional regulators should become a standard for the field of synthetic biology, just as other common protein parts such as the Lac repressor and LuxR quorum-sensing transcription activator have been.

The high activity of T7 RNA polymerase and its derivatives can potentially lead to toxicity, especially if multiple polymerases are expressed in the same cell. One possible solution to this problem is to limit the total number of active polymerases in a given cell. As promoter specificity is conferred by the C-terminal domain, each polymerase may be split into a common "core" N-terminal domain and a $\sigma$-factor like C-terminal domain. A fixed number of N-terminal fragments sets a cap on the total number of full length polymerases that can

form and thus limits toxicity, and by altering the relative levels of each C-terminal fragment one can tune the expression from each promoter. Indeed, one of the polymerases described has already been used in the instantiation of such a "resource allocation" system.[30]

## MATERIALS AND METHODS

**Library Design and Selection.** For the selection of an RNAP mutant capable of recognizing $P_{CTGA}$, site saturation mutagenesis was used to randomize the residues R746, L747, N748, R756, L757, and Q758 of the T7 RNAP promoter specificity loop. The T7 RNAP open reading frame was amplified by PCR using degenerate oligonucleotides. An oligonucleotide spanning the promoter specificity loop (RAH.01, Supplementary Table 1) was synthesized, with degeneracy introduced by the use of degenerate trimer phosphoramidites (Glen Research). This oligonucleotide was PCR amplified using primers AJM.01 and AJM.02, and the resulting product was used to assemble the full length open reading frame. The full length T7 RNAP library was digested with *Bam*HI and *Hin*dIII and ligated into the pQE-RSS backbone (in which a strong T5-lac promoter drives T7 RNAP), as previously detailed.[24]

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 driving *Taq* DNA polymerase were generated using Mega-primer PCR or isothermal assembly. BL21 gold cells (Agilent) were transformed with pACYC-Taq (or its derivative with altered promoter) and grown overnight. Two hundred and fifty microliters of this culture was subcultured into 20 mL of 2× YT growth medium and grown at 37 °C for 2 h (reaching an $OD_{600}$ of approximately 0.5). The culture was then centrifuged for 10 min 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).

One hundred microliters of overnight transformation cultures were subcultured in 2 mL of 2× YT medium, grown for 1 h ($OD_{600-0.6}$) and induced with 0.05 mM IPTG at 37 °C for 4 h. Two hundred microliters of the induced culture was centrifuged for 10 min at 5000*g*. The supernatant was removed and cells were gently resuspended in 20 μL of 10× PCR buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3, 15 mM $MgCl_2$) 10 μL of dNTP mix (4 mM each), 4 μL of each primer (AJM.03 and AJM.04 20 uM), and 162 μL of water. Emulsification was performed by slowly adding resuspended cells to 600 μL of spinning oil mix (438 μL of Tegosoft DEC (Evonik), 42 μL of AbilWE09 (Evonik), and 120 μL of mineral oil (Sigma)). The oil mixture was constantly spun in a tube (Sarstedt, 13 mL, 95 mm × 16.8 mm) on ice using a stirbar (Spinplus, 9.5 mm × 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 1 min interval and spun for an additional 4 min. The emulsified cells were thermal cycled (95 °C, 3 min, 20 cycles [95 °C, 30 s; 55 °C, 30 s; 72 °C, 5 min]; 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. First, it was centrifuged for 5 min at 10 000*g*, and the oil (upper) phase was removed. Second, 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 of 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 for 2 min at 16 000*g*, 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, we used primers with 5′ biotin groups in the emulsion PCR step. Biotinylated PCR product was purified using streptavidin coated beads (MyOne Streptavidin C1 Dynabeads, Invitrogen) and used as a template for reamplification using nested primers (AJM.05 and AJM.06). This PCR product was gel purified and used in an assembly PCR, thus regenerating the full-length T7 RNAP ORF. This was followed by digestion with *Bam*HI and *Hin*dIII and ligation into the pQE-RSS backbone.

In rounds 8 to 13 of the $P_{CTGA}$ selection (after isolation of CTGA-R7-1) 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) using primers AJM.07 and AJM.08. 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 with *Bam*HI and *Hin*dIII and ligation into the pLUV-RSS backbone (in which the strong T5 promoter was replaced with the moderate-strength promoter, LacUV5).

Error prone PCR was performed on the larger region prior to $P_{CTGA}$ 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 (AJM.07 and AJM.08), 2 ng/μL template, and 0.8 U/μL Taq DNAP (New England Biolabs) and was thermal cycled (95 °C, 4 min, 25 cycles [95 °C, 30 s; 55 °C, 30 s; 72 °C, 2 min]; 72 °C, 5 min). This achieved the expected one mutation per 500 bp. 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 with *Bam*HI and *Hin*dIII and ligation into the pLUV-RSS backbone.

Individual variants from CTGA-R7, CTGA-R15, T3-R5, K1F-R5, and N4-R5 were sequenced and analyzed using Geneious software (Biomatters, Ltd.).

***In Vivo* GFP Assay.** Combinations of consensus mutations were made using Mega-primer PCR or isothermal assembly. Purified T7 RNA polymerase plasmid was electroporated into BL21-gold cells containing pACYC derivatives in which T7 promoter variants drive GFP production. Transformations were grown at 37 °C overnight. One hundred microliters of the culture was grown in 2 mL of 2× YT medium at 37 °C for 1 h ($OD_{600-0.6}$) and induced with 0.05 mM IPTG for 4 h. 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 $OD_{600}$ on a Synergy-HT plate reader (Bio-Tek) and GFP fluorescence (excitation 481 nm/emission 507 nm) on a Safire monochromator (Tecan). Images of T7 RNA polymerase-driven GFP expression shown in Figures 2a and Supplementary Figure 10 were generated by pelleting 2 mL of induced culture, decanting the supernatant, and resuspending cells in 500 μL of PBS. The resuspended cells

were excited with a transilluminator (475 nm excitation) and imaged with a FluorChem Q (Cy3-filter, Protein Simple).

**T7 RNA Polymerase Purification.** For *in vitro* transcription assays, T7 RNAP variants were purified by standard Ni-NTA 6× His (N-terminal) methods. The plasmid pQE-T7RNAP (or a derivative thereof for T7 RNA polymerase mutants) was transformed in BL21-gold (Agilent). Cells were grown in 2× YT media at 37 °C until an $OD_{600-0.7-0.8}$ was reached. Cells were induced for 4 h 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, and 5 mM imidazole). Resuspended cells were lysed via sonication on ice using 50% probe amplitude for 3 min (one second on; one second off). Cell debris was pelleted by centrifugation for 30 min at 10 000*g*. His-tagged T7 RNAP was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over a 1 mL (bead volume) Ni-NTA 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, and 20 mM imidazole). T7 RNAP was eluted off the column by the addition of four column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, and 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.[25] Templates were prepared by PCR and agarose gel purification using QIAquick Gel Extraction Kit (Qiagen). Each template was made from a universal reverse primer (AJM.09) and a unique forward primer (AJM.10−AJM.15).

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[25] with 0.5 $\mu$M of the appropriate T7 RNAP and 0.5 $\mu$M of the appropriate DNA template. Reactions were incubated for 30 min at 37 °C with Spinach fluorescence (excitation 469 nm/emission 501 nm) reading taken every 3 min in a Safire monochromator (Tecan).

## ■ ASSOCIATED CONTENT

### Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*Tel: 512-471-6445. E-mail: andy.ellington@mail.utexas.edu.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Welch Foundation (F-1654) and the National Security Science and Engineering Faculty Fellowship (FA9550-10-1-0169). We would like to thank Randall Hughes for the expert synthesis of the degenerate trimer containing oligonucleotide.

## ■ REFERENCES


(1) Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., and Holliger, P. (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol. 22*, 755−9.

(2) Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci. 20*, 448−455.

(3) Nevoigt, E., Fischer, C., Mucha, O., Matthäus, F., Stahl, U., and Stephanopoulos, G. (2007) Engineering promoter regulation. *Biotechnol. Bioeng. 96*, 550−8.

(4) Levy, M., and Ellington, A. D. (2008) Directed evolution of streptavidin variants using in vitro compartmentalization. *Chem. Biol. 15*, 979−89.

(5) Temme, K., Zhao, D., and Voigt, C. A. (2012) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. U.S.A. 109*, 7085−90.

(6) 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 Res. 40*, 8773−81.

(7) Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature 460*, 894−8.

(8) Kim, J., and Winfree, E. (2011) Synthetic in vitro transcriptional oscillators. *Mol. Syst. Biol. 7*, 465.

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

(10) Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol. 189*, 113−30.

(11) Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. (1996) A T7 promoter-specific, inducible protein expression system for Bacillus subtilis. *Mol. Gen. Genet. 250*, 230−236.

(12) 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. *J. Biotechnol. 122*, 5−15.

(13) Wirtz, E., Hoek, M., and Cross, G. A. (1998) Regulated processive transcription of chromatin by T7 RNA polymerase in Trypanosoma brucei. *Nucleic Acids Res. 26*, 4626−34.

(14) Davidson, E. A., Meyer, A. J., Ellefson, J. W., Levy, M., and Ellington, A. D. (2012) An in vitro autogene. *ACS Synth. Biol. 1*, 190−196.

(15) 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. *Proc. Natl. Acad. Sci. U.S.A. 83*, 8122−6.

(16) 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. *Proc. Natl. Acad. Sci. U.S.A. 91*, 7301−5.

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

(18) 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.

(19) 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.

(20) Raskin, C. A., Diaz, G. A., and McAllister, W. T. (1993) T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl. Acad. Sci. U.S.A. 90*, 3147−51.

(21) Rong, M., He, B., McAllister, W. T., and Durbin, R.K. (1998) Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A. 95*, 515−9.

(22) Chelliserrykattil, J., Cai, G., and Ellington, A. D. (2001) A combined in vitro/in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnol. 1*, 13.
(23) Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules. *Nature 472*, 499−503.
(24) 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. *Nat. Biotechnol. 32*, 97−101.
(25) Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011) RNA mimics of green fluorescent protein. *Science 333*, 642−646.
(26) 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. *Nat. Commun. 3*, 1257.
(27) Matsumura, I., and Ellington, A. D. (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol. 305*, 331−9.
(28) Karanicolas, J., Corn, J. E., Chen, I., Joachimiak, L. A., Dym, O., Peck, S. H., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., Montelione, G. T., Spiegel, C. P., Liu, D. R., and Baker, D. (2011) A de novo protein binding pair by computational design and directed evolution. *Mol. Cell 42*, 250−60.
(29) Chelliserrykattil, J., and Ellington, A. D. (2004) Evolution of a T7 RNA polymerase variant that transcribes 2′-O-methyl RNA. *Nat. Biotechnol. 22*, 1155−60.
(30) Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D., and Voigt, C. A. (2014) A "resource allocator" for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol. 10*, 742.