

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

**A Method for Cost-Effective and Rapid Characterization of Engineered T7-based Transcription Factors by Cell-Free Protein Synthesis Reveals Insights into the Regulation of T7 RNA Polymerase-Driven Expression**

John B. McManus<sup>a,b</sup>, Richard M. Murray<sup>b</sup>, and Peter A. Emanuel<sup>c</sup>, Matthew W. Lux<sup>c,\*</sup>

<sup>a</sup> US Army Combat Capabilities Development Command, Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, MD 20783

<sup>b</sup> California Institute of Technology, Biology and Biological Engineering, 1200 East California Blvd, Pasadena, CA 91125

<sup>c</sup> US Army Combat Capabilities Development Command Chemical Biological Center, 8567 Ricketts Point Road, APG, MD 21010

\*Corresponding Author: [matthew.w.lux.civ@mail.mil](mailto:matthew.w.lux.civ@mail.mil)

24 **Abstract**

25 The T7 bacteriophage RNA polymerase (T7 RNAP) serves as a model for understanding RNA synthesis, as  
26 a tool for protein expression, and as an actuator for synthetic gene circuit design in bacterial cells and  
27 cell-free extract. T7 RNAP is an attractive tool for orthogonal protein expression in bacteria owing to its  
28 compact single subunit structure and orthogonal promoter specificity. Understanding the mechanisms  
29 underlying T7 RNAP regulation is important to the design of engineered T7-based transcription factors,  
30 which can be used in gene circuit design. To explore regulatory mechanisms for T7 RNAP-driven  
31 expression, we developed a rapid and cost-effective method to characterize engineered T7-based  
32 transcription factors using cell-free protein synthesis and an acoustic liquid handler. Using this method,  
33 we investigated the effects of the tetracycline operator's proximity to the T7 promoter on the regulation  
34 of T7 RNAP-driven expression. Our results reveal a mechanism for regulation that functions by  
35 interfering with the transition of T7 RNAP from initiation to elongation and validates the use of the  
36 method described here to engineer future T7-based transcription factors.

37 **Highlights**

- 38 • Development of a rapid and cost-effective method for screening synthetic promoters.
- 39 • Insights into the regulation of engineered T7-based transcription factors and T7 RNAP enzyme  
40 kinetics.
- 41 • Validation of this method by comparison with the T7 RNAP kinetic model.

42

43

44

45 **Key words:** cell-free protein synthesis, TXTL, T7 RNA Polymerase, tetracycline

## 46 Introduction

47 Since its isolation in 1970 [1], the T7 bacteriophage RNA polymerase (T7 RNAP) has become a model for  
48 understanding RNA synthesis, as well as emerging as an important tool for protein expression [2] and as  
49 an actuator for synthetic gene circuit design in bacterial cells and cell-free extract [3,4]. The T7 RNAP is  
50 a 98 kDa, single-subunit enzyme that requires no additional protein factors to perform the complete  
51 transcriptional cycle [5]. This transcriptional cycle can be broken into three phases: binding, initiation,  
52 and elongation. During binding, T7 RNAP specifically recognizes the T7 promoter [5], and it has little  
53 affinity for other sequences, even the closely-related T3 promoter [6]. T7 RNAP then performs several  
54 rounds of abortive transcription, producing transcripts 10-12 nucleotides in length [7,8]. The enzyme  
55 undergoes a conformational change that marks its transition from initiation to the highly-processive  
56 elongation phase [9], transcribing RNA at a rate of 43 nucleotides per second [10], and producing  
57 transcripts greater than 10 kb in size [11]. Comparing these characteristics with the native *Escherichia*  
58 *coli* RNAP, which requires multiple subunits and recognizes orthogonal promoters, demonstrates the  
59 utility of T7 RNAP as an orthogonal tool. T7 RNAP was exploited early and often in synthetic gene circuit  
60 design owing to its ability to partially insulate circuit function from the host metabolism [3,4,12–14]. On  
61 average, there are only 2,000 native RNAP molecules per *E. coli* cell [15], and thus, the fluctuation of  
62 intracellular resources and variation in drag on metabolism, corresponding with growth phase and  
63 conditions, can complicate the predicted function of a gene circuit. Having an orthogonal tool set of  
64 well-characterized actuators, such as T7 RNAP, is vital to the ability to accurately predict gene circuit  
65 function [12].

66 One major drawback of the T7 RNAP system is the lack of regulatory mechanisms, beyond regulating the  
67 expression level of T7 RNAP, and a small set of synthetic promoters [4]. This stands in stark contrast to  
68 the large libraries of available and regulatable native promoters, including sets that are highly  
69 characterized and lack crosstalk [16]. Understanding regulatory mechanisms for T7 RNAP is important

70 to the discovery of new engineered T7-based transcription factors that can be used in gene circuit  
71 design. Here we developed a rapid and cost-effective method to characterize promoter-operator  
72 combinations using cell-free protein synthesis and an acoustic liquid handler. We chose, as a model to  
73 test this method, the tetracycline regulatory (*tet*) system. The *tet* system functions to regulate protein  
74 expression when the homodimeric tetracycline repressor protein (TetR) binds to the tetracycline  
75 operator (*tetO*) sequence, resulting in downregulated transcription [17,18]. The *tet* system has been  
76 exploited in synthetic gene circuit design to regulate both host RNAP and T7 RNAP-driven expression  
77 [4,17,19–22]. In the case of host RNAP regulation, the *tetO* sequence can be placed between the -10  
78 and -35 regions, resulting in relatively greater dynamic ranges when compared with T7 RNAP regulation  
79 by the same system [23]. Due to the single binding region of the T7 promoter, the *tetO* sequence can  
80 only be placed up or downstream from the T7 promoter. Thus, we investigated the effects of proximity  
81 of the *tetO* sequence to the T7 promoter on the regulation of T7 RNAP-driven expression.

82 We show that, irrespective of the position of *tetO* within the first 13 bp downstream of the T7 promoter,  
83 T7 RNAP-driven expression is downregulated equally by TetR. Conversely, placing the *tetO* sequence  
84 upstream from the T7 promoter shows nearly equal expression in the presence or absence of TetR. Our  
85 results suggest that *tet* regulation of T7 RNAP occurs by interfering with the initiation phase of T7 RNAP.  
86 We believe that this finding reveals characteristics regarding regulation of T7 RNAP that can be used in  
87 the engineering of T7-based transcription factors, and that such engineered transcription factors can be  
88 rapidly characterized by the methods described herein.

## 89 **Materials and Methods**

### 90 ***PCR-Amplification of Linear Template***

91 Linear template for use in cell-free protein synthesis was amplified in two rounds of PCR using a single  
92 universal reverse primer and a set of two overlapping forward primers (one for each round of PCR)

93 containing different positional pairings of T7-*tetO* sequences. PCR products were isolated by the Qiagen  
94 gel purification kit (Qiagen), quantified using a nanodrop spectrophotometer, diluted to 20 nM, and  
95 stored at -20°C until use (Fig. S2).

#### 96 ***Cloning of Linear Constructs into PY71***

97 Linear templates for each *tetO* position were PCR-amplified using forward and reverse primers to add  
98 homology regions to the 5' and 3' ends in order to facilitate seamless cloning into the pY71 vector (Fig.  
99 S1). Linear templates were cloned into the pY71 vector using the In-fusion seamless cloning kit  
100 (Clontech). Circular templates were amplified in *E. coli* DH5 $\alpha$  cells and isolated by Qiagen miniprep kit  
101 (Qiagen), quantified using a nanodrop spectrophotometer, diluted to 20 nM, and stored at -20°C until  
102 use.

#### 103 ***Preparation of Cell-Free Extract and Cell-Free Reactions***

104 Cell-free extract was prepared according to Sun et al. [24] with the modification that cells were lysed by  
105 French pressure cell at 10,000 psi rather than by bead beating. Cell-free reactions were prepared  
106 according to Sun et al. [24]. Except where noted, a reaction contained energy buffer (3.3  $\mu$ L), extract  
107 (4.2  $\mu$ L), T7 polymerase (0.12  $\mu$ L from 13 mg/mL stock), malachite green (0.2  $\mu$ L from 10 mM stock),  
108 GamS (0.15  $\mu$ L from 207  $\mu$ M stock), DNA template (1  $\mu$ L), and water (0.03  $\mu$ L). Reactions were  
109 distributed by electronic pipette and TetR dilutions were distributed into cell-free reactions in volumes  
110 of 1  $\mu$ L using an Echo 525 acoustic liquid handler (Labcyte Inc.).

111 sfGFP expression, in a total volume of 10  $\mu$ L, was measured in black, clear-bottom 384 well plates  
112 (Greiner). Reactions were performed at 30°C and terminated after 12 h. sfGFP expression was  
113 measured by fluorescence in a Biotek H1 plate reader at 415 nm (ex) / 528 nm (em). Where indicated,  
114 RFU values were converted to protein concentration using calibration curves generated with purified  
115 sfGFP (a gift from Scott Walper, Naval Research Lab).

### 116 ***Purification of TetR Protein***

117 The *tetR* gene was PCR amplified from *E. coli* DH5 $\alpha$  total DNA and seamlessly cloned into a pET-22B  
118 vector containing a C-terminal hexahistidine tag and expressed in *E. coli* BL21(DE3) cells. Cells (1 g) were  
119 resuspended in lysis buffer (5 mL) (50 mM Tris-Cl, 500 mM NaCl, 5 mM imidazole, pH 8.0) and lysed by  
120 sonication. The lysate was clarified by centrifugation at 15,000 x g for 30 minutes at 4°C. The  
121 supernatant (5 mL) was mixed with Ni-NTA resin (1 mL) (Sigma Aldrich), incubated at 4°C for 1 h, and  
122 loaded into a column. The resin was washed with 10 column volumes of wash buffer (50 mM Tris-Cl,  
123 500 mM NaCl, 25 mM imidazole, pH 8.0). TetR was collected with three column volumes of elution  
124 buffer (50 mM Tris-Cl, 500 mM NaCl, 250 mM imidazole, pH 8.0) and concentrated to 1.5 mL using a 3  
125 kDa cut-off centrifugal concentrator (Millipore). TetR was dialyzed against 2 L of dialysis buffer (50 mM  
126 NaHPO<sub>4</sub>, 100 mM NaCl, 2% DMSO, pH 7.5) for 1 h at 4°C then 2 L of dialysis buffer overnight at 4°C. TetR  
127 was then centrifuged at 14,000 x g for 10 minutes to remove precipitate and quantified by absorbance  
128 at 280 nm using the molar extinction coefficient (15,845 M<sup>-1</sup> cm<sup>-1</sup>). Working TetR dilutions were first  
129 prepared, using dialysis buffer, by serial dilution, then flash frozen to reduce error and stored at -80°C  
130 until use.

### 131 ***Purification of GamS Protein***

132 The GamS protein was expressed from the pBad vector according to Sun et al. [25] and purified by nickel  
133 affinity as described for the TetR protein. GamS was dialyzed against 2 L of dialysis buffer (50 mM  
134 NaHPO<sub>4</sub>, 1 mM DTT, 1 mM EDTA, 100 mM NaCl, 2% DMSO, pH 7.5) for 1 h at 4°C then 2 L of dialysis  
135 buffer overnight at 4°C. GamS was then centrifuged at 14,000 x g for 10 minutes to remove precipitate  
136 and quantified by absorbance at 280 nM using the molar extinction coefficient (11,460 M<sup>-1</sup> cm<sup>-1</sup>). GamS  
137 was diluted in dialysis buffer to 207  $\mu$ M, flash frozen in liquid nitrogen, and stored at -80°C until use.

### 138 ***Purification of T7 RNA Polymerase***

139 T7 RNAP was expressed and purified according to Swartz et al. [26]. T7 RNAP was diluted to 13 mg/mL,  
140 flash frozen in liquid nitrogen, and stored at -80°C until use.

#### 141 ***Curve Fitting and Statistical Analysis***

142 sfGFP expression curves were fit to a sigmoidal equation:

$$y = \frac{a}{1 + e^{-(x-x_0)/b}}$$

143  
144 Four parameter logistic regressions were performed using Prism software (Graphpad). The maximum  
145 repression value for each construct was expressed as the difference between the top and bottom values  
146 determined by the four parameter logistic fit.

147 Statistical analysis was performed by a one-way ANOVA test, a two-way ANOVA test, or by Welch's t-  
148 test, as indicated in the results and figures, using Prism software (Graphpad). For ANOVA tests, Tukey's  
149 method was applied to determine statistical significance between data at each *tetO* position.

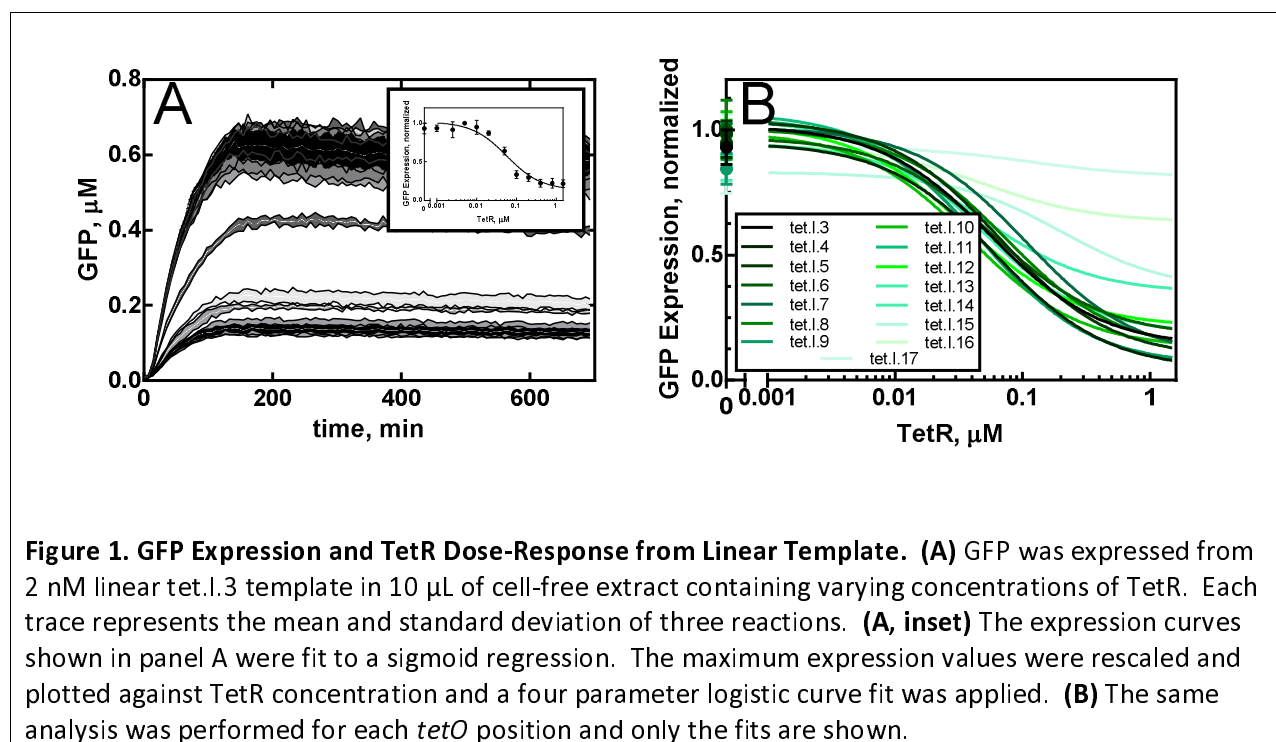
#### 150 **Results**

##### 151 ***TetO Represses T7 RNAP Equally when It is Positioned within the First 13 Bases of the T7 Transcript***

152 In order to effectively explore and understand the design space of engineered T7-based transcription  
153 factors, we aimed to develop a method that would allow for rapid and cost-effective characterization of  
154 promoter-operator combinations. To do this, we generated linear template by PCR-amplifying the sfGFP  
155 gene from the sfGFP-containing pY71-GFP plasmid using one universal reverse primer and different  
156 forward primers containing spatial combinations of T7-*tetO* across a stretch of standard base pairs (Fig.  
157 S2 and S3). We chose the *tetO*<sub>2</sub> sequence for this work because TetR has shown approximately twice  
158 the affinity for *tetO*<sub>2</sub> over *tetO*<sub>1</sub> [27]. We chose cell-free extract, which is prepared from *E. coli* cells, in-  
159 house, as a rapid and cost-effective medium for the measurement of protein expression [24,25]. This

160 method circumvents the need to clone, transform, and measure expression in whole cells, allowing for  
161 the characterization of many promoter-operator combinations in a single day. Additionally, it allowed  
162 us to directly probe the effects of the *tetO* position using purified TetR. For nomenclature purposes,  
163 constructs were designated according to the number of base pairs that the 5' end of the *tetO* sequence  
164 lies away from the transcriptional start site of the T7 promoter sequence (or the length of the transcript,  
165 in bases) (Fig. S3).

166 In order to prevent the degradation of linear template, GamS protein was added to the reaction mixture  
167 [25]. Different concentrations of TetR were added to reaction wells using an acoustic liquid handler.  
168 Cell-free reactions were run for 12 h, expression curves were fit using a sigmoidal regression, and the  
169 maximum sfGFP expression values were used for further evaluation (Fig. 1A).



170

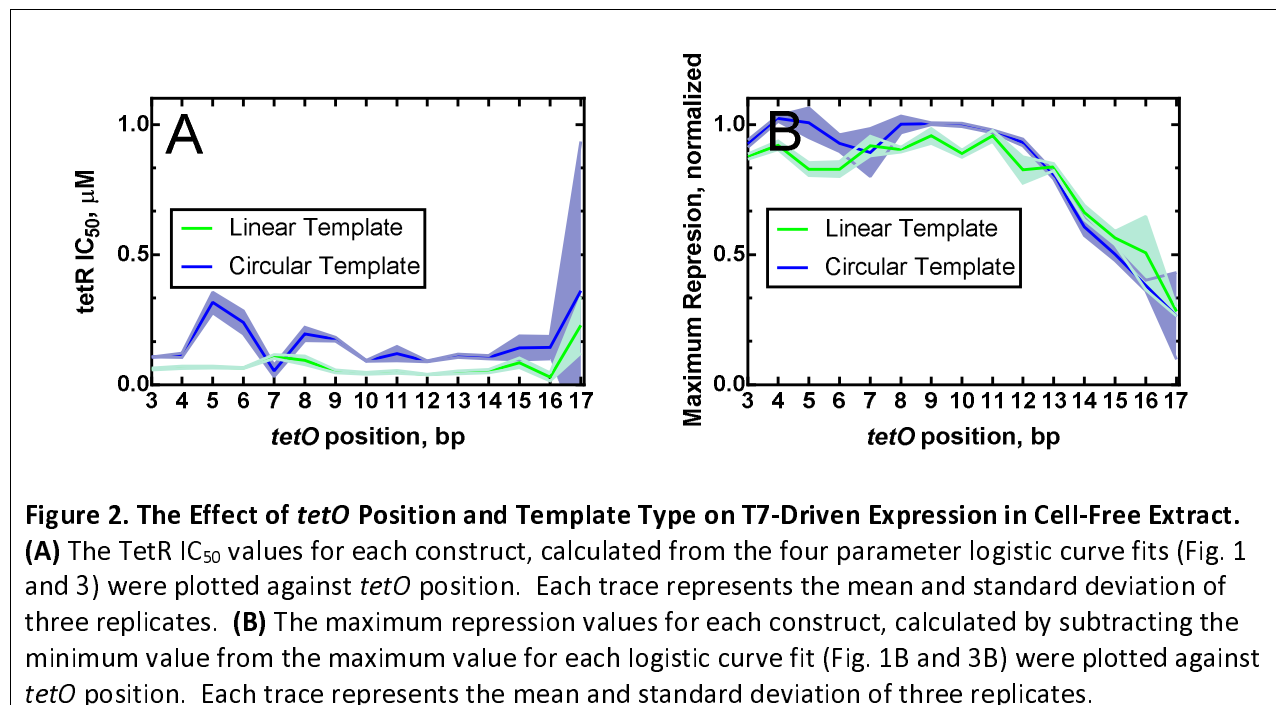
171 Despite that the same concentration of template was added to each reaction, we observed variation in  
172 the maximum sfGFP expression using different templates, ranging from 0.4 µM to 1.4 µM, in the



173 absence of TetR (Fig. S4A). To determine whether these fluctuations were simply due to the positioning  
174 of the *tetO* sequence, we tested two different preparations of each template with no TetR present (Fig.  
175 S4A). When a one-way ANOVA test was applied to the expression data from lot 1, no apparent pattern  
176 that might indicate an effect of the *tetO* position on expression became apparent (Fig. S4C). A two-way  
177 ANOVA test was applied to the expression data of both lots in order to determine statistical significances  
178 in the difference in expression at each *tetO* position. A pattern indicating no significance, running along  
179 the diagonal in Figure S4E, would illustrate the reproducibility of the variation in expression between  
180 lots (Fig. S4A). However, such a pattern does not emerge, suggesting that variation cannot be explained  
181 by the *tetO* position alone.

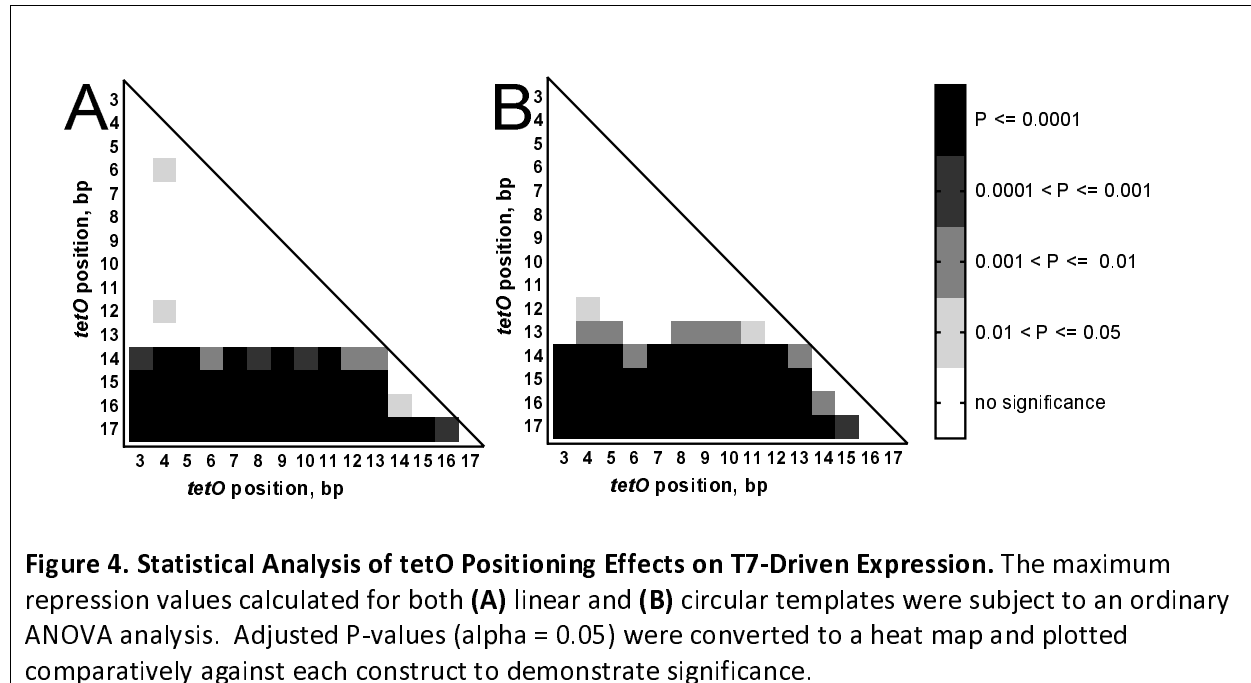
182 In order to facilitate direct comparisons between each *tetO* position, all mean expression values were  
183 divided by the greatest mean expression value for that position. This transformation rescaled the  
184 expression data to a maximum value of one for each position. Rescaled expression values were then  
185 plotted against the TetR concentration to generate dose-response profiles for each *tetO* position (Fig.  
186 1A, inset). Each dose-response profile was fit to a four parameter logistic regression curve (Fig. 1B) in  
187 order to obtain maximum repression values and 1/2 inhibitory concentration ( $IC_{50}$ ) values for TetR for  
188 each construct (Table S1).

189 A one-way ANOVA test comparing  $IC_{50}$  values at each position, revealed no significant difference ( $P \leq$   
190 0.0001), with the exception being that of tet.l.17, comparatively to other *tetO* positions (Fig. 2A).  
191 However, this may be explained by the difficulty in fitting the dose-response profile for tet.l.17, resulting  
192 in a large standard deviation (Fig. 2A). These results suggest that there is no effect of the *tetO* position  
193 on TetR binding, using linear template.



194

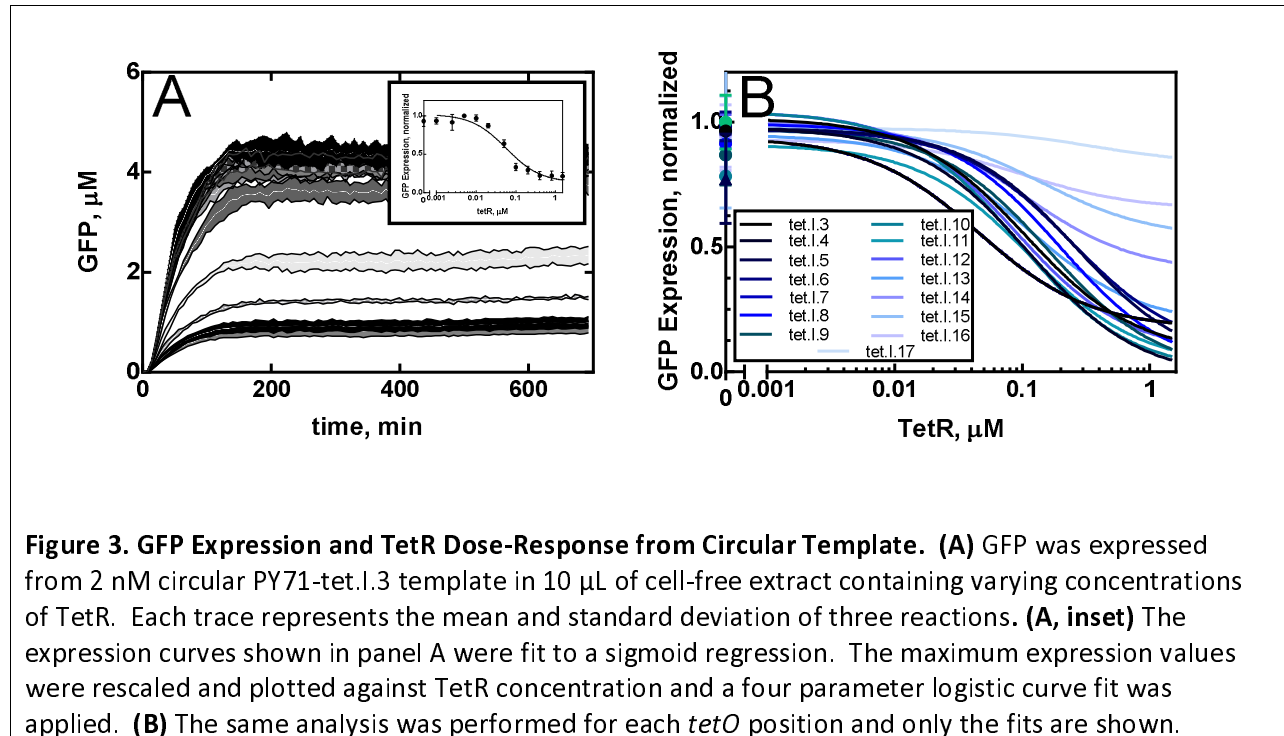
195 When a one-way ANOVA test to compare maximum repression values at each *tetO* position was applied,  
196 only tet.l.15 through tet.l.17 showed a significant difference ( $P \leq 0.0001$ ) on downregulation of T7  
197 RNAP-driven expression (tet.l.14 falls at a transitional position ( $P \leq 0.01$ )) (Fig. 2B). There is little, if any,  
198 observable effect of the *tetO* position on downregulation when the *tetO* sequence is less than 14 bp  
199 downstream from the T7 promoter. This phenomenon is best illustrated by the heat map in Figure 4A.  
200 Our results suggest one of two mechanisms for regulation may be at play: (1) that TetR blocks the  
201 binding of T7 RNAP, equally, up through position 13 or (2) that TetR prevents T7 polymerase from  
202 transitioning from initiation to elongation.



203

#### 204 **Template Format Impacts TetR Binding to the TetO Sequence but Not Downregulation**

205 To investigate the effects of template format (linear versus circular) on *tet* regulation of T7 RNAP-driven  
206 expression, each construct was seamlessly cloned into the pY71 expression vector, as described in  
207 Materials and Methods. Circular templates are differentiated from linear templates using the prefix  
208 pY71. Each circular template was evaluated for expression in cell-free extract, as described for linear  
209 templates. Different concentrations of TetR were added to reaction wells using an acoustic liquid  
210 handler. Cell-free reactions were run for 12 h, expression curves were fit using a sigmoidal regression,  
211 and the maximum sfGFP values were used for further evaluation (Fig. 3A)



212

213 We again observed variation in maximum sfGFP expression between templates, ranging from 0.5 µM to  
214 4.4 µM in the absence of TetR. Again, we tested two different preparations of each template with no  
215 TetR present (Fig. S4B). When a one-way ANOVA test was applied to the expression data from lot 1, no  
216 pattern that might indicate an effect of *tetO* position on expression became apparent (Fig. S4D). A two-  
217 way ANOVA test was applied to the expression data of both lots in order to determine statistical  
218 significances in the difference in expression at each *tetO* position (Fig. S4F). In contrast to linear  
219 template, a pattern of no significance running along the diagonal in Figure S4F was observed (with the  
220 exception of pY71-tet.l.17). However, at most positions there is no statistically significant difference in  
221 expression. This may be attributed to the large standard deviations for the expression means of lot 2.  
222 Therefore, it is difficult to conclude that the variation in expression is necessarily due to *tetO* position  
223 alone, in circular template.

224 Expression values for circular template were rescaled as described for linear template. This was also  
225 useful for the direct comparison of data from circular template with those from linear template.  
226 Rescaled expression values were plotted against TetR concentration to generate dose-response profiles  
227 for each *tetO* position. Each dose-response profile was fit to a four parameter logistic regression curve  
228 (Fig. 3B) in order to obtain maximum repression values and IC<sub>50</sub> values for TetR for each position (Table  
229 S2).

230 As with linear template, a one-way ANOVA test comparing IC<sub>50</sub> values for all positions, in circular  
231 templates, revealed no significant difference ( $P \leq 0.0001$ ), except for the IC<sub>50</sub> values for PY71-tet.l.5 and  
232 PY71-tet.l.6 comparatively to each other and the other *tetO* positions. Interestingly, a two-way ANOVA  
233 test, comparing IC<sub>50</sub> values from circular and linear template, revealed a statistically significant  
234 difference. Further, with the exception of tet.l.15 and tet.l.17, a t-test of individual IC<sub>50</sub> values between  
235 the two template formats, at each *tetO* position, shows a statistically significant difference ( $P \leq 0.05$ ).  
236 This suggests that, while *tetO* position does not influence TetR binding, template format may affect TetR  
237 binding. This is illustrated, qualitatively, by Figure 2A, which shows the relatively small variation in IC<sub>50</sub>  
238 values along *tetO* position within template format compared with the relatively greater variation in IC<sub>50</sub>  
239 values observed between template format.

240 Even though TetR displays an unusually low background affinity for DNA ( $\sim 10^5 \text{ M}^{-1}$ ) [28], we reasoned  
241 that some non-specific DNA binding might well account for the increased IC<sub>50</sub> values in circular template,  
242 increasing the effective concentration required for downregulation. In order to test this hypothesis, we  
243 added 2 nM PCR-amplified linear pY71 backbone to 2 nM linear tet.l.5 template in cell-free reactions,  
244 while varying TetR concentration (Fig. S5). Our results showed that the IC<sub>50</sub> value increased from  $0.010$   
245  $\pm 0.002 \mu\text{M}$  to  $0.029 \pm 0.002 \mu\text{M}$  upon the addition of pY71 backbone DNA, a statistically significant  
246 increase (Welch's t-test,  $p = 0.0006$ ). The IC<sub>50</sub> value for circular template in this experiment was  $0.073 \pm$   
247  $0.047 \mu\text{M}$ , which is not statistically different from the IC<sub>50</sub> value for linear template with linear pY71

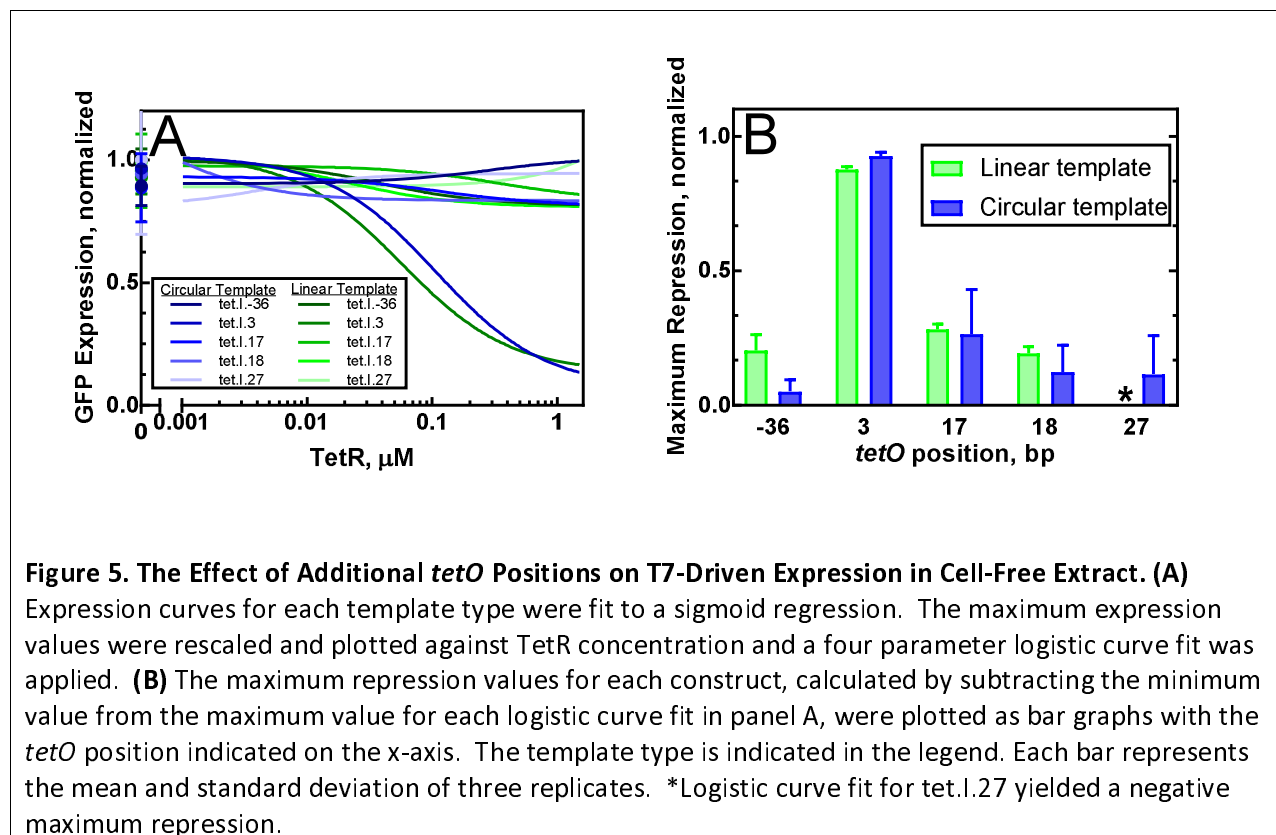
248 backbone DNA (Welch's t-test,  $p = 0.24$ ). These results suggest that non-specific binding of TetR to the  
249 vector backbone may indeed be responsible for the observed differences between linear and circular  
250 templates. However, a t-test, applied to the  $IC_{50}$  values of tet.l.5 and pY71-tet.l.5 also shows no  
251 statistically significant difference (Welch's t-test,  $p = 0.15$ ). Thus, we cannot rule out contribution from  
252 other factors.

253 A one-way ANOVA test comparing TetR downregulation in circular template yielded results similar to  
254 that for linear template. This, again, is best illustrated by the heat map in Figure 4B, showing no  
255 statistically significant ( $P \leq 0.0001$ ) reduction in repression when *tetO* is upstream of position 14. A  
256 qualitative comparison of the heat maps in Figure 4 and the traces in Figure 2B show remarkably similar  
257 trends for both template formats. This suggests that template format has little, if any, impact on *tet*  
258 regulation of T7 RNAP-driven expression, verifying our use of linear template to evaluate engineered T7-  
259 based transcription factors.

260 ***TetR Acts to Regulate T7-Driven Expression by Interfering with the Transition of T7 RNAP from***  
261 ***Initiation to Elongation***

262 In order to further investigate the *tet* regulatory mechanism for T7 RNAP, we placed the *tetO* sequence  
263 at positions 18, 27, and -36. For positions 18 and 27, we predicted that the trend of decreased  
264 regulation at positions further downstream from position 14 would continue. As expected, we observed  
265 nearly no downregulation at positions 18 and 27 (Fig. 5). Evaluation of position -36, which is  
266 immediately upstream of the T7 promoter sequence, was intended to differentiate the aforementioned  
267 competing hypotheses: (1) that the mechanism of regulation was either due to competitive binding  
268 between TetR and T7 RNAP near the promoter region, or (2) that TetR prevents T7 RNAP from  
269 transitioning from initiation to elongation. Interestingly, we found that, when the *tetO* sequence was  
270 placed immediately upstream from the T7 promoter sequence (tet.l.-36), TetR did not significantly

271 downregulate T7 RNAP-driven expression (Fig. 5), suggesting that *tet* regulates T7 RNAP-driven  
272 expression by preventing T7 RNAP from transitioning from initiation to elongation, rather than blocking  
273 T7 RNAP binding. Our observations are supported by the findings of Iyer et al. [4] and are consistent  
274 with the kinetic model published by Skinner et al. [10].  
275



276

## 277 Discussion

278 Due to its orthogonality to bacterial host machinery, T7 RNAP is a powerful tool for gene circuit design,  
279 and regulating its activity is central to fine-tuning gene circuit function. As such, understanding  
280 regulatory mechanisms for T7 RNAP are important to the design of engineered T7-based transcription  
281 factors that can be used in synthetic gene circuits.

282 Here we describe a rapid and cost-effective method to characterize promoter-operator combinations  
283 using cell-free protein synthesis and an acoustic liquid handler. Using this method, we investigated the  
284 effect of proximity of the *tetO* sequence to the T7 promoter on the regulation of T7 RNAP-driven  
285 expression. We observed that the absolute expression levels of sfGFP varied between templates, even  
286 when the same amount of template was added. Therefore, we prepared a second lot of each template,  
287 and measured sfGFP expression in cell lysate. While it may be tempting to conclude that the *tetO*  
288 position is accountable for the variation in expression based on a qualitative assessment from the  
289 patterns in the traces of Figures S4E and S4F, statistical analyses of the expression data (Fig. S4) indicate  
290 that the *tetO* position is not alone responsible for the variation in expression. There are numerous  
291 factors, beyond minor variations in template sequence, that may be responsible for the observed  
292 variation. One such culprit is template preparation. Templates were prepared by the mini prep  
293 (Qiagen) method and small variations in the amount of salts carried over during template preparation  
294 may account, in part, for the variation in expression. It is known that the salts magnesium and  
295 potassium, which is contained as 0.9 M potassium acetate in the neutralization buffer of the Qiagen  
296 miniprep kit, are among the most important parameters, along with template concentration, that affect  
297 the efficiency of cell-free protein synthesis [24,29]. Despite variation in expression, when the expression  
298 values were rescaled, consistent patterns emerged that are useful in promoter characterization.

299 Our data comparing the  $IC_{50}$  values for TetR in both template formats raises the question of whether  
300 format influences the  $IC_{50}$  value and, by extension, TetR binding to the *tetO* sequence (Fig. 2A and S3).  
301 While our experiment of adding linear backbone DNA to linear template appears to at least partially  
302 explain the differences in  $IC_{50}$  values, a second explanation may be in play as well. The crystal structure  
303 of the TetR homodimer reveals binding to the *tetO* sequence via N-terminal alpha helices, which occupy  
304 the major grooves of the operator, engaging with all but three base pairs of the sequence [30]. In its  
305 relaxed state, the periodicity of the DNA helix is 10.4 bp per turn [31]. However, under supercoiled

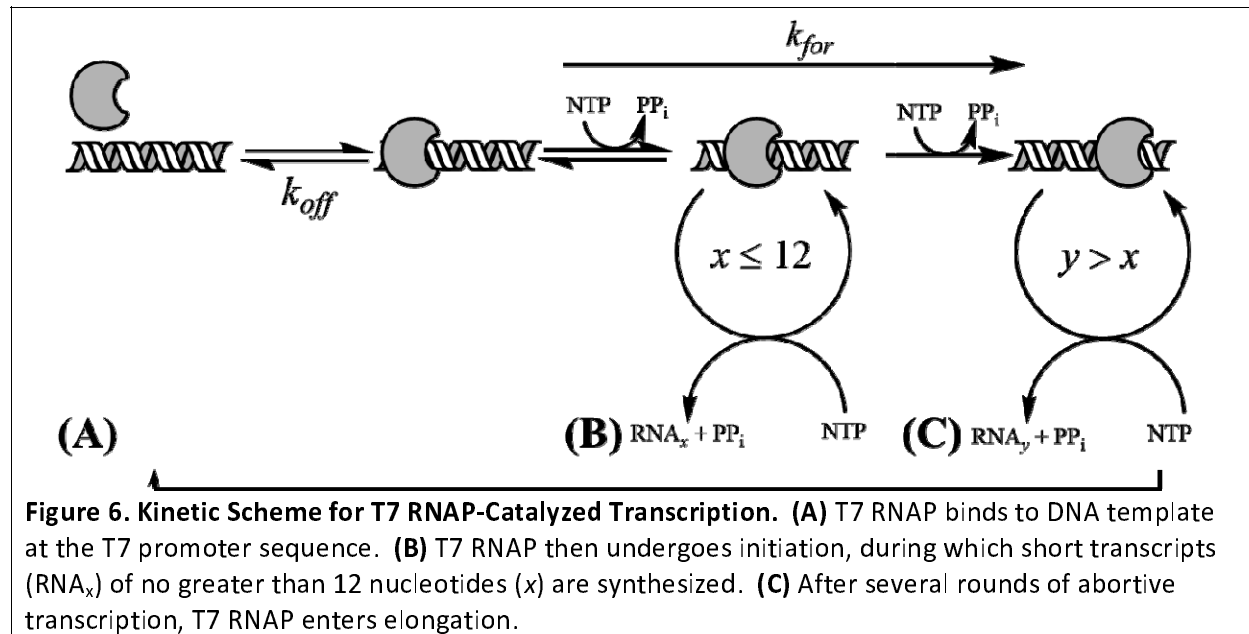


306 conditions, such as with circular DNA [32], the periodicity can vary between 10 to 11 bp turn [33], thus  
307 changing the width of the DNA grooves. It is plausible that the difference in  $IC_{50}$  values we observed  
308 between template formats can reasonably be attributed to the use of circular *versus* linear template.  
309 The contribution from experimental variation, however, makes it difficult to identify a single conclusive  
310 explanation.

311 In probing *tetO* position effects, we observed that T7 RNAP-driven expression is downregulated to the  
312 same degree when the *tetO* sequence is within 13 bp downstream from the T7 transcriptional start site  
313 (Fig. 3B) and that nearly no TetR downregulation is observed if the *tetO* sequenced is placed  
314 immediately upstream of the T7 promoter sequence (Fig. 5B). These results suggest that the *tet*  
315 regulatory mechanism for T7 RNAP operates by disrupting the transcriptional cycle, as described by  
316 Skinner et al. [10], at the initiation phase.

317 The transcriptional cycle, as described by Skinner et al. [10], proceeds through three phases: binding,  
318 initiation, and elongation (Fig. 6). During the binding phase, T7 RNAP recognizes the T7 promoter.  
319 Binding to the promoter sequence is close to the diffusion-controlled limit, indicating a relatively strong  
320 affinity of T7 RNAP for the T7 promoter [34] (Fig. 6A). Helix melting then occurs rapidly with the binding  
321 of the second ribonucleotide. Single molecule kinetic studies [10] on T7 RNAP revealed that, during  
322 initiation, T7 RNAP undergoes several rounds of abortive transcription across the first 12 bases of the  
323 template (Fig 6B), producing short RNA transcripts. Further, single molecule kinetics have shown that,  
324 during initiation, T7 RNAP favors dissociation ( $k_{off} = 2.9 \text{ s}^{-1}$ ) over transitioning to elongation ( $k_{for} = 0.36 \text{ s}^{-1}$ ) [10].  
325 During initiation, T7 RNAP accommodates only three base pairs of the DNA-RNA heteroduplex  
326 within the active site of the enzyme [35]. This explains the relatively weak affinity of the T7 RNAP for  
327 the DNA template throughout the 12 bases that constitute initiation. As T7 RNAP transitions from  
328 initiation to elongation (Fig. 6C), it undergoes a conformational change: the collapse of the promoter  
329 binding site and the formation of a channel, around the active site, that accommodates seven base pairs

330 of the DNA-RNA heteroduplex [9,36], as well as the formation of an N-terminal tunnel, allowing for the  
 331 egress of the nascent RNA transcript [9]. Throughout elongation, processivity is increased significantly,  
 332 indicating a relatively strong affinity of the T7 RNAP for the DNA template.



**Figure 6. Kinetic Scheme for T7 RNAP-Catalyzed Transcription.** (A) T7 RNAP binds to DNA template at the T7 promoter sequence. (B) T7 RNAP then undergoes initiation, during which short transcripts ( $RNA_x$ ) of no greater than 12 nucleotides ( $x$ ) are synthesized. (C) After several rounds of abortive transcription, T7 RNAP enters elongation.

333

334 Our results are in good agreement with the transcriptional model presented by Skinner et al. [10], and  
 335 illustrated in Figure 6. Consistent with the relatively weak affinity of T7 RNAP for the template during  
 336 initiation, we observed that downregulation of T7 RNAP-driven expression is the strongest and identical,  
 337 irrespective of the *tetO* position, as long as it is within the first 13 bp downstream from the T7 promoter  
 338 (Fig. 3B and Fig. 4). TetR became less effective as *tetO* moved further downstream, consistent with T7  
 339 RNAP entering elongation following base 12 of its transcript [10], and the enzyme's strong affinity for  
 340 the DNA template during this phase. Finally, TetR was also not effective at downregulating expression  
 341 when *tetO* was placed immediately upstream of the T7 promoter sequence (Fig. 5B), consistent with T7  
 342 RNAP's strong affinity for the T7 promoter.

343 Understanding the mechanism of repression for T7 RNAP using well characterized systems, such as the  
344 *tet* system, will allow for the design of more effective engineered T7-based transcription factors. Our  
345 results suggest that the design of new repressor-based, T7-based transcription factors would be best  
346 narrowed to the initiation phase of T7 RNAP. Indeed, since *tet* repression is one of the most effective in  
347 native promoter systems [23], our results suggest that placing operator sites outside of the 12 bp stretch  
348 consistent with initiation is likely futile unless additional mechanisms such as DNA looping [4] are  
349 employed. We also showed that the method developed here, utilizing cell-free protein synthesis and  
350 linear template, can be used to rapidly evaluate any such new engineered T7-based transcription  
351 factors. These results will assist in expanding the pallet of engineered T7-based transcription factors for  
352 the design of gene circuits.

### 353 **Bibliography**

- 354 [1] M. Chamberlin, J. McGrath, L. Waskell, New RNA Polymerase from Eschericia coli Infected with  
355 Bacteriophage T7, *Nature*. 228 (1970) 227–231.
- 356 [2] S. Tabor, Expression Using the T7 RNA Polymerase / Promoter System, in: *Curr. Protoc. Mol. Biol.*,  
357 1990: pp. 1–11. doi:10.1002/0471142727.mb1602s11.
- 358 [3] K. Temme, R. Hill, T.H. Segall-Shapiro, F. Moser, C.A. Voigt, Modular Control of Multiple Pathways  
359 Using Engineered Orthogonal T7 Polymerases, *Nucleic Acids Res.* 40 (2012) 8773–8781.  
360 doi:10.1093/nar/gks597.
- 361 [4] S. Iyer, D.K. Karig, S.E. Norred, M.L. Simpson, M.J. Doktycz, Multi-Input Regulation and Logic with  
362 T7 Promoters in Cells and Cell-Free Systems, *PLoS One*. 8 (2013) 1–12.  
363 doi:10.1371/journal.pone.0078442.
- 364 [5] V.L. Tunitskaya, S.N. Kochetkov, Structural-Functional Analysis of Bacteriophage T7 RNA  
365 Polymerase, *Biochem.* 67 (2002) 1124–1135. doi:10.1023/A:1020911223250.
- 366 [6] J.F. Klement, M.B. Moorefield, E. Jorgensen, J.E. Brown, S. Risman, W.T. McAllister,  
367 Discrimination between Bacteriophage T3 and T7 Promoters by the T3 and T7 RNA Polymerases  
368 Depends Primarily upon a Three Base-Pair Region Located 10 to 12 Base-Pairs Upstream from the  
369 Start Site, *J. Mol. Biol.* 215 (1990) 21–29. doi:10.1016/S0022-2836(05)80091-9.
- 370 [7] C.T. Martin, D.K. Muller, J.E. Coleman, Processivity in Early Stages of Transcription by T7 RNA  
371 Polymerase, *Biochemistry.* 27 (1988) 3966–3974. doi:10.1021/bi00411a012.
- 372 [8] C. Liu, C.T. Martin, Promoter Clearance by T7 RNA Polymerase, *J. Biol. Chem.* 277 (2002) 2725–  
373 2731. doi:10.1074/jbc.M108856200.
- 374 [9] Y.W. Yin, T.A. Steitz, Structural Basis for the Transition from Initiation to Elongation Transcription

- 375 in T7 RNA Polymerase, *Science* (80-. ). 298 (2002) 1387–1396. doi:10.1126/science.1077464.
- 376 [10] G.M. Skinner, C.G. Baumann, D.M. Quinn, J.E. Molloy, J.G. Hoggett, Promoter Binding, Initiation,  
377 and Elongation by Bacteriophage T7 RNA Polymerase, *J. Biol. Chem.* 279 (2004) 3239–3244.  
378 doi:10.1074/jbc.M310471200.
- 379 [11] N. Arnaud-barbe, V. Cheynet-sauvion, G. Oriol, B. Mandrand, Transcription of RNA Templates by  
380 T7 RNA Polymerase, *Nucleic Acids Res.* 26 (1998) 3550–3554.
- 381 [12] M. Kushwaha, H.M. Salis, A Portable Expression Resource for Engineering Cross-Species Genetic  
382 Circuits and Pathways, *Nat. Commun.* 6 (2015) 7832. doi:10.1038/ncomms8832.
- 383 [13] T.H. Segall-Shapiro, A.J. Meyer, A.D. Ellington, E.D. Sontag, C.A. Voigt, A “Resource Allocator” for  
384 Transcription Based on a Highly Fragmented T7 RNA Polymerase, *Mol. Syst. Biol.* 10 (2014) 1–15.  
385 doi:10.15252/msb.20145299.
- 386 [14] D.L. Shis, M.R. Bennett, Synthetic biology: The Many Facets of T7 RNA Polymerase., *Mol. Syst.*  
387 *Biol.* 10 (2014) 745. doi:10.15252/msb.20145492.
- 388 [15] A. Ishihama, Functional Modulation of Escherichia coli RNA Polymerase, *Annu. Rev. Microbiol.* 54  
389 (2000) 499–518.
- 390 [16] A.J. Meyer, T. Segall, C.A. Voigt, Marionette<sup>®</sup>: E. coli containing 12 highly-optimized small  
391 molecule sensors, *Bioarchives.* (2018) 1–27.
- 392 [17] W. Saenger, P. Orth, C. Kisker, W. Hillen, W. Hinrichs, The Tetracycline Repressor - A Paradigm for  
393 a Biological Switch, *Angew. Chemie - Int. Ed.* 39 (2000) 2042–2052.
- 394 [18] R. Bertram, W. Hillen, L.M. Genetik, E. Karls, The Application of Tet Repressor in Prokaryotic Gene  
395 Regulation and Expression, *Microb. Biotechnol.* 1 (2008) 2–16. doi:10.1111/j.1751-  
396 7915.2007.00001.x.
- 397 [19] D.K. Karig, S. Iyer, M.L. Simpson, M.J. Doktycz, Expression Optimization and Synthetic Gene  
398 Networks in Cell-Free Systems, *Nucleic Acids Res.* 40 (2012) 3763–3774.  
399 doi:10.1093/nar/gkr1191.
- 400 [20] J. Hasty, D. McMillen, J.J. Collins, Engineered Gene Circuits, *Nature.* 420 (2002) 224.  
401 <https://doi.org/10.1038/nature01257>.
- 402 [21] K. Brenner, D.K. Karig, R. Weiss, F.H. Arnold, Engineered Bidirectional Communication Mediates a  
403 Consensus in a Microbial Biofilm Consortium., *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 17300–  
404 17304. doi:10.1073/pnas.0704256104.
- 405 [22] A. Becskei, L. Serrano, Engineering Stability in Gene Networks by Autoregulation, *Nature.* 405  
406 (2000) 590. <https://doi.org/10.1038/35014651>.
- 407 [23] K.P. Bertrand, R.E. Lenski, Effects of Carriage and Expression of the Tn10 Tetracycline-Resistance  
408 Operon on the Fitness of Escherichia coli K12, *Mol. Biol. Evol.* 6 (1989) 213–225.
- 409 [24] Z.Z. Sun, C.A. Hayes, J. Shin, F. Caschera, R.M. Murray, V. Noireaux, Protocols for Implementing  
410 an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology, *J. Vis. Exp.* 79  
411 (2013) 1–14. doi:10.3791/50762.
- 412 [25] Z.Z. Sun, E. Yeung, C.A. Hayes, V. Noireaux, R.M. Murray, Linear DNA for Rapid Prototyping of

- 413 Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System, ACS Synth. Biol. 3  
414 (2014) 387–397. doi:10.1021/sb400131a.
- 415 [26] J.R. Swartz, M.C. Jewett, K.A. Woodrow, Cell-Free Protein Synthesis With Prokaryotic Combined  
416 Transcription-Translation, in: Methods Mol. Biol., 2004: pp. 169–182.
- 417 [27] S. Grkovic, M.H. Brown, R.A. Skurray, Regulation of Bacterial Drug Export Systems, Microbiol.  
418 Mol. Biol. Rev. 66 (2002) 671–701. doi:10.1128/MMBR.66.4.671.
- 419 [28] C. Berens, W. Hillen, Gene Regulation by Tetracyclines, Eur. J. Biochem. 270 (2003) 3109–3121.  
420 doi:10.1046/j.1432-1033.2003.03694.x.
- 421 [29] J. Shin, V. Noireaux, Efficient Cell-Free Expression with the Endogenous E. Coli RNA Polymerase  
422 and Sigma Factor 70, J. Biol. Eng. 4 (2010) 1–9. doi:10.1186/1754-1611-4-8.
- 423 [30] P. Orth, D. Schnappinger, W. Hillen, W. Saenger, W. Hinrichs, Structural Basis of Gene Regulation  
424 by the Tetracycline Inducible Tet Repressor – Operator System, Nat. Struct. Biol. 7 (2000) 215–  
425 219.
- 426 [31] J.C. Wang, Helical Repeat of DNA in Solution Biochemistry, Proc. Natl. Acad. Sci. 76 (1979) 200–  
427 203.
- 428 [32] N.P. Higgins, A. V Vologodskii, Topological Behavior of Plasmid DNA, Microbiol. Spectr. 3 (2016)  
429 1–49. doi:10.1128/microbiolspec.PLAS-0036-2014.Topological.
- 430 [33] P. Schieg, H. Herzel, Periodicities of 10 – 11 bp as Indicators of the Supercoiled State of Genomic  
431 DNA, J. Mol. Biol. 343 (2004) 891–901. doi:10.1016/j.jmb.2004.08.068.
- 432 [34] A. Újvári, C.T. Martin, Thermodynamic and Kinetic Measurements of Promoter Binding by T7 RNA  
433 Polymerase, Biochemistry. 35 (1996) 14574–14582. doi:10.1021/bi961165g.
- 434 [35] G.M.T. Cheetham, T.A. Steitz, Structure of a Transcribing T7 RNA Polymerase Initiation Complex,  
435 Science (80-. ). 286 (1999) 2305–2309. doi:10.1126/science.286.5448.2305.
- 436 [36] J. Huang, R. Sousa, T7 RNA Polymerase Elongation Complex Structure and Movement, J. Mol.  
437 Biol. 303 (2000) 347–358. doi:10.1006/jmbi.2000.4150.
- 438
- 439