



# A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase

Thomas H Segall-Shapiro<sup>1</sup>, Adam J Meyer<sup>2</sup>, Andrew D Ellington<sup>2</sup>, Eduardo D Sontag<sup>3</sup> & Christopher A Voigt<sup>1,\*</sup>

## Abstract

Synthetic genetic systems share resources with the host, including machinery for transcription and translation. Phage RNA polymerases (RNAPs) decouple transcription from the host and generate high expression. However, they can exhibit toxicity and lack accessory proteins ( $\sigma$  factors and activators) that enable switching between different promoters and modulation of activity. Here, we show that T7 RNAP (883 amino acids) can be divided into four fragments that have to be co-expressed to function. The DNA-binding loop is encoded in a C-terminal 285-aa ' $\sigma$  fragment', and fragments with different specificity can direct the remaining 601-aa 'core fragment' to different promoters. Using these parts, we have built a resource allocator that sets the core fragment concentration, which is then shared by multiple  $\sigma$  fragments. Adjusting the concentration of the core fragment sets the maximum transcriptional capacity available to a synthetic system. Further, positive and negative regulation is implemented using a 67-aa N-terminal ' $\alpha$  fragment' and a null (inactivated)  $\sigma$  fragment, respectively. The  $\alpha$  fragment can be fused to recombinant proteins to make promoters responsive to their levels. These parts provide a toolbox to allocate transcriptional resources via different schemes, which we demonstrate by building a system which adjusts promoter activity to compensate for the difference in copy number of two plasmids.

**Keywords** genetic circuit; resource allocation; split protein; synthetic biology; T7 RNA polymerase

**Subject Categories** Synthetic Biology & Biotechnology; Methods & Resources

**DOI** 10.15252/msb.20145299 | Received 21 March 2014 | Revised 5 June 2014 | Accepted 24 June 2014

**Mol Syst Biol.** (2014) **10**: 742

See also: **DL Shis & MR Bennett** (July 2014)

## Introduction

Cells must control the production of RNA polymerase (RNAP) and ribosomes to balance their biosynthetic cost with the needs of cell

growth and maintenance (Warner, 1999). As such, RNAP and ribosome synthesis is under stringent regulatory control, both to coordinate their levels with respect to cellular and environmental cues for growth (Nierlich, 1968; Hayward *et al*, 1973; Iwakura & Ishihama, 1975; Bedwell & Nomura, 1986; Bremer & Dennis, 2008; Schaechter *et al*, 1958; Lempiäinen & Shore, 2009; Gausing, 1977; Schneider *et al*, 2003) and to balance the expression of their components for proper assembly into functional machines (Warner, 1999; Ishihama, 1981; Nierhaus, 1991; Fatica & Tollervey, 2002). This sets a resource budget that must be shared in the transcription of approximately 4,000 genes and translation of  $\sim 10^6$  nucleotides of mRNA in *E. coli* (Bremer & Dennis, 1996). The budget is not large; on average, there are 2,000 RNAP and 10,000 ribosomes per cell (Ishihama *et al*, 1976; Bremer & Dennis, 1996; Ishihama, 2000). Mathematical models often assume these budgets to be constant (Shea & Ackers, 1985; Gardner *et al*, 2000; Elowitz & Leibler, 2000), but the numbers can vary significantly in different growth phases and nutrient conditions, ranging from 1,500 to 11,400 RNAPs and 6,800 to 72,000 ribosomes per cell (Bremer & Dennis, 1996; Klumpp & Hwa, 2008). The fluctuations in resources can lead to global changes in expression levels and promoter activities (Keren *et al*, 2013; De Vos *et al*, 2011).

This poses a problem when a synthetic genetic system is introduced. When it relies on the transcription and translation machinery of the host, it becomes implicitly embedded in their regulation, making it sensitive to changes that occur during cell growth and function. As a result, the system can be fragile because the strengths of its component parts (promoters and ribosome binding sites) will vary with the resource budgets (Moser *et al*, 2012; Arkin & Fletcher, 2006; Kittleson *et al*, 2012). For example, changes in the RNAP concentration can impact the expression from constitutive promoters by fivefold (Bremer & Dennis, 1996; Liang *et al*, 1999; Klumpp *et al*, 2009; Liang *et al*, 2000; Klumpp & Hwa, 2008). These changes can reduce the performance of a system that requires precise balances in expression levels (Temme *et al*, 2012b; Moser *et al*, 2012; Moon *et al*, 2012). This has emerged as a particular problem in obtaining reliable expression levels and gene circuit performance during industrial scale-up, where each phase is associated with different growth and media conditions (Moser *et al*, 2012).

<sup>1</sup> Department of Biological Engineering, Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup> Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, USA

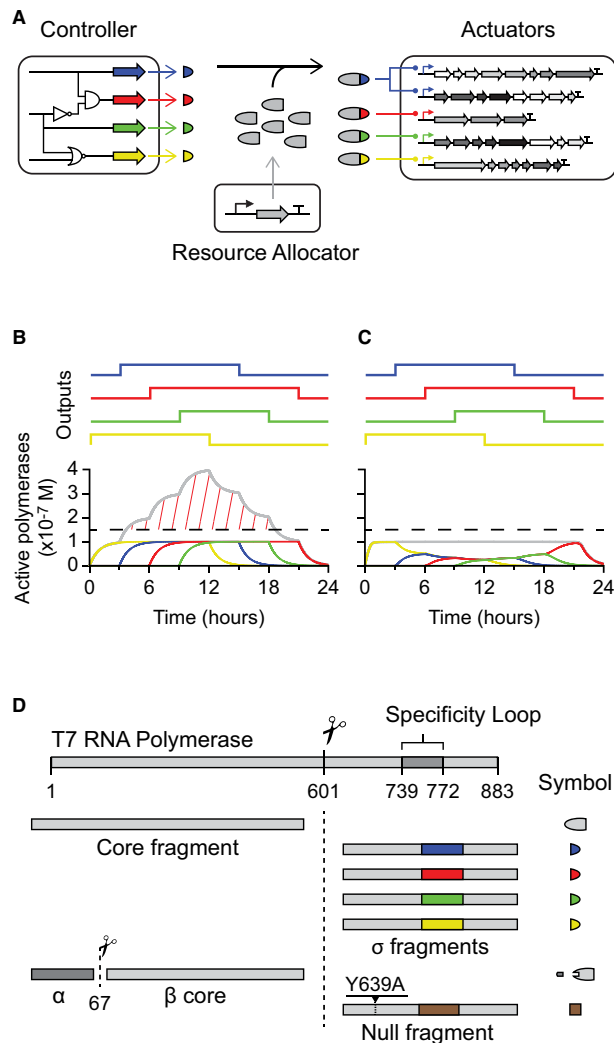
<sup>3</sup> Department of Mathematics, Rutgers University, Piscataway, NJ, USA

\*Corresponding author. Tel: +1 617 324 4851; E-mail: cvoigt@gmail.com

Another problem is that synthetic systems often place high demands on host transcription and translation resources and this can have global consequences in maintaining growth and responding to stress (Hoffmann & Rinas, 2004; Birnbaum & Bailey, 1991). Proteins and pathways expressed at very high levels place a burden on cells that can reach up to 30% of total cellular proteins and utilize 50% of translation capacity (Dong *et al*, 1995; Scott *et al*, 2010; Carrera *et al*, 2011). The competition with native genes can cause a decrease in their expression and a reduction or cessation of growth (Dong *et al*, 1995; Scott *et al*, 2010; Carrera *et al*, 2011; Tabor *et al*, 2008). In addition, because of the small numbers of RNAP and ribosomes, the expression of recombinant genes can become coupled, where a high level of expression of one gene titrates a resource and reduces the expression of another gene. In the context of synthetic signaling networks, this has been referred to as 'retroactivity', where downstream targets can impart a load on the upstream signaling pathway (Jiang *et al*, 2011; Jayanthi *et al*, 2013; Del Vecchio *et al*, 2008; Del Vecchio & Murray, 2014).

These challenges were recognized early in biotechnology and a partial solution emerged by using the RNAP from T7 phage to decouple transcription from the host machinery (Chamberlin *et al*, 1970; Studier & Moffatt, 1986; Alexander *et al*, 1992). Heterologous T7 RNAP was patented in 1984 (Studier *et al*, 1990) and since then has been the basis for expression systems across many organisms (Elroy-Stein & Moss, 1990; Brunschwig & Darzins, 1992; McBride *et al*, 1994; Conrad *et al*, 1996). An advantage cited for this system was that it could achieve high expression levels by adding an inhibitor of *E. coli* RNAP, thus directing metabolic resources to recombinant protein production (Tabor & Richardson, 1985). However, there are also some challenges with using T7 RNAP. While the polymerase itself is not toxic, when it is combined with a strong promoter, it can cause severe growth defects. The origin of this toxicity is not clear, but it could be related to the rate of transcription of T7 RNAP, which is eightfold faster than *E. coli* RNAP and could expose naked mRNA (Iost *et al*, 1992; Miroux & Walker, 1996). Toxicity can be ameliorated by introducing a mutation near the active site and by selecting parts to lower polymerase expression (Temme *et al*, 2012a,b). Beyond the RNAP from T7, many polymerases have been identified from different phage and directed evolution experiments have yielded variants that recognize different promoter sequences (Temme *et al*, 2012a; Ellefson *et al*, 2013; Carlson *et al*, 2014).

Phage polymerases are central to our organization of larger genetic systems (Temme *et al*, 2012a,b; Smanski *et al*, 2014). We separate the regulation of a system (on a plasmid we refer to as the 'controller') from those genes encoding pathways or cellular functions ('actuators') (Fig 1A). The controller contains synthetic sensors and circuits, whose outputs are phage polymerases specific to the activation of the actuators. This organization has several practical advantages. First, it avoids evolutionary pressure when manipulating the actuators because the promoters are tightly off in the absence of phage polymerase. Thus, they can be carried in an inactive state until the controller is introduced into the cell. Actuators often require many genes and assembled parts, making re-verification of their sequence expensive. Second, it allows the regulation of the actuators to be changed quickly. Controllers can be swapped to change the conditions and dynamics of expression, so long as they produce the same dynamic range in output polymerase expression. In the same way, the controllers can also be



**Figure 1. The resource allocator.**

- A Complex synthetic genetic systems are broken down into three modules. The core fragment of RNAP is expressed from the resource allocator. Each output from the controller results in the expression of a different  $\sigma$  fragment (colored half-circles), which share the core fragment and turn on different actuators.
- B Dynamic simulations of resource allocation are shown, where the outputs from the controller are turned on and off at different times (colored lines) (Supplementary information Section IV.A). A hypothetical toxicity threshold is shown with the dashed horizontal line. When the outputs of the controller are complete RNAPs, their sum crosses the threshold (gray line and red hash).
- C With resource allocation, the outputs of the controller are  $\sigma$  fragments that must share the core fragment, thus ensuring that their sum transcriptional activity does not cross the threshold.
- D The complete toolbox of phage RNAP fragments is shown.

characterized independently using surrogate fluorescent reporters prior to being combined with the actuators.

With these large and complex synthetic systems, problems can arise as the host is subjected to significant perturbation and load. Simultaneously activating a number of actuators requires expressing multiple polymerases that might collectively cross the threshold for toxicity (Fig 1B). While lowering expression rates throughout the

system could avoid toxicity, it would needlessly constrain expression when only one actuator is active. To address this issue, we aimed to create an allocation system that allows independently setting the total desired polymerase activity and allocating this resource to the various actuators as needed. With this organization, a single actuator can be expressed to full strength, but expression of multiple actuators is attenuated to avoid overexpression (Fig 1C). In effect, we are proposing to add another layer to the organization of genetic designs, where a separate 'resource allocator' is responsible for the maintenance of a desired level of orthogonal transcriptional machinery (Fig 1A).

Prokaryotes solve the problem of partitioning a budget of RNAP to different cellular processes through the action of  $\sigma$  factors, which bind to the core RNAP ( $\alpha 2$ ,  $\beta$ ,  $\beta'$ , and  $\omega$  subunits) and direct it to promoter sequences (Gruber & Gross, 2003; El-Samad *et al*, 2005). Core RNAP itself only has the ability to non-specifically bind to DNA, whereas the  $\sigma$  factor contains the DNA recognition domains for the  $-35$  and  $-10$  regions of promoters. Different  $\sigma$  factors bind to distinct promoter recognition sequences. In *E. coli*, there is one 'housekeeping'  $\sigma$  factor ( $\sigma^{70}$ ) that is expressed at a constant level of 500–700 molecules/cell, independent of growth phase or stress, and 6 alternate  $\sigma$  factors that control various stress responses (e.g., heat shock) and cellular functions (e.g., flagella assembly) (Jishage *et al*, 1996).  $\sigma$  factors can range in size;  $\sigma^{70}$  is 613 amino acids and the average alternative  $\sigma$  is  $\sim 200$  amino acids (Burton *et al*, 1981; Staroń *et al*, 2009; Rhodius *et al*, 2013). These alternative  $\sigma$ s can be embedded in complex regulatory networks that implement signal integration and feedback regulation that mimics engineering control architectures (Lange & Hengge-Aronis, 1994; Hengge-Aronis, 2002; Kurata *et al*, 2001). In this way, the level of core RNAP dictates the total transcriptional potential in the cell, while the relative levels of  $\sigma$  factors determine how this resource is allocated between growth and stress resistance (Nyström, 2004; Maharjan *et al*, 2013). Bacteria with more diverse lifestyles can have significantly more  $\sigma$  factors, for example, *Streptomyces* and *Bacteroides* species can have greater than 50 (Lange & Hengge-Aronis, 1994; Hengge-Aronis, 2002; Kurata *et al*, 2001). All of these  $\sigma$ s compete to bind to the core RNAP (Ishihama, 2000; Gruber & Gross, 2003).

In this manuscript, we have created an analogous system by fragmenting T7 RNAP. We used a transposon method to identify five regions where the polymerase can be bisected and retain function. One of these splits produces a 285 amino acid fragment that we refer to as the ' $\sigma$  fragment' because it contains the region that binds to the promoter (Fig 1D). We find that variants of this fragment with different promoter specificities can bind to the remaining 'core fragment' and direct it to different promoters. The expression level of the core fragment dictates the maximum number of active polymerases. The outputs of the controller are different  $\sigma$  fragments, which are used to turn on different actuators. If the pool of core fragments is saturated by  $\sigma$  fragments, the total number of active polymerases in the system will remain constant regardless of the levels of  $\sigma$  fragments being produced (Fig 1C). In this way, a desired transcriptional load can be specified and then dynamically allocated to different actuators as the conditions require. Negative regulators can be built by creating null  $\sigma$  fragments that titrate the core fragment but do not support transcription. Additionally, the core fragment can be positively regulated using the N-terminal bisection point to separate an ' $\alpha$  fragment' that is required for activity. These regulators

could be used to implement feedback loops that control the amount of active RNAP complexes under different conditions or the dynamics of signal progression from the controller to the actuators.

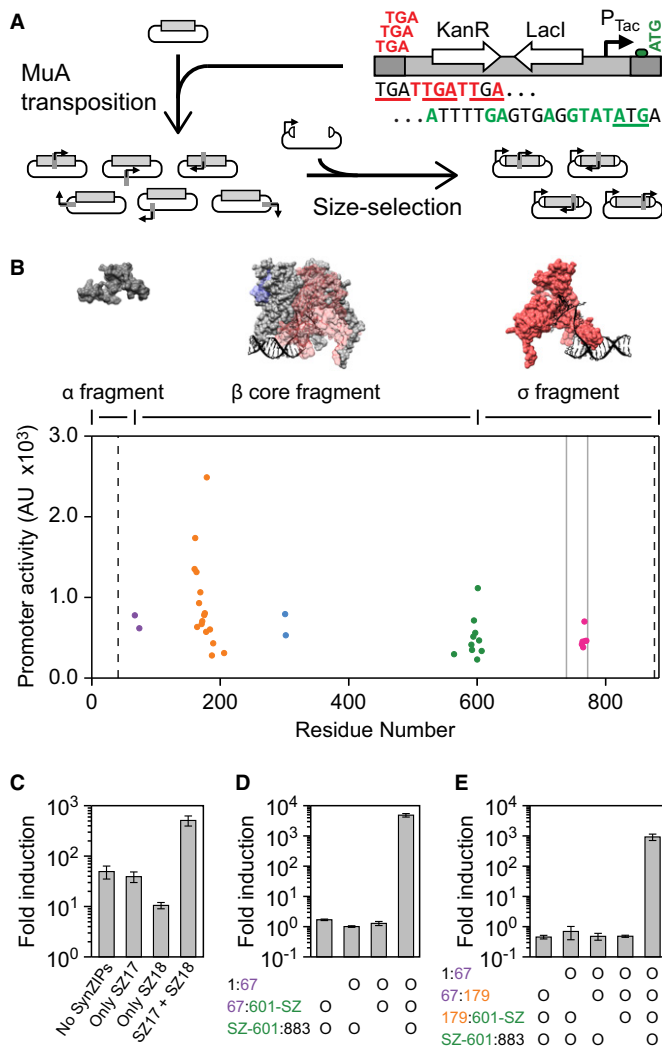
## Results

### Bisection mapping of T7 RNA polymerase

Our first objective was to identify all of the places T7 RNAP could be split to yield two fragments that can be co-expressed to produce a functional protein. To do this, we developed a transposase-based method that uses a novel transposon to split proteins, which we refer to as a 'splitposon'. Previous methods have been published to generate libraries of split proteins or domain insertions that are based on incremental truncation (Ostermeier *et al*, 1999; Paschon & Ostermeier, 2004), multiplex inverse PCR (Kanwar *et al*, 2013), DNase cleavage (Guntas & Ostermeier, 2004; Chen *et al*, 2009), and transposon insertion (Segall-Shapiro *et al*, 2011; Mahdavi *et al*, 2013). The transposon-based approaches are able to generate large libraries and do not require sensitive DNase steps, but they leave  $\sim 10$  added amino acids at the split site. To improve on this approach, the splitposon is a Mu transposon in which one terminal transposon recognition end is altered to contain a non-disruptive ribosome binding site (RBS) and start codon (Fig 2A). We further modified the transposon to add the remaining necessary regulation to divide a protein into two fragments (stop codon— $P_{Tac}$  IPTG-inducible system—RBS—start codon). The MuA transposase efficiently yields random insertions of the splitposon throughout a DNA molecule, producing a library of split proteins flanked by just three additional amino acids for in-frame insertions (Supplementary Fig S1).

With the splitposon, a bisection library for any protein can be generated in two steps (Fig 2A). First, the splitposon is transposed *in vitro* into a plasmid containing the DNA within which bisections are desired (e.g., a gene or segment of a gene). Second, the target region is digested from the plasmid backbone and size selected for fragments containing an inserted transposon. These fragments are ligated into an expression plasmid containing an upstream inducible promoter. The final library will contain only plasmids with a single transposon insertion in the region of interest and can be induced and screened for function.

The splitposon method was applied to generate a library of bisections of a variant of T7 RNAP (T7\* RNAP). This gene contains the R632S mutant, which reduces host toxicity (Temme *et al*, 2012a). To avoid trivial truncations of the termini, we directed transposon insertions to the region of the gene corresponding to amino acids 41 through 876 of the polymerase. Both fragments are induced with IPTG from  $P_{Tac}$ . The library was co-transformed with a screening plasmid that contains a T7 RNAP dependent promoter and red fluorescent protein (RFP) (Temme *et al*, 2012a), and 384 clones were picked by eye from agar plates, re-assayed in liquid media, and the best 192 sequenced. From these, 36 unique in-frame split sites were identified (Fig 2B). The split sites cluster into five distinct seams that correspond to six potential fragments if they were all implemented simultaneously. The seam around position 179 corresponds to a previously identified split site that yields a functional T7 RNAP (Ikeda & Richardson, 1987a,b; Muller *et al*, 1988; Shis & Bennett, 2013).



**Figure 2. Bisection mapping of T7\* RNAP.**

- A** The splitposon is based on a modified mini-Mu transposon mutated to contain staggered stop codons in one recognition end (red) and an RBS & start codon in the other (green). An internal inducible system (LacI and  $P_{Tac}$ ) has been added. Bisection mapping includes two cloning steps. First, the splitposon is transposed randomly into a gene using MuA transposase. Second, the library is size selected for inserts that contain one transposon insertion and cloned into an expression plasmid.
- B** Each point represents a unique in-frame split location in T7\* RNAP, where the residue number is the final residue in the N-terminal fragment. The promoter activity is the mean  $P_{T7}$  activity for all recovered clones at each split point, from four independent assays (10  $\mu$ M IPTG induction). Bisection points are clustered into five 'seams', which are color-coded. The vertical dashed lines show the region where bisections were allowed in the library, and the gray vertical lines show the location of the promoter specificity loop. Surface models are shown for the three fragments used for the resource allocator (PDB:1QLN (Cheetham & Steitz, 1999), visualized using UCSF Chimera (Pettersen *et al*, 2004)). The model for the  $\beta$  core fragment shows the position of the  $\alpha$  and  $\sigma$  fragments in transparent blue and red, respectively. More views of the surface model are shown in Supplementary Fig S4.
- C** The fragments created from splitting T7 RNAP at residue 601 were assayed with and without SynZIP domains at low expression levels (4  $\mu$ M IPTG). When SynZIP 17 (SZ17) is fused to the N-terminal fragment and SynZIP 18 (SZ18) is fused to the C-terminal fragment, a large increase in the induction of  $P_{T7}$  is observed. Fold induction is calculated as the  $P_{T7}$  promoter activity in induced cells divided by the promoter activity of cells that contain the reporter plasmid but no polymerase fragments.
- D** Data are shown for the expression of the three fragments corresponding to the  $\alpha$  fragment (1:67),  $\beta$  core fragment (67:601-SZ), and  $\sigma$  fragment (SZ-601:883). An 'o' indicates the presence of a fragment in an operon that is expressed with 100  $\mu$ M IPTG.
- E** Data are shown for the induction of four fragments, as in (D), with an additional split of the  $\beta$  core fragment at residue 179.

Data information: For the graphs in (C–E), the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

Source data are available online for this figure.

### Division of T7 RNAP into multiple fragments

All of the discovered split seams occur in surface-exposed regions of the T7\* RNAP, and the largest seam corresponds to a large surface-exposed loop known as the 'Flap' in the 3-dimensional structure (Supplementary Fig S3) (Tahirov *et al*, 2002). This implies that additional functional domains can be inserted at these positions. We hypothesized that the addition of protein–protein interaction domains could improve the affinity of the fragments. To this end, two leucine zipper domains that bind in an antiparallel orientation were chosen from the SynZIP toolbox (variants 17 and 18) (Reinke *et al*, 2010; Thompson *et al*, 2012). Addition of either SynZIP at the 601 split site with a short flexible linker is tolerated by the split polymerase, and adding both is beneficial and improves activity by greater than tenfold at low expression levels (Fig 2C).

The outcome of the bisection mapping experiment also implied that it might be possible to divide T7\* RNAP into more than two fragments. First, the protein was divided into three fragments based on the split points at residues 67 and 601, including the added SynZIPs at the 601 split. These three fragments were expressed as a single inducible operon and compared to versions lacking each of

the single fragments. RNAP activity (4,000-fold induction) is only detected when all three fragments are expressed and there is no activity in the absence of any fragment (Fig 2D). We also tested a four fragment version, which includes a split at position 179 (Fig 2E). The expression of these four fragments yields active RNAP (900-fold induction), and there is no detectable activity if any of the fragments are not expressed.

While the four and three-piece polymerases do lead to a reduction in cell growth when expressed at high levels, this effect is more pronounced when expressing the full-length protein (Supplementary Fig S12). Splitting the polymerase into five or six fragments was not attempted due to the attenuation of activity and growth impact of high expression with four fragments.

### Construction of ' $\sigma$ fragments' with different promoter specificities

The C-terminal fragment generated by the split site at residue 601 (601–883) contains the DNA-binding loop that determines promoter specificity (Cheetham *et al*, 1999). Thus, we refer to this as the ' $\sigma$  fragment' as it functions analogously to  $\sigma$  factors that bind to *E. coli* RNAP and is approximately the same size. Following this analogy, the 601 amino acid N-terminal fragment is referred to as the 'core fragment'. Note that this fragment is much smaller than the  $\alpha$ 2/ $\beta$ /

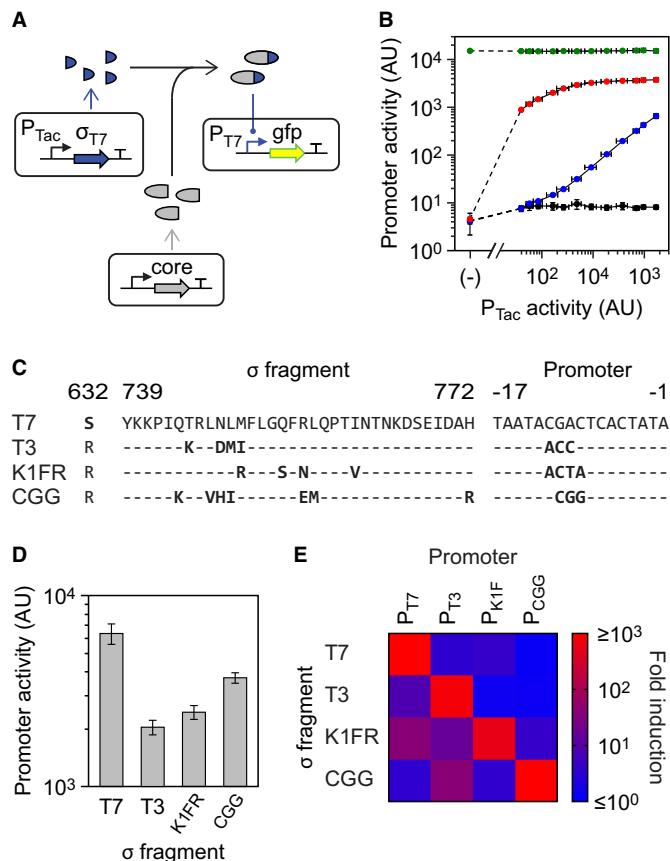
$\beta'/\omega$  subunits of *E. coli* RNAP (329/1342/1407/91 amino acids) and they assemble into a very different 3-dimensional structure (Sousa et al, 1993; Vassilyev et al, 2002; Opalka et al, 2010).

A simple resource allocator was built based on the core and  $\sigma$  fragments (Fig 3A), retaining the amino acids added by the splitposon method and the SynZIP 18 domain on the  $\sigma$  fragment. The core fragment is expressed from the constitutive promoter  $P_{J23105}$ , tuned to a low level such that expressing full-length polymerase in its place is not toxic. The  $\sigma$  fragment is expressed at varying levels using an IPTG-inducible  $P_{Tac}$  promoter. Polymerase activity is measured using  $P_{T7}$  driving green fluorescent protein (GFP) (Materials and Methods). The  $\sigma$  fragment, core fragment, and reporter are carried on three separate plasmids (p15A\*, BAC, pSC101) to mimic the controller, resource allocator, and actuator organization (Fig 1A).

For the resource allocation scheme to function correctly,  $\sigma$  fragments need to saturate the core fragment, causing total RNAP activity to plateau above a certain total concentration of  $\sigma$  fragments. The maximum level of polymerase activity is then set by the concentration of the core fragment, independent of changes in  $\sigma$  fragment expression (Fig 1C). Core fragment expression, and thus overall maximum functional polymerase expression, can be modulated by selecting constitutive promoters and RBSs of different strengths. This saturation behavior is observed when the core fragment is fused to the SynZIP 17 domain (Fig 3B, red points). The RNAP activity saturates approximately fourfold below that obtained with the expression of full-length T7\* RNAP in place of the core fragment, which does not change as a function of  $\sigma$  fragment expression (green points). Since the full-length T7\* RNAP is expressed at a level equivalent to the core fragment, this indicates that the split polymerase with SynZIPs has about one quarter the activity of full-length T7\* RNAP. Without the SynZIP domain on the core fragment, the  $\sigma$  fragment binds with much lower affinity and does not reach saturation even at high levels of expression (blue points). Because the desired saturation of the core fragment is obtained only with the SynZIPs, they were used in all further experiments.

A key feature of the allocator is to be able to direct transcriptional resources to different actuators. This requires multiple  $\sigma$  fragments that can bind to the core fragment to change its promoter affinity. These  $\sigma$  fragments need to be orthogonal, that is, they cannot cross-react with each other's promoters. Initially, we attempted to base the orthogonal  $\sigma$  fragments on a set of specificity loop mutations previously shown to generate orthogonal variants of full-length T7 RNAP (Temme et al, 2012a). These specificity loops are based on polymerases from the T3, K1F, and N4 phages. We tested the corresponding  $\sigma$  fragments and mutated promoters. Unfortunately, of these variants, only the  $\sigma$  fragment containing the T3 specificity loop and corresponding promoter (Fig 3C) generated an activity comparable to that of the T7  $\sigma$  fragment (Fig 3D).

The  $\sigma$  fragments based on the K1F and N4 specificity loops did have some residual activity. This was used as a basis to apply error-prone PCR to the  $\sigma$  fragments to search for mutations that increase activity (Materials and Methods). One mutation was found for the K1F loop (K1FR: M750R) that recovered activity to a sufficient level, but similar efforts with the N4 loop proved unsuccessful (Supplementary Information Section III.A.). An additional  $\sigma$  fragment was built based on an orthogonal T7 RNAP variant (CGG-R12-KIR) that was identified from directed evolution experiments (Ellefson et al,



**Figure 3. Activation of the core fragment via  $\sigma$  fragments.**

- A A schematic of the induction system is shown; the core fragment is expressed at a constant level from a constitutive promoter.
- B The T7  $\sigma$  fragment (SZ-601:883) is induced in the presence of different core fragments, and the activity of  $P_{T7}$  is measured. Red and blue points show the induction in the presence and absence of the SynZIP, respectively (core fragments 1:601-SZ and 1:601). The activity of full-length T7\* RNAP is shown as a positive control (green). A negative control with no core fragment is shown (black). The leftmost point (marked '(–)') represents cells that did not encode the T7  $\sigma$  fragment. From left to right, the remaining points represent induction levels of: 0, 1, 2, 4, 6, 3, 10, 16, 25, 40, 63, 100, and 1,000  $\mu$ M IPTG.
- C The variations between the  $\sigma$  fragments and promoters are shown. Position 632 indicates the mutation made in T7\* RNAP that reduces toxicity, and positions 739–772 show the DNA-binding loop.
- D The activities of each of the four  $\sigma$  fragments are shown with their cognate promoters when expressed to saturation (100  $\mu$ M IPTG) with the core fragment.
- E The cross-reactivity of each  $\sigma$  fragment with each promoter is shown (100  $\mu$ M IPTG induction of the  $\sigma$  fragments and constant core fragment expression). The underlying activity levels and variation for this assay are shown in Supplementary Fig S5.

Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

Source data are available online for this figure.

2013). This produced a comparable activity to the other  $\sigma$  fragments (Fig 3D). In total, four  $\sigma$  fragment variants (T7, T3, K1FR, and CGG) and cognate promoters were built. It is noteworthy that the  $\sigma$  fragments only differ in sequence by 5–10 amino acids (Fig 3C). Expression of each  $\sigma$  fragment with its cognate promoter and the

same level of core fragment shows that their activities fall into a similar range with less than a fourfold difference between the strongest (T7) and weakest (T3)  $\sigma$  fragments (Fig 3D). The four  $\sigma$  fragments were also found to be orthogonal (Fig 3E), and their expression to saturation with the core fragment does not lead to growth defects (Supplementary Fig S10).

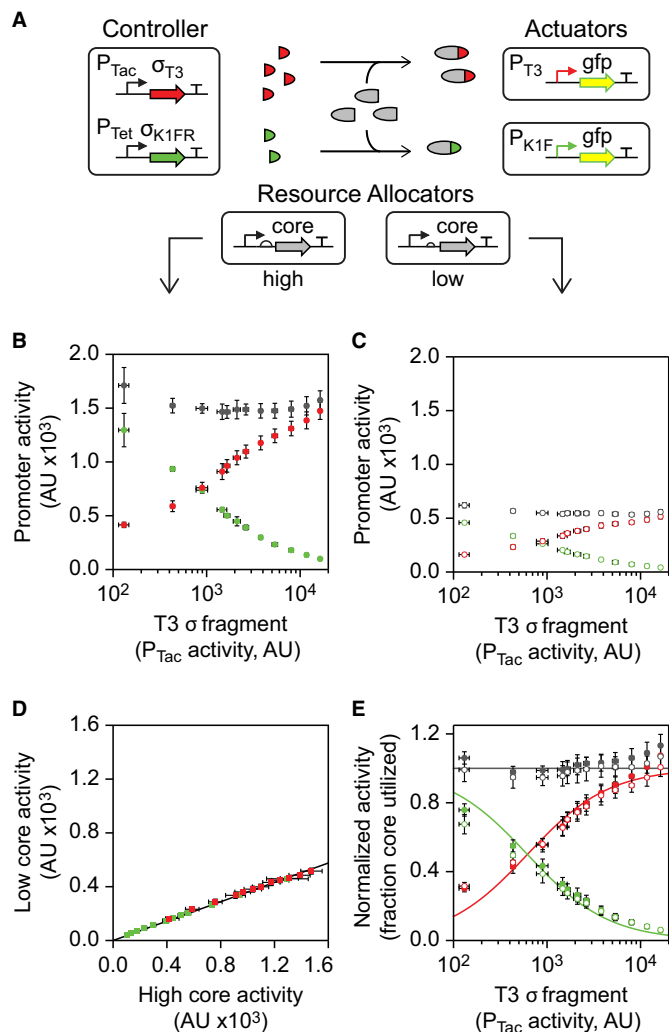
### Setting and sharing the transcriptional budget

The expression level of the core fragment from the resource allocator sets the maximum number of active RNAPs in the synthetic system. This budget has to be shared between  $\sigma$  fragments that are expressed simultaneously (Fig 1C). To test this, we built a plasmid where the K1FR  $\sigma$  fragment is expressed from  $P_{Tet}$  and the T3  $\sigma$  fragment is expressed from  $P_{Tac}$  (Fig 4A). By inducing the system with IPTG, the level of expression of the T3  $\sigma$  fragment is varied while the K1FR  $\sigma$  fragment is maintained at a constant level ( $P_{Tet}$  is uninduced but has leaky expression). In essence, this captures the scenario where one output of a controller is constantly on at a saturating level and then another output turns on and competes for the RNAP resource. To report how much of each type of polymerase complex is present in the system, reporter plasmids that express GFP from  $P_{T3}$  and  $P_{K1F}$  were used. The activity of the  $\sigma_{T3}:P_{T3}$  and  $\sigma_{K1FR}:P_{K1F}$  pairs are very similar (Fig 3D), making it possible to compare their expression levels.

Core fragment expression was driven by the  $P_{J23105}$  promoter with RBSs of different strengths. Initially, a strong RBS was chosen that sets a high expression level of the core fragment (Fig 4B). The K1FR  $\sigma$  fragment utilizes the majority of the core fragment budget before the T3  $\sigma$  fragment is induced. As the T3  $\sigma$  fragment is induced, it competes for the core fragment. At high concentrations, it saturates the pool of core fragment, almost completely titrating it from binding to the K1FR  $\sigma$  fragment. The sum of the  $P_{K1F}$  and  $P_{T3}$  promoter activities (gray points) remains constant and is independent of the expression of either  $\sigma$  fragment. The competition experiment was repeated with the core fragment expressed at a lower level from a weaker RBS (Fig 4C). Importantly, the expression level of the K1F  $\sigma$  fragment and the induction of the T3  $\sigma$  fragment remain unchanged. As before, the sum of activities from the  $P_{T3}$  and  $P_{K1F}$  promoters remains constant. Both of these competition systems are tolerated by cells with little growth impact at the induction levels used (Supplementary Fig S11).

The shapes of the curves are essentially identical when compared for high and low concentrations of the core fragment. The similarity is shown by plotting the  $P_{T3}$  and  $P_{K1F}$  promoter activities with low core fragment expression against their activities with high core fragment expression (Fig 4D). This results in a linear relationship, meaning that all promoter activities scale equally with the amount of core fragment expressed. The slope of this line indicates that the low level of core fragment yields approximately 36% of the activity compared to the high level. Hence, the budget is shared identically between the  $\sigma$  fragments at each core fragment expression level. This property means that the proportional outputs of the resource allocator can be set independently from the level of resource being produced.

To correct for the slight activity difference between the T3 and K1FR systems, we normalized the  $P_{T3}$  and  $P_{K1F}$  activity values by the activity when each individual  $\sigma$  fragment is expressed to saturation with the appropriate resource allocator (Fig 4E). Assuming



**Figure 4. Competition between  $\sigma$  fragments to bind the core fragment.**

- A The genetic system used for the competition assays is shown. Two resource allocator plasmids were built that generate high and low core fragment expression levels via a strong or weak RBS and constitutive promoter.
- B Data for the high resource allocator are shown. The K1FR  $\sigma$  fragment was expressed at a constant level (no induction of  $P_{Tet}$ ), and the T3  $\sigma$  fragment was induced with 0, 2, 4, 6.3, 7.4, 8.6, 10, 13, 16, 20, 25, and 32  $\mu$ M IPTG. The activities of  $P_{T3}$  (red circles) and  $P_{K1F}$  (green circles) were measured, and the sum of their activities computed (gray circles).
- C Data for the low resource allocator are shown, as in (B).
- D Each point represents promoter activity (red:  $P_{T3}$ , green:  $P_{K1F}$ ) at a specific level of inducer. The x and y values show the activity with high and low levels of core fragment expression, respectively. The line shows a linear regression, with the intercept fixed to 0.
- E Each  $\sigma$  fragment was expressed to saturation (100  $\mu$ M IPTG) with the high and low resource allocators, and the measured promoter activities were used to normalize the data shown in (B) and (C) (solid and hollow circles, respectively). The 'fraction core utilized' represents the proportion of the core fragment present in the system that is bound by either  $\sigma$  fragment, assuming a linear correlation with promoter activity. The solid lines show a simplified model of competition fit to the normalized data.

Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

Source data are available online for this figure.

that promoter activity is linearly proportional to the number of active polymerases, these normalized values represent the proportion of the available core fragment bound by each of the  $\sigma$  fragments. A mathematical model of the system was built and its dynamics analyzed (Supplementary Information Section IV.B.). When the core fragment is fully saturated by  $\sigma$  fragments, the model predicts that the proportion of the core fragment bound by each  $\sigma$  fragment should depend solely on the relative expression levels of each  $\sigma$  fragment. The simplified model has only one parameter not measured in the normalized data set: the relative expression of the K1FR  $\sigma$  fragment (Supplementary Information Section IV.C, Equations 29-30). Fitting this parameter yields a good agreement between the theory and experimental data (Fig 4E, Supplementary Equations 31-33).

### Positive and negative regulation of the core fragment

The resource allocators shown in Figs 3 and 4 maintain a constant level of core fragment. It is desirable to be able to dynamically shift the budget up or down, for example, to control the maximum transcriptional capacity as a function of media or growth phase. To do this, we used additional splits and mutations to create positive and negative regulators. These regulators could also be used to design feedback or feedforward circuits to implement control algorithms that act on the signal from the controller plasmid to the actuators.

The negative regulator is based on a 'null'  $\sigma$  fragment that binds to the core fragment but does not support transcription. This functions to sequester the core fragment in the same way as an active  $\sigma$  fragment, making less of it available to the other competing  $\sigma$  fragments. Sequestration has emerged as a generalizable method to tune the threshold and ultrasensitivity of genetic circuits by setting a concentration of sequestering molecule that must be outcompeted before the circuit turns on (Buchler & Louis, 2008; Buchler & Cross, 2009; Chen & Arkin, 2012; Rhodius *et al*, 2013). The null fragment was identified by testing amino acid substitutions and deletions identified from the literature to disrupt T7 RNAP function (Bonner *et al*, 1992; Mookhtiar *et al*, 1991). These mutations were selected to disrupt transcription activity without impacting the ability of the  $\sigma$  fragment to bind and sequester the core fragment (Supplementary Table S4). Based on the screen, we identified the Y638A mutation in the CGG  $\sigma$  fragment as having the strongest effect when sequestering the core fragment. This fragment was confirmed to carry no residual activity for its original promoter (Supplementary Fig S6).

A system was constructed to test the ability of the null fragment to titrate the core fragment and reduce its availability to the  $\sigma$  fragments (Fig 5A). For this, the  $\sigma$  fragments were expressed using a constitutive promoter derived from  $P_{J23119}$  and the null fragment was placed under  $P_{Tac}$  IPTG-inducible control on a separate plasmid. When expressed with the T7  $\sigma$  fragment, the null fragment decreases the activity from  $P_{T7}$  as it is induced (Fig 5B). The null fragment is able to compete with all of the  $\sigma$  fragments and reduces each of their activities by at least tenfold when fully induced (Fig 5C).

The positive regulator is based on further splitting the core fragment at the most N-terminal split site (Fig 2B and D). This divides the core fragment into two pieces: a short 67 amino acid ' $\alpha$  fragment' and a larger 586 amino acid ' $\beta$  core fragment' (including the SynZIP). The  $\alpha$  fragment can be expressed separately and is required

for activity. It can be used to modulate the fraction of the polymerase pool that is active. Note that it still does not enable more transcriptional activity than is set by the amount of  $\beta$  core fragment that is expressed. Thus, the maximum can be set and then the  $\alpha$  fragment used to modulate the amount that is available at any given time.

A system was constructed to assay the  $\alpha$  fragment's ability to regulate the polymerase budget (Fig 5D). The  $\beta$  core fragment is expressed from the  $P_{J23105}$  constitutive promoter on a low copy plasmid, while the T7  $\sigma$  fragment is expressed from a constitutive promoter derived from  $P_{J23119}$  on a high copy plasmid. The  $\alpha$  fragment is expressed from  $P_{Tac}$ . There is no T7 RNAP activity without the  $\alpha$  fragment and activity increases as it is induced (Fig 5E).

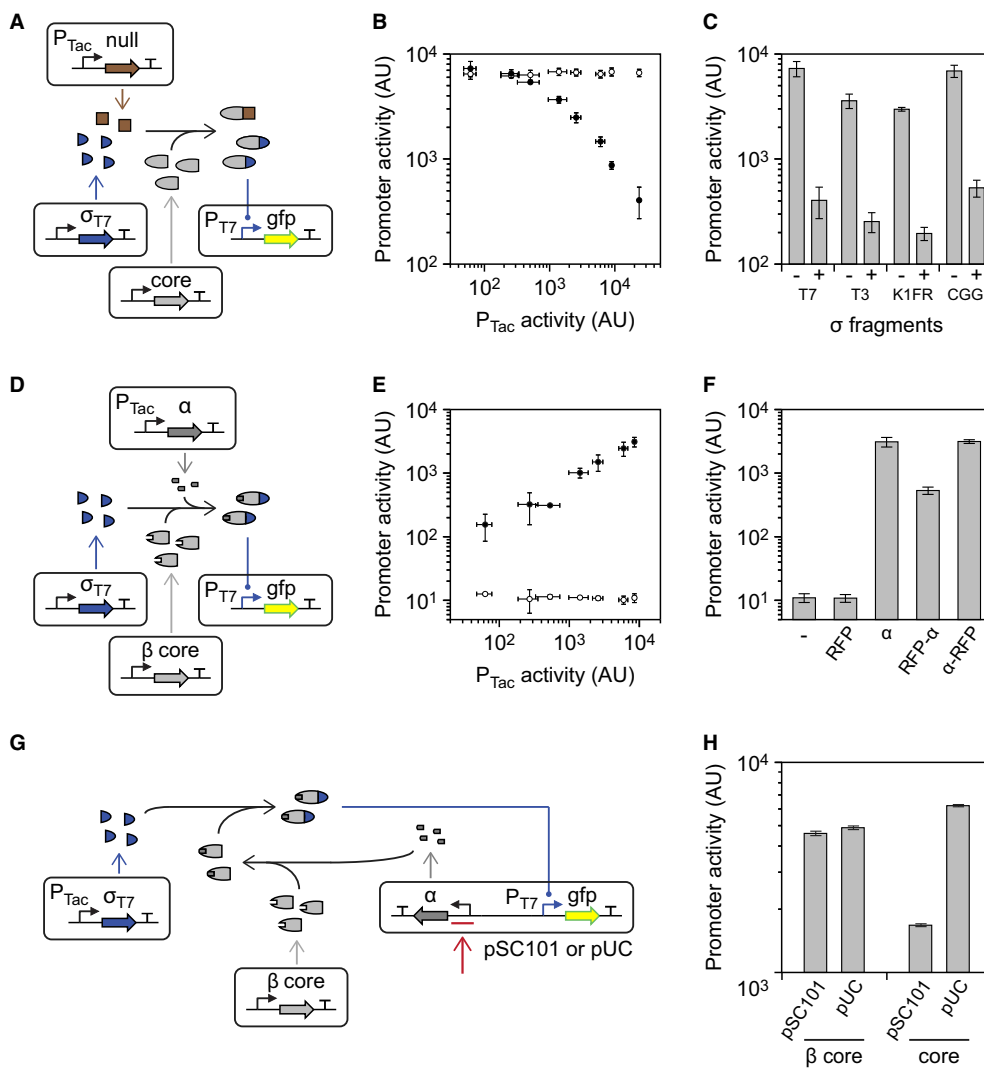
### Coupling RNAP activity to the concentration of arbitrary $\alpha$ fragment tagged proteins

Since the  $\alpha$  fragment is relatively small (67 aa) and required for polymerase function, we hypothesized that it would be useful as a protein tag to activate transcription proportional to the level of an arbitrary protein of interest. While the C-terminus of T7 RNAP catalyzes transcription and is highly sensitive to alteration, the N-terminus (where the  $\alpha$  fragment is located) is much more tolerant to modifications (Dunn *et al*, 1988). The  $\alpha$  fragment was fused to proteins of interest via a GGSGG flexible linker. Fusion to either the N- and C-terminus of RFP or GFP makes polymerase activity responsive to the level of fluorescent protein expression (Fig 5F and Supplementary Fig S7). This may be used to tag proteins in a synthetic system or the host, enabling the readout of an internal or cell state.

### Application of the $\alpha$ fragment to compensate for differences in copy number

A challenge in building genetic systems is that regulatory parts will change their activity depending on the copy number of the system. For example, a constitutive promoter will produce a high level of expression when it is placed on a high copy plasmid and a low level of activity with placed at single copy on a bacterial artificial chromosome (Kittleson *et al*, 2011). The  $\alpha$  fragment could be used to regulate the activity of the polymerase to adjust the activity of promoters and compensate for the copy number at which they are carried due to different plasmid origins (or in the genome). The idea is to combine the phage promoter(s) with an expression cassette including the  $\alpha$  fragment that is expressed at a level inversely proportional to the copy number (Fig 5G). In other words, a strong promoter and RBS would be selected to drive the expression of the  $\alpha$  fragment from a low copy plasmid and vice versa.

Plasmids were constructed on pSC101 and pUC backbones that contain a  $P_{T7}$  promoter driving GFP expression and a  $\alpha$  fragment expression cassette. We mutagenized the RBSs and altered the promoters and start codon of the  $\alpha$  fragment expression cassettes to identify a strong cassette that would be carried on the pSC101 plasmid and weak cassette that would be carried on the pUC plasmid (Materials and Methods). With these different levels of  $\alpha$  fragment expression, we were able to achieve nearly identical activities for  $P_{T7}$  in the different plasmid contexts when they are used with the  $\beta$  core fragment (Fig 5H). In contrast, when the plasmids are used



**Figure 5. Positive and negative post-transcriptional regulation of the core fragment.**

- A Null fragment sequestration of the core fragment.
- B The core fragment and T7  $\sigma$  fragment are expressed constitutively, while null fragment expression is induced from  $P_{Tac}$  (induction from left to right is: 0, 2, 4, 10, 16, 25, 40, and 1000  $\mu$ M IPTG). The effect of the expression of the null fragment on  $P_{T7}$  activity is shown as black circles. The activity of  $P_{T7}$  under the same conditions lacking the inducible null fragment cassette is shown as white circles.
- C The null fragment is shown in competition with each of the four  $\sigma$  fragments. Data are shown when the null fragment is uninduced (–, 0  $\mu$ M IPTG) and induced (+, 1000  $\mu$ M IPTG).
- D Activation of the  $\beta$  core fragment through the expression of the  $\alpha$  fragment.
- E The impact of expressing the  $\alpha$  fragment from the  $P_{Tac}$  promoter is shown. The black and white circles show induction in the presence and absence of the  $\alpha$  fragment cassette, respectively (from left to right: 0, 2, 4, 10, 16, 25, and 40  $\mu$ M IPTG). The high level for uninduced is due to leaky expression from  $P_{Tac}$ .
- F The ability of  $\alpha$  fragment : RFP fusions to complement the  $\beta$  core fragment (with the T7  $\sigma$  fragment) is shown. From left to right: (–), no inducible cassette; RFP, expression of unmodified RFP;  $\alpha$ , expression of free  $\alpha$  fragment; RFP- $\alpha$ , expression of a C-terminal fusion of  $\alpha$  fragment to RFP;  $\alpha$ -RFP, expression of an N-terminal fusion. Each system was induced with 40  $\mu$ M IPTG.
- G A genetic system is shown that uses  $\alpha$  fragment expression from a constitutive promoter to compensate for the effects of differences in copy number. A strong constitutive promoter and RBS controlling  $\alpha$  expression (red arrow) are selected at low copy (pSC101), while a weaker promoter and RBS are used at high copy (pUC).
- H Data are shown for a pair of pSC101 and pUC plasmids carrying tuned  $\alpha$  fragment cassettes and a  $P_{T7}$  promoter driving GFP. '  $\beta$  core ' indicates that the  $\beta$  core fragment and T7  $\sigma$  fragment are co-expressed. ' core ' indicates that the core fragment and T7  $\sigma$  fragment are co-expressed.

Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

with the full core fragment, which does not need the  $\alpha$  fragment to function, high expression is seen from the high copy pUC backbone and low expression is seen from the low copy pSC101 backbone.

One of the values of this approach is that it enables actuators that require multiple phage promoters to be moved to different copy

number contexts without having to change and rebalance each of the promoters. For example, actuators that produce deoxychromoviridans, nitrogenase, and lycopene require 2, 4, and 5 phage promoters (Temme *et al*, 2012a,b). These could be moved to different copy number backbones without changing their genetics by



changing the expression level of the  $\alpha$  fragment from that backbone. One can also imagine harnessing feedback or feedforward loops that self-adjust the level of  $\alpha$  fragment to maintain constant promoter activity independent of context, similar to systems that have been implemented in mammalian cells (Bleris *et al*, 2011).

## Discussion

As a means to organize and control large genetic engineering projects, we propose to introduce a separate resource allocator module. The allocator is responsible for providing resources that are orthogonal to those required by the host for growth and maintenance. To that end, this manuscript focuses on budgeting transcriptional resources through the control of phage polymerase activity and promoter specificity. Thinking ahead, this approach can be extended to budget additional resources. For example, translational resources could be incorporated by controlling a orthogonal rRNA (Rackham & Chin, 2005; An & Chin, 2009) (specific to RBSs only in the synthetic system) or even introducing an entire second ribosome. Extending this idea, it may be possible to incorporate orthogonal tRNAs (Liu *et al*, 1997; Chin, 2014), DNA replication machinery (Ravikumar *et al*, 2014), protein degradation machinery (Grilly *et al*, 2007), carbon precursors (Pfeifer *et al*, 2001), and organelle structures (Moon *et al*, 2010; Bonacci *et al*, 2012). While this never completely decouples the synthetic system from the host, it systematically reduces its dependence on host resources and genetic idiosyncrasies. This approaches the concept of a 'virtual machine' for cells, where synthetic systems would bring all of the necessary cellular machinery with them. This concept will become critical as designs become larger, moving toward the scale of genomes and requiring the simultaneous control over many multi-gene actuators.

This work demonstrates an incredible tolerance of the T7 RNAP structure for division into multiple proteins without disrupting its function. To our knowledge, this is the first time that a protein has been artificially divided into four fragments that can be functionally co-expressed. This tolerance is surprising because T7 RNAP is known to undergo large-scale conformational changes as it proceeds from promoter binding to transcription elongation (Ma *et al*, 2002; Guo *et al*, 2005). The residues involved in these conformational changes occur toward the N-terminal region but are distributed across the first three fragments of the 4-fragment polymerase (Fig 2E). All of the RNAP split points were discovered simultaneously using a new experimental method, which we refer to as a 'splitposon'. This approach is faster, simpler, and produces more accurate split proteins than previous methods. Split proteins have applications in genetic circuits (Shis & Bennett, 2013; Mahdavi *et al*, 2013), plasmid maintenance with fewer antibiotics (Schmidt *et al*, 2012), and biosensors (Johnsson & Varshavsky, 1994; Galarneau *et al*, 2002; Hu & Kerppola, 2003; Michnick *et al*, 2007; Camacho-Soto *et al*, 2014).

The fragments of T7 RNAP are used to implement regulatory control. A C-terminal fragment contains the DNA-binding loop and we demonstrate that fragments with different specificities can direct the RNAP to different promoters. For this reason, and because of its size, we draw a loose analogy to the role of  $\sigma$  factors in native prokaryotic transcription. However, there are notable differences between our  $\sigma$  fragments compared to natural  $\sigma$  factors. First, core *E. coli* RNAP binds to DNA in a non-specific manner and this is

titrated away by the  $\sigma$  factors (Grigorova *et al*, 2006; Bratton *et al*, 2011). It is unlikely that our T7 RNAP core fragment binds to DNA. Second, a prokaryotic  $\sigma$  factor only recruits the RNAP to the promoter and once transcription initiation is complete, the  $\sigma$  factor dissociates during transcription (Travers & Burgess, 1969; Raffaele *et al*, 2005). Thus, the ratio of  $\sigma$  factors to core RNAP is low (~50%) because they only have to compete to bind to free (non-transcribing) polymerase (Ishihama, 2000). Our system requires larger ratios, because the  $\sigma$  fragments must remain associated with the core fragment during transcription. Third, while the size of a  $\sigma$  factor and the  $\sigma$  fragment are about the same, their 3-dimensional structure and mechanism of binding to core and DNA are different (Vassylyev *et al*, 2002). Finally, recent results suggest that the *B. subtilis* core RNAP is shared by  $\sigma$  factors in time as opposed to concentration (Levine *et al*, 2013). In other words, the  $\sigma$  factors pulse in a mutually exclusive manner to take turns fully utilizing the pool of core RNAP. In contrast, our  $\sigma$  fragments compete for the core fragment following mass action kinetics. This is similar to the previous understanding, where differences in  $\sigma$  factor binding affinities are a means that cells prioritize and order different responses (Lord *et al*, 1999; Maeda *et al*, 2000; Grigorova *et al*, 2006).

Resource allocation also occurs in natural regulatory networks. In bacteria, alternative  $\sigma$  factors can redirect RNAP to different condition-specific promoters. Factors such as ppGpp and 6S RNA also regulate the pool of active free RNAP (Jensen & Pedersen, 1990; Wassarman & Storz, 2000; Klumpp & Hwa, 2008). Using up this resource has been observed and shown to result in a slower growth rate (Farewell *et al*, 1998). Further, the competition between  $\sigma$  factors for core RNAP has been quantified (De Vos *et al*, 2011; Grigorova *et al*, 2006). Keren and co-workers measured the activity of thousands of native *E. coli* and *S. cerevisiae* promoters under different environmental conditions (Keren *et al*, 2013). They found that while changes in conditions have a global impact on many promoters, they shift by a linear factor that is characteristic of each condition. This factor ranges from 0.51 to 1.68 with M9 + glucose being the reference condition. They found that a simple model that treats overall promoter activity as a fixed resource explains their data. Overall promoter activity is equivalent to the total active RNAP concentration that forms the backbone of our resource allocator and the ratio of 0.36 shown in Fig 4D is analogous to their linear factor when moving from the high to the low resource allocator.

In the context of synthetic signaling networks, retroactivity occurs when downstream regulation impacts an upstream process. For example, the titration of ribosomes or proteases by one branch of the network can influence the network as a whole (Cookson *et al*, 2011). This is viewed as an undesirable effect that must be buffered against in order to maintain computational integrity (Del Vecchio & Murray, 2014). In contrast, the resource allocator harnesses retroactivity in order to budget transcription to different pathways without surpassing a limit. As an allocation mechanism, retroactivity is an ideal means of distributing a budgeted resource. Currently, this is limited to dividing the core fragment among the  $\sigma$  fragments in a way that is proportional to their expression levels. Building on this, more complex dynamics could be introduced that implement signal processing between the output of the controller plasmid and the actuators that are being regulated. For instance, it may be desirable to control several actuators via a mutually exclusive or analog relationship, for example to slow down a metabolic pathway as a

molecular machine is being built. Other actuators may require graded or ultrasensitive responses, for example the all-or-none commitment to flagellum construction versus simply changing the level of an enzyme. The toolbox presented in this paper provides a means to rationally design such control that can be implemented on the signal from the output of circuitry encoded on a controller to the actuators.

## Materials and Methods

### Strains and media

*Escherichia coli* DH10B was used for all routine cloning and characterization. ElectroMAX competent cells (Life Technologies) were used for library cloning steps as noted. LB-Miller media was used for assays and strain propagation, 2YT media was used for strain propagation, and SOC media was used for transformation recovery. Antibiotics were used as necessary for plasmid maintenance, with ampicillin at 100 µg/ml, spectinomycin at 100 µg/ml, kanamycin at 50 µg/ml, and chloramphenicol at 17 µg/ml. IPTG (isopropyl β-D-1-thiogalactopyranoside) was used as an inducer at concentrations up to 1 mM.

### Plasmids and parts

Plasmids with the ColE1 origin were based off of the plasmid pSB1C3 from the Registry of Standard Biological Parts, which has a pUC19 (Yanisch-Perron *et al*, 1985) derived origin. Plasmids with the pUC origin were based off of a pUC19 (Yanisch-Perron *et al*, 1985) vector. Plasmids with the p15A\* origin were based off of plasmid pSB3C5 (Shetty *et al*, 2008) from the Registry. This origin appears to maintain at a higher copy number than standard for p15A. Plasmids with the pSC101 origin were based on pUA66 (Zaslaver *et al*, 2006). Plasmids with the BAC origin were based on pBACr-Mgr940 (Anderson *et al*, 2007) (BBa\_J61039), which has an F plasmid derived origin. A P<sub>Tac</sub> promoter system derived from pEXT20 (Dykxhoorn *et al*, 1996) modified to contain a symmetric LacI binding site or a shortened version of this expression system was used in all systems that required inducible expression. Constitutive protein expression was driven by promoter P<sub>J23105</sub> (BBa\_J23105) or P<sub>J23109</sub> (BBa\_J23109), by a modified P<sub>Tet</sub> expression system (Moon *et al*, 2012) (uninduced), and by promoters selected from libraries derived from P<sub>J23119</sub> (BBa\_J23119) through degenerate PCR. RBSs were either generated using the RBS calculator, taken from the Registry (BBa\_B0032 and BBa\_B0034 (Elowitz & Leibler, 2000)), or selected from libraries generated using degenerate PCR. The RiboJ insulator (Lou *et al*, 2012) was used between P<sub>Tac</sub> or P<sub>Tet</sub> and the RBS in all constructs when titration curves were run. mRFP1 (Campbell *et al*, 2002) and sfGFP (Pédrelacq *et al*, 2006) were used as fluorescent reporters. Representative plasmid maps are shown in Supplementary Figs S2, S9, and S13 through S19. A list of new plasmids is given in Supplementary Table S6. Select constructs from this study will be made available online through Addgene ([http://www.addgene.org/Christopher\\_Voigt/](http://www.addgene.org/Christopher_Voigt/)).

### Bisection mapping T7 RNA polymerase

The splitposon was generated by modifying the HyperMu <KAN-1> transposon (Epicentre Biotechnologies). Examining previously

described variants of the MuA transposon system (Goldhaber-Gordon *et al*, 2002; Poussu *et al*, 2004, 2005; Jones, 2006; Hoeller *et al*, 2008), a number of terminal bases were identified that could be altered while maintaining transposition activity. The RBS calculator (Salis, 2011) was used to design a strong terminal RBS and start codon while staying within these alterations. This modified end was combined with a previously built end containing terminal stop codons (Poussu *et al*, 2005). A P<sub>Tac</sub> promoter and constitutive LacI expression cassette were inserted into the transposon to drive transcription at the end with the RBS and start codon. Finally, point mutations were made to remove restriction sites that would interfere with downstream cloning steps. A region of the T7\* RNA polymerase CDS encoding aa 41–876 was flanked by BsaI sites in a ColE1 AmpR backbone. The splitposon (KanR) was transposed into this plasmid with MuA transposase (300 ng target DNA, 200 ng transposon, MuA buffer, 1.1 U HyperMuA transposase (Epicentre Biotechnologies), 30°C 8 h, 75°C 10 min), DNA clean and concentrated (Zymo), electroporated into ElectroMAX cells and plated on LB + Kan/Amp plates to obtain > 700,000 colonies. The colonies were scraped from the plates, pooled, and miniprepmed to obtain DNA of the transposon insertion library. The transposon insertion library was digested with BsaI, run on an agarose gel, and a band of ~5.7 kb (representing the section of the T7 CDS plus transposon) was excised, gel-purified (Zymo), and DNA clean and concentrated. A plasmid containing an inducible P<sub>Tac</sub> system and the remainder of the T7 CDS (aa 1–40 and 877–883) with internal BsaI sites on a p15A\* SpecR backbone was digested with BsaI and the size-selected fragment ligated into it. This reaction was DNA clean and concentrated, electroporated into ElectroMAX cells plated on LB + Spec/Kan plates to obtain > 600,000 colonies, and the colonies were scraped, pooled, and miniprepmed as before to obtain the bisected library. This library was electroporated into *E. coli* DH10B cells with a plasmid containing a P<sub>T7</sub>-RFP cassette on a pSC101 CamR backbone (Nif\_489 (Temme *et al*, 2012a)), plated on LB + Spec/Kan/Cam, and visually red colonies were picked after 16 h of growth for analysis in liquid media. More information on the splitposon method and T7 RNAP bisection mapping are included in Supplementary Information Sections I and II.

### Assay protocol

All promoter activity assays except the initial assay of T7 bisection mapping were performed as follows. Cells containing the plasmids of interest were inoculated from glycerol stocks into 0.5 ml LB-Miller media plus antibiotics in a 2-ml 96-deepwell plate (USA Scientific) sealed with an AeraSeal film (Excel Scientific) and grown at 37°C, 900 rpm overnight (~14–16 h) in a deepwell shaker. These overnights were diluted 200-fold into 150 µl LB-M with antibiotics plus varying concentrations of IPTG in 300-µl 96-well V-bottom plates (Thermo Scientific Nunc) sealed with an AeraSeal film and grown at 37°C, 1,000 rpm for 6 h. 5 µl of each sample was removed and diluted in 195 µl PBS + 2 mg/ml kanamycin to halt protein production. Cells diluted in PBS were either characterized immediately with flow cytometry or stored at 4°C until characterization. The initial T7 bisection mapping assays were performed similarly except the overnight cultures were grown in 2YT, and the overnight cultures were diluted 1:10 into 150 µl induction media.

## Flow cytometry characterization

All fluorescence characterization was performed on a BD LSR Fortessa flow cytometer with HTS attachment and analyzed using FlowJo vX (TreeStar). Cells diluted in PBS + kanamycin were run at a rate of 0.5  $\mu\text{l/s}$  until up to 100,000 events were captured (at least 50,000 events were recorded in all cases). The events were gated by forward scatter and side scatter to reduce false events and by time to reduce carry-over events. Gating was determined by eye and was kept constant for all analysis within each triplicate experiment. For all assays except the initial characterization of T7 bisection mapping, the geometric mean value of fluorescence was calculated for each sample, using a biexponential transform with a width basis of  $-10.0$  to allow calculations with negative values. Finally, white-cell fluorescence measured concurrently from cells lacking fluorescent protein was subtracted from measured fluorescence to yield the Promoter activity (AU) values presented in the figures. The initial T7 bisection mapping assay was characterized identically, except that white-cell values were not subtracted.

Where fold induction calculations were required, fluorescence measurements were made of cells containing the appropriate reporter construct and lacking a functional polymerase, grown in the same conditions as the test cells. The fold induction is reported as the ratio of the white-cell-corrected test cell fluorescence to the white-cell-corrected fluorescence of the reporter-only cells.

To obtain relative expression levels for the polymerase fragments driven by  $P_{T_{ac}}$ , constructs were made that express GFP after  $P_{T_{ac}}$  and RiboJ (Supplementary Fig S9). For each assay, cells with this construct were induced under the same conditions as the test cells, and their fluorescence measured (Supplementary Fig S8). The  $P_{T_{ac}}$  activity value in each plot represents the geometric mean white-cell-corrected fluorescence of these cells for that assay, and the horizontal error bars show the standard deviation of those measurements.

## Measuring the growth impact of split polymerase expression

Cells containing the plasmids of interest were inoculated from colonies on agar plates into 0.5 ml LB-Miller media plus antibiotics in a 2-ml 96-deepwell plate, sealed with an AeraSeal film, and grown at 37°C, 900 rpm overnight (~14–16 h) in a deepwell shaker. These overnights were diluted 200-fold into 150  $\mu\text{l}$  LB-M with antibiotics plus varying concentrations of IPTG in 300- $\mu\text{l}$  96-well V-bottom plates, sealed with an AeraSeal film, and grown at 37°C, 1,000 rpm for 6 h. 20  $\mu\text{l}$  of each sample were added to 80  $\mu\text{l}$  LB in a 96-well optical plate (Thermo Scientific Nunc), and the  $OD_{600}$  of each diluted sample was measured using a BioTek Synergy H1 plate reader. These measurements were normalized by dividing by the  $OD_{600}$  of samples containing plasmids with the same backbones but expressing none of the proteins of interest (polymerase fragments or GFP) at each level of IPTG induction. Growth data are shown in Supplementary Figs S10, S11 and S12.

## Error-prone PCR of $\sigma$ fragment variants

Sections of the K1F and N4 T7 RNAP variants (Temme *et al*, 2012a) were amplified using GoTaq (Promega) in 1 $\times$  GoTaq buffer plus  $\text{MgCl}_2$  to a final concentration of 6.5 mM  $\text{Mg}^{2+}$ . The amplified

fragments were cloned into a  $\sigma$  fragment expression plasmid including any necessary flanking RNAP sequence and the N-terminal SynZIP 18 domain. These mutated  $\sigma$  fragments were expressed with the core fragment and the appropriate promoter driving GFP. Colonies with visually improved GFP production were picked from plates, re-assayed to confirm activity, and sequenced to identify their mutations (Supplementary Tables S2 and S3). Promising variants were reconstructed to isolate their effects and the resulting new  $\sigma$  fragments assayed for activity.

## Tuning $\alpha$ fragment expression to compensate for copy number

An  $\alpha$  fragment expression cassette consisting of the constitutive promoter  $P_{J23105}$ , RiboJ, and B0032 RBS driving the  $\alpha$  fragment was inserted in the reverse direction before the  $P_{T7}$ : GFP cassette on a pSC101 reporter plasmid. These two cassettes were also inserted into a pUC19 backbone, with the weaker constitutive promoter  $P_{J23109}$  and start codon (GTG instead of ATG) in the  $\alpha$  fragment cassette. Degenerate PCR was used to randomize the RBS in each plasmid at five nucleotides, and the resulting libraries were screened for fluorescence in the presence of the  $\sigma_{T7}$  and either core or  $\beta$  core fragments. Sets of pSC101 and pUC plasmids were selected that had similar levels of activity with the  $\beta$  core fragment, but retained different levels of activity with the core fragment. These plasmids were isolated, sequenced, re-assayed, and the pair of pSC101 and pUC plasmids with the closest levels of expression in the presence of the  $\beta$  core fragment was selected.

Supplementary information for this article is available online: <http://msb.embopress.org>

## Acknowledgements

This work was supported by the United States Office of Naval Research (N00014-13-1-0074), the United States National Institutes of Health (5R01GM095765), and the US National Science Foundation Synthetic Biology Engineering Research Center (SA5284-11210). THSS was supported by the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program and by a Fannie and John Hertz Foundation Fellowship.

## Author contributions

THSS and CAV conceived of the study. THSS carried out experiments. AJM and ADE developed the CCG T7 RNAP variant. THSS and EDS modeled and analyzed the system. THSS and CAV wrote the manuscript with input and contributions from all of the authors.

## Conflict of interest

A patent application has been filed on some aspects of this work, with THSS and CAV as inventors.

## References

- Alexander WA, Moss B, Fuerst TR (1992) Regulated expression of foreign genes in vaccinia virus under the control of bacteriophage T7 RNA polymerase and the *Escherichia coli* lac repressor. *J Virol* 66: 2934–2942
- An W, Chin JW (2009) Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci* 106: 8477–8482

- Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. *Mol Syst Biol* 3: 133
- Arkin AP, Fletcher DA (2006) Fast, cheap and somewhat in control. *Genome Biol* 7: 114
- Bedwell DM, Nomura M (1986) Feedback regulation of RNA polymerase subunit synthesis after the conditional overproduction of RNA polymerase in *Escherichia coli*. *Mol Gen Genet* 204: 17–23
- Birnbaum S, Bailey JE (1991) Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol Bioeng* 37: 736–745
- Bleris L, Xie Z, Glass D, Adadey A, Sontag E, Benenson Y (2011) Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol* 7: 519
- Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver PA, Savage DF (2012) Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci* 109: 478–483
- Bonner G, Patra D, Lafer EM, Sousa R (1992) Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J* 11: 3767–3775
- Bratton BP, Mooney RA, Weisshaar JC (2011) Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. *J Bacteriol* 193: 5138–5146
- Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella: cellular and molecular biology*, Neidhardt FC, Curtis III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE (eds), 2nd edn, pp 1553–1569. Washington, DC: ASM Press
- Bremer H, Dennis P (2008) Feedback control of ribosome function in *Escherichia coli*. *Biochimie* 90: 493–499
- Brunschwig E, Darzins A (1992) A two-component T7 system for the overexpression of genes in *Pseudomonas aeruginosa*. *Gene* 111: 35–41
- Buchler NE, Louis M (2008) Molecular titration and ultrasensitivity in regulatory networks. *J Mol Biol* 384: 1106–1119
- Buchler NE, Cross FR (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol* 5: 272
- Burton Z, Burgess RR, Lin J, Moore D, Holder S, Gross CA (1981) The nucleotide sequence of the cloned rpoD gene for the RNA polymerase sigma subunit from *E. coli* K12. *Nucleic Acids Res* 9: 2889–2903
- Camacho-Soto K, Castillo-Montoya J, Tye B, Ghosh I (2014) Ligand-gated split-kinases. *J Am Chem Soc* 136: 3995–4002
- Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) A monomeric red fluorescent protein. *Proc Natl Acad Sci* 99: 7877–7882
- Carlson JC, Badran AH, Guggiana-Nilo DA, Liu DR (2014) Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol* 10: 216–222
- Carrera J, Rodrigo G, Singh V, Kirov B, Jaramillo A (2011) Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnol J* 6: 773–783
- Chamberlin M, Mcgrath J, Waskell L (1970) New RNA Polymerase from *Escherichia coli* infected with Bacteriophage T7. *Nature* 228: 227–231
- Cheetham GM, Jeruzalmi D, Steitz TA (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399: 80–83
- Cheetham GM, Steitz TA (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286: 2305–2309
- Chen Y, Li S, Chen T, Hua H, Lin Z (2009) Random dissection to select for protein split sites and its application in protein fragment complementation. *Protein Sci* 18: 399–409
- Chen D, Arkin AP (2012) Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Mol Syst Biol* 8: 620
- Chin JW (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem* 83: 379–408
- Conrad B, Savchenko RS, Breves R, Hofemeister J (1996) A T7 promoter-specific, inducible protein expression system for *Bacillus subtilis*. *Mol Gen Genet* 250: 230–236
- Cookson NA, Mather WH, Danino T, Mondragón-Palomino O, Williams RJ, Tsimring LS, Hasty J (2011) Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol Syst Biol* 7: 561
- De Vos D, Bruggeman FJ, Westerhoff HV, Bakker BM (2011) How molecular competition influences fluxes in gene expression networks. *PLoS ONE* 6: e28494
- Del Vecchio D, Murray RM (2014) *Biomolecular Feedback Systems*. Princeton, NJ: Princeton University Press
- Del Vecchio D, Ninfa AJ, Sontag ED (2008) Modular cell biology: retroactivity and insulation. *Mol Syst Biol* 4: 161
- Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J Bacteriol* 177: 1497–1504
- Dunn JJ, Krippel B, Bernstein KE, Westphal H, Studier FW (1988) Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene* 68: 259–266
- Dykxhoorn DM, St. Pierre R, Linn T (1996) A set of compatible tac promoter expression vectors. *Gene* 177: 133–136.
- Ellefsen JW, Meyer AJ, Hughes RA, Cannon JR, Brodbelt JS, Ellington AD (2013) Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol* 32: 97–101
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–338
- Elroy-Stein O, Moss B (1990) Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *Proc Natl Acad Sci* 87: 6743–6747
- El-Samad H, Kurata H, Doyle JC, Gross CA, Khammash M (2005) Surviving heat shock: control strategies for robustness and performance. *Proc Natl Acad Sci USA* 102: 2736–2741
- Farewell A, Kvint K, Nyström T (1998) Negative regulation by RpoS: a case of sigma factor competition. *Mol Microbiol* 29: 1039–1051
- Fatica A, Tollervey D (2002) Making ribosomes. *Curr Opin Cell Biol* 14: 313–318
- Galarneau A, Primeau M, Trudeau L-E, Michnick SW (2002)  $\beta$ -Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat Biotechnol* 20: 619–622
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339–342
- Gausing K (1977) Regulation of ribosome production in *Escherichia coli*: synthesis and stability of ribosomal RNA and of ribosomal protein messenger RNA at different growth rates. *J Mol Biol* 115: 335–354
- Goldhaber-Gordon I, Early MH, Gray MK, Baker TA (2002) Sequence and positional requirements for DNA sites in a Mu transpososome. *J Biol Chem* 277: 7703–7712
- Grigороva IL, Phleger NJ, Mutalik VK, Gross CA (2006) Insights into transcriptional regulation and  $\sigma$  competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci* 103: 5332–5337

- Grilly C, Stricker J, Pang WL, Bennett MR, Hasty J (2007) A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*. *Mol Syst Biol* 3: 127
- Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 57: 441–466
- Guntas G, Ostermeier M (2004) Creation of an allosteric enzyme by domain insertion. *J Mol Biol* 336: 263–273
- Guo Q, Nayak D, Brieba LG, Sousa R (2005) Major conformational changes during T7RNAP transcription initiation coincide with, and are required for, promoter release. *J Mol Biol* 353: 256–270
- Hayward RS, Tittawella IP, Scaife JG (1973) Evidence for specific control of RNA polymerase synthesis in *Escherichia coli*. *Nat New Biol* 243: 6–9
- Hengge-Aronis R (2002) Signal transduction and regulatory mechanisms involved in control of the  $\sigma$  S (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* 66: 373–395
- Hoffmann F, Rinas U (2004) Stress induced by recombinant protein production in *Escherichia coli*. *Adv Biochem Eng Biotechnol* 89: 73–92
- Hu C-D, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* 21: 539–545
- Hoeller BM, Reiter B, Abad S, Graze I, Glieder A (2008) Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). *J Microbiol Methods* 75: 251–257
- Ikeda RA, Richardson CC (1987a) Interactions of a proteolytically nicked RNA polymerase of bacteriophage T7 with its promoter. *J Biol Chem* 262: 3800–3808
- Ikeda RA, Richardson CC (1987b) Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. *J Biol Chem* 262: 3790–3799
- lost I, Guillerez J, Dreyfus M (1992) Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol* 174: 619–622
- Ishihama A, Taketo M, Saitoh T, Fukuda R (1976) Control of formation of RNA polymerase in *Escherichia coli*. *Fac Res* 1970–1979: 485–502
- Ishihama A (1981) Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv Biophys* 14: 1–35
- Ishihama A (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* 54: 499–518
- Iwakura Y, Ishihama A (1975) Biosynthesis of RNA polymerase in *Escherichia coli*. II. control of RNA polymerase synthesis during nutritional shift up and down. *Mol Gen Genet* 142: 67–84
- Jayanthi S, Nilgiriwala KS, Del Vecchio D (2013) Retroactivity controls the temporal dynamics of gene transcription. *ACS Synth Biol* 2: 431–441
- Jensen KF, Pedersen S (1990) Metabolic growth rate control in *Escherichia coli* may be a consequence of subsaturation of the macromolecular biosynthetic apparatus with substrates and catalytic components. *Microbiol Rev* 54: 89–100
- Jiang P, Ventura AC, Sontag ED, Merajver SD, Ninfa AJ, Del Vecchio D (2011) Load-Induced Modulation of Signal Transduction Networks. *Sci Signal* 4: ra67
- Jishage M, Iwata A, Ueda S, Ishihama A (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol* 178: 5447–5451
- Jonhsson N, Varshavsky A (1994) Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci* 91: 10340–10344
- Jones D (2006) European Patent: EP1838851 - Polypeptide mutagenesis method.
- Kanwar M, Wright RC, Date A, Tullman J, Ostermeier M (2013) Protein switch engineering by domain insertion. *Methods Enzymol* 523: 369–388
- Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, Aidelberg G, Bren A, Zeevi D, Weinberger A, Alon U, Milo R, Segal E (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol* 9: 701
- Klumpp S, Hwa T (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci USA* 105: 20245–20250
- Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139: 1366–1375
- Kittleson JT, Cheung S, Anderson JC (2011) Rapid optimization of gene dosage in *E. coli* using DIAL strains. *J Biol Eng* 5: 10
- Kittleson JT, Wu GC, Anderson JC (2012) Successes and failures in modular genetic engineering. *Curr Opin Chem Biol* 16: 329–336
- Kurata H, El-Samad H, Yi T-M, Khammash M, Doyle J (2001) Feedback regulation of the heat shock response in *E. coli*. In *Proceedings of the 40th IEEE Conference on Decision and Control*, Vol. 1, pp 837–842
- Lange R, Hengge-Aronis R (1994) The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev* 8: 1600–1612
- Lempäinen H, Shore D (2009) Growth control and ribosome biogenesis. *Curr Opin Cell Biol* 21: 855–863
- Levine JH, Lin Y, Elowitz MB (2013) Functional roles of pulsing in genetic circuits. *Science* 342: 1193–1200
- Liang S-T, Bipatnath M, Xu Y-C, Chen S-L, Dennis P, Ehrenberg M, Bremer H (1999) Activities of constitutive promoters in *Escherichia coli*. *J Mol Biol* 292: 19–37
- Liang S-T, Xu Y-C, Dennis P, Bremer H (2000) mRNA composition and control of bacterial gene expression. *J Bacteriol* 182: 3037–3044
- Liu DR, Magliery TJ, Pastrnak M, Schultz PG (1997) Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc Natl Acad Sci* 94: 10092–10097
- Lord M, Barillà D, Yudkin MD (1999) Replacement of vegetative  $\zeta$ A by sporulation-specific  $\zeta$ F as a component of the RNA polymerase holoenzyme in sporulating bacillus subtilis. *J Bacteriol* 181: 2346–2350
- Lou C, Stanton B, Chen Y-J, Munsky B, Voigt CA (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol* 30: 1137–1142
- Ma K, Temiakov D, Jiang M, Anikin M, McAllister WT (2002) Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase. *J Biol Chem* 277: 43206–43215
- Maeda H, Fujita N, Ishihama A (2000) Competition among seven *Escherichia coli*  $\sigma$  subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* 28: 3497–3503
- Maharjan R, Nilsson S, Sung J, Haynes K, Beardmore RE, Hurst LD, Ferenci T, Gudelj I (2013) The form of a trade-off determines the response to competition. *Ecol Lett* 16: 1267–1276
- Mahdavi A, Segall-Shapiro TH, Kou S, Jindal GA, Hoff KG, Liu S, Chitsaz M, Ismagilov RF, Silberg JJ, Tirrell DA (2013) A genetically encoded and gate for cell-targeted metabolic labeling of proteins. *J Am Chem Soc* 135: 2979–2982
- McBride KE, Schaaf DJ, Daley M, Stalker DM (1994) Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *Proc Natl Acad Sci* 91: 7301–7305
- Michnick SW, Ear PH, Manderson EN, Remy I, Stefan E (2007) Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov* 6: 569–582
- Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260: 289–298

- Mookhtiar KA, Peluso PS, Muller DK, Dunn JJ, Coleman JE (1991) Processivity of T7 RNA polymerase requires the C-terminal Phe882-Ala883-COO- or 'foot'. *Biochemistry* 30: 6305–6313
- Moon TS, Dueber JE, Shiue E, Prather KLJ (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab Eng* 12: 298–305
- Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. *Nature* 491: 249–253
- Moser F, Broers NJ, Hartmans S, Tamsir A, Kerkman R, Roubos JA, Bovenberg R, Voigt CA (2012) Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth Biol* 1: 555–564
- Muller DK, Martin CT, Coleman JE (1988) Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry* 27: 5763–5771
- Nierhaus KH (1991) The assembly of prokaryotic ribosomes. *Biochimie* 73: 739–755
- Nierlich DP (1968) Amino acid control over RNA synthesis: a re-evaluation. *Proc Natl Acad Sci* 60: 1345–1352
- Nyström T (2004) MicroReview: growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Mol Microbiol* 54: 855–862
- Opalka N, Brown J, Lane WJ, Twist K-AF, Landick R, Asturias FJ, Darst SA (2010) Complete structural model of *Escherichia coli* RN polymerase from a hybrid approach. *PLoS Biol* 8: e1000483
- Ostermeier M, Nixon AE, Shim JH, Benkovic SJ (1999) Combinatorial protein engineering by incremental truncation. *Proc Natl Acad Sci* 96: 3562–3567
- Paschon DE, Ostermeier M (2004) Construction of protein fragment complementation libraries using incremental truncation. *Methods Enzymol* 388: 103–116
- Pédélecq J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* 24: 79–88
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612
- Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*. *Science* 291: 1790–1792
- Poussu E, Vihinen M, Paulin L, Savilahti H (2004) Probing the  $\alpha$ -complementing domain of *E. coli*  $\beta$ -galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu *in vitro* DNA transposition. *Proteins* 54: 681–692
- Poussu E, Jäntti J, Savilahti H (2005) A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu *in vitro* transposition-based approach. *Nucleic Acids Res* 33: e104
- Rackham O, Chin JW (2005) A network of orthogonal ribosome-mRNA pairs. *Nat Chem Biol* 1: 159–166
- Raffaella M, Kanin EI, Vogt J, Burgess RR, Ansari AZ (2005) Holoenzyme switching and stochastic release of sigma factors from RNA polymerase *in vivo*. *Mol Cell* 20: 357–366
- Ravikumar A, Arrieta A, Liu CC (2014) An orthogonal DNA replication system in yeast. *Nat Chem Biol* 10: 175–177
- Reinke AW, Grant RA, Keating AE (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J Am Chem Soc* 132: 6025–6031
- Rhodus VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, Burkhardt DH, Clancy K, Peterson TC, Gross CA, Voigt CA (2013) Design of orthogonal genetic switches based on a crosstalk map of  $\sigma$ , anti- $\sigma$ , and promoters. *Mol Syst Biol* 9: 702
- Salis HM (2011) The ribosome binding site calculator. *Methods Enzymol* 498: 19–42
- Schaechter M, Maaløe O, Kjeldgaard NO (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19: 592–606
- Schmidt CM, Shis DL, Nguyen-Huu TD, Bennett MR (2012) Stable maintenance of multiple plasmids in *E. coli* using a single selective marker. *ACS Synth Biol* 1: 445–450
- Schneider DA, Ross W, Gourse RL (2003) Control of rRNA expression in *Escherichia coli*. *Curr Opin Microbiol* 6: 151–156
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099–1102
- Segall-Shapiro TH, Nguyen PQ, Dos Santos ED, Subedi S, Judd J, Suh J, Silberg JJ (2011) Mesophilic and hyperthermophilic adenylate kinases differ in their tolerance to random fragmentation. *J Mol Biol* 406: 135–148
- Shea MA, Ackers GK (1985) The OR control system of bacteriophage lambda: a physical-chemical model for gene regulation. *J Mol Biol* 181: 211–230
- Shetty RP, Endy D, Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2: 5
- Shis DL, Bennett MR (2013) Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci* 110: 5028–5033
- Smanski MJ, Bhatia S, Park Y, Woodruff L, Zhao D, Giannoukos G, Ciulla D, Busby M, Calderon J, Nicol R, Gordon B, Densmore D, Voigt CA (2014) Combinatorial design and assembly of refactored gene clusters. *Nat Biotechnol* In press
- Sousa R, Chung YJ, Rose JP, Wang B-C (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364: 593–599
- Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF)  $\sigma$  factor protein family. *Mol Microbiol* 74: 557–581
- Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113–130
- Studier FW, Davanloo P, CH, Rosenberg AH, Moffatt BA, Dunn JJ (1990) United States Patent: 4952496 – Cloning and expression of the gene for bacteriophage T7 RNA polymerase.
- Tabor S, Richardson CC (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci* 82: 1074–1078
- Tabor JJ, Bayer TS, Simpson ZB, Levy M, Ellington AD (2008) Engineering stochasticity in gene expression. *Mol Biosyst* 4: 754–761
- Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassilyev DG, Yokoyama S (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420: 43–50
- Temme K, Hill R, Segall-Shapiro TH, Moser F, Voigt CA (2012a) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res* 40: 8773–8781

- Temme K, Zhao D, Voigt CA (2012b) Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci* 109: 7085–7090
- Thompson KE, Bashor CJ, Lim WA, Keating AE (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth Biol* 1: 118–129
- Travers AA, Burgess RR (1969) Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222: 537–540
- Vassilyev DG, Sekine S, Laptenko O, Lee J, Vassilyeva MN, Borukhov S, Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417: 712–719
- Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24: 437–440
- Wassarman KM, Storz G (2000) 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* 101: 613–623
- Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* 33: 103–119
- Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* 3: 623–628



**License:** This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.