SOURCE DATA

TRANSPARENT PROCESS

OPEN ACCESS

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

Thomas H Segall-Shapiro[1], Adam J Meyer[2], Andrew D Ellington[2], Eduardo D Sontag[3] & Christopher A Voigt[1,*]

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 (σ 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 'σ 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 σ 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 'α fragment' and a null (inactivated) σ fragment, respectively. The α 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.

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 ~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).

1 Department of Biological Engineering, Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA
2 Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, USA
3 Department of Mathematics, Rutgers University, Piscataway, NJ, USA
    *Corresponding author. Tel: +1 617 324 4851; E-mail: cavoigt@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 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
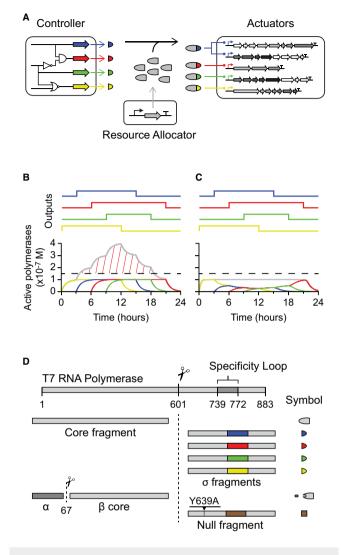

Figure 1. The resource allocator.

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 σ 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 σ 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 σ factors, which bind to the core RNAP ($\alpha 2$, β, β′, and ω 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 σ factor contains the DNA recognition domains for the −35 and −10 regions of promoters. Different σ factors bind to distinct promoter recognition sequences. In *E. coli*, there is one 'housekeeping' σ factor (σ^{70}) that is expressed at a constant level of 500–700 molecules/cell, independent of growth phase or stress, and 6 alternate σ factors that control various stress responses (e.g., heat shock) and cellular functions (e.g., flagella assembly) (Jishage *et al*, 1996). σ factors can range in size; σ^{70} is 613 amino acids and the average alternative σ is ~200 amino acids (Burton *et al*, 1981; Staroń *et al*, 2009; Rhodius *et al*, 2013). These alternative σs 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 σ 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 σ 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 σs 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 'σ 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 σ fragments, which are used to turn on different actuators. If the pool of core fragments is saturated by σ fragments, the total number of active polymerases in the system will remain constant regardless of the levels of σ 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 σ 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 'α 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.
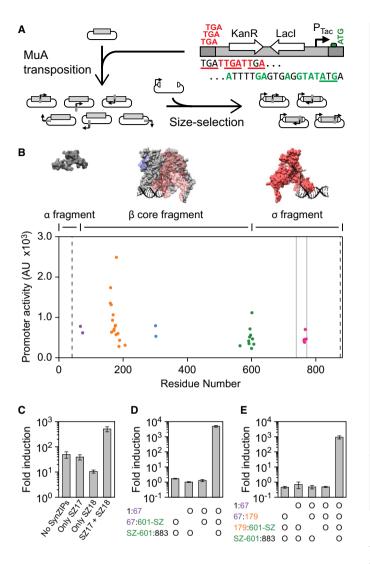
Results

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 ~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 μM 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 β core fragment shows the position of the α and σ 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 μM 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 α fragment (1:67), β core fragment (67:601-SZ), and σ fragment (SZ-601:883). An 'o' indicates the presence of a fragment in an operon that is expressed with 100 μM IPTG.

E   Data are shown for the induction of four fragments, as in (D), with an additional split of the β 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 'σ 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 'σ fragment' as it functions analogously to σ 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 α2/β/

β'/ω 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; Vassylyev *et al*, 2002; Opalka *et al*, 2010).

A simple resource allocator was built based on the core and σ fragments (Fig 3A), retaining the amino acids added by the splitposon method and the SynZIP 18 domain on the σ 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 σ 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 σ 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, σ fragments need to saturate the core fragment, causing total RNAP activity to plateau above a certain total concentration of σ fragments. The maximum level of polymerase activity is then set by the concentration of the core fragment, independent of changes in σ 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 σ 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 σ 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 σ fragments that can bind to the core fragment to change its promoter affinity. These σ fragments need to be orthogonal, that is, they cannot cross-react with each other's promoters. Initially, we attempted to base the orthogonal σ 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 σ fragments and mutated promoters. Unfortunately, of these variants, only the σ fragment containing the T3 specificity loop and corresponding promoter (Fig 3C) generated an activity comparable to that of the T7 σ fragment (Fig 3D).

The σ fragments based on the K1F and N4 specificity loops did have some residual activity. This was used as a basis to apply errorprone PCR to the σ 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 σ fragment was built based on an orthogonal T7 RNAP variant (CGG-R12-KIR) that was identified from directed evolution experiments (Ellefson *et al*,
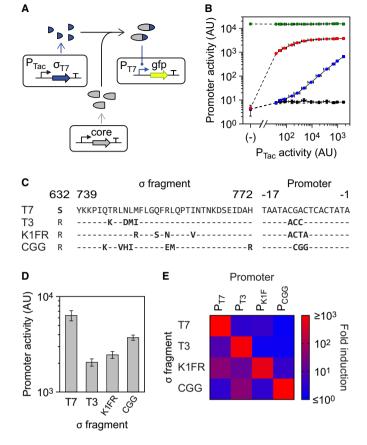
Figure 3. Activation of the core fragment via σ fragments.

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 σ 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 σ 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 μM IPTG.

C The variations between the σ 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 σ fragments are shown with their cognate promoters when expressed to saturation (100 μM IPTG) with the core fragment.

E The cross-reactivity of each σ fragment with each promoter is shown (100 μM IPTG induction of the σ 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 σ fragments (Fig 3D). In total, four σ fragment variants (T7, T3, K1FR, and CGG) and cognate promoters were built. It is noteworthy that the σ fragments only differ in sequence by 5–10 amino acids (Fig 3C). Expression of each σ 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) σ fragments (Fig 3D). The four σ 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 σ fragments that are expressed simultaneously (Fig 1C). To test this, we built a plasmid where the K1FR σ fragment is expressed from P_{Tet} and the T3 σ fragment is expressed from P_{Tac} (Fig 4A). By inducing the system with IPTG, the level of expression of the T3 σ fragment is varied while the K1FR σ 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_{KIF} were used. The activity of the σ_{T3}:P_{T3} and σ_{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 σ fragment utilizes the majority of the core fragment budget before the T3 σ fragment is induced. As the T3 σ 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 σ fragment. The sum of the P_{K1F} and P_{T3} promoter activities (gray points) remains constant and is independent of the expression of either σ 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 σ fragment and the induction of the T3 σ 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 σ 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 σ fragment is expressed to saturation with the appropriate resource allocator (Fig 4E). Assuming
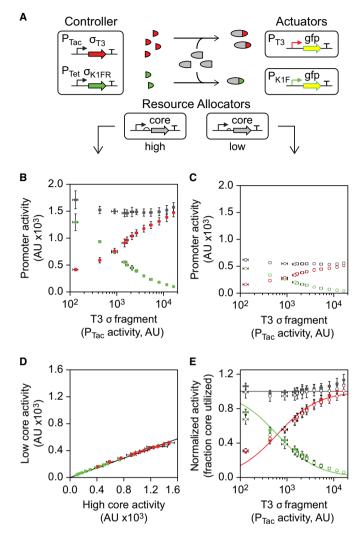

Figure 4. Competition between σ fragments to bind the core fragment.

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 σ fragment was expressed at a constant level (no induction of P_{Tet}), and the T3 σ fragment was induced with 0, 2, 4, 6.3, 7.4, 8.6, 10, 13, 16, 20, 25, and 32 μM 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 σ fragment was expressed to saturation (100 μM 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 σ 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 σ 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 σ fragments, the model predicts that the proportion of the core fragment bound by each σ fragment should depend solely on the relative expression levels of each σ fragment. The simplified model has only one parameter not measured in the normalized data set: the relative expression of the K1FR σ 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' σ 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 σ fragment, making less of it available to the other competing σ 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 σ fragment to bind and sequester the core fragment (Supplementary Table S4). Based on the screen, we identified the Y638A mutation in the CGG σ 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 σ fragments (Fig 5A). For this, the σ 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 σ 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 σ 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 'α fragment' and a larger 586 amino acid 'β core fragment' (including the SynZIP). The α 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 β core fragment that is expressed. Thus, the maximum can be set and then the α fragment used to modulate the amount that is available at any given time.

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

Coupling RNAP activity to the concentration of arbitrary α fragment tagged proteins

Since the α 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 α fragment is located) is much more tolerant to modifications (Dunn *et al*, 1988). The α 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 α 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 α 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 α 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 α 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 α fragment expression cassette. We mutagenized the RBSs and altered the promoters and start codon of the α 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 α fragment expression, we were able to achieve nearly identical activities for P_{T7} in the different plasmid contexts when they are used with the β core fragment (Fig 5H). In contrast, when the plasmids are used
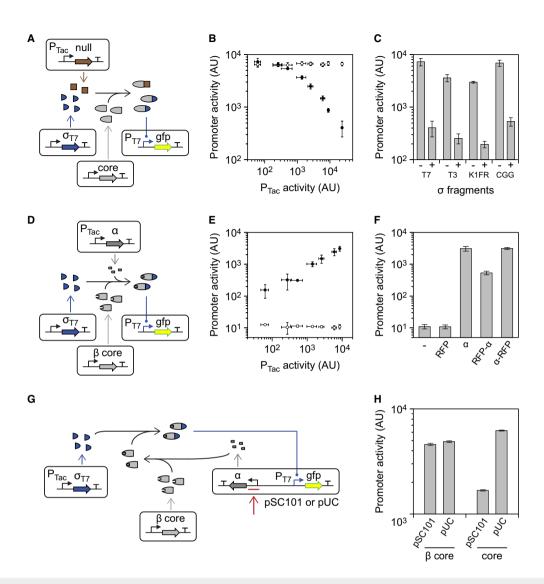
Figure 5. Positive and negative post-transcriptional regulation of the core fragment.

A   Null fragment sequestration of the core fragment.

B   The core fragment and T7 σ 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 μM 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 σ fragments. Data are shown when the null fragment is uninduced (−, 0 μM IPTG) and induced (+, 1000 μM IPTG).

D   Activation of the β core fragment through the expression of the α fragment.

E   The impact of expressing the α fragment from the P_{Tac} promoter is shown. The black and white circles show induction in the presence and absence of the α fragment cassette, respectively (from left to right: 0, 2, 4, 10, 16, 25, and 40 μM IPTG). The high level for uninduced is due to leaky expression from P_{Tac}.

F   The ability of α fragment : RFP fusions to complement the β core fragment (with the T7 σ fragment) is shown. From left to right: (−), no inducible cassette; RFP, expression of unmodified RFP; α, expression of free α fragment; RFP-α, expression of a C-terminal fusion of α fragment to RFP; α-RFP, expression of an N-terminal fusion. Each system was induced with 40 μM IPTG.

G   A genetic system is shown that uses α fragment expression from a constitutive promoter to compensate for the effects of differences in copy number. A strong constitutive promoter and RBS controlling α 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 α fragment cassettes and a P_{T7} promoter driving GFP. 'β core' indicates that the β core fragment and T7 σ fragment are co-expressed. 'core' indicates that the core fragment and T7 σ 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 α 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 α fragment from that backbone. One can also imagine harnessing feedback or feedforward loops that self-adjust the level of α 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 σ factors in native prokaryotic transcription. However, there are notable differences between our σ fragments compared to natural σ factors. First, core *E. coli* RNAP binds to DNA in a non-specific manner and this is titrated away by the σ factors (Grigorova *et al*, 2006; Bratton *et al*, 2011). It is unlikely that our T7 RNAP core fragment binds to DNA. Second, a prokaryotic σ factor only recruits the RNAP to the promoter and once transcription initiation is complete, the σ factor dissociates during transcription (Travers & Burgess, 1969; Raffaelle *et al*, 2005). Thus, the ratio of σ 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 σ fragments must remain associated with the core fragment during transcription. Third, while the size of a σ factor and the σ 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 σ factors in time as opposed to concentration (Levine *et al*, 2013). In other words, the σ factors pulse in a mutually exclusive manner to take turns fully utilizing the pool of core RNAP. In contrast, our σ fragments compete for the core fragment following mass action kinetics. This is similar to the previous understanding, where differences in σ 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 σ 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 σ 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 σ 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_{Tac} 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_{J23105} (BBa_J23105) or P_{J23109} (BBa_J23109), by a modified P_{Tet} expression system (Moon *et al*, 2012) (uninduced), and by promoters selected from libraries derived from P_{J23119} (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_{Tac} or P_{Tet} and the RBS in all constructs when titrations curves were run. mRFP1 (Campbell *et al*, 2002) and sfGFP (Pédelacq *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/).

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_{Tac} 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 miniprepped 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_{Tac} 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 miniprepped as before to obtain the bisected library. This library was electroporated into *E. coli* DH10B cells with a plasmid containing a P_{T7}-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 μl/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_{Tac}, constructs were made that express GFP after P_{Tac} 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_{Tac} 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 μl LB-M with antibiotics plus varying concentrations of IPTG in 300-μl 96-well V-bottom plates, sealed with an AeraSeal film, and grown at 37°C, 1,000 rpm for 6 h. 20 μl of each sample were added to 80 μl 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 σ fragment variants

Sections of the K1F and N4 T7 RNAP variants (Temme *et al*, 2012a) were amplified using GoTaq (Promega) in 1× GoTaq buffer plus MgCl$_2$ to a final concentration of 6.5 mM Mg^{2+}. The amplified fragments were cloned into a σ fragment expression plasmid including any necessary flanking RNAP sequence and the N-terminal SynZIP 18 domain. These mutated σ 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 σ fragments assayed for activity.

Tuning α fragment expression to compensate for copy number

An α fragment expression cassette consisting of the constitutive promoter P_{J23105}, RiboJ, and B0032 RBS driving the α 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 α 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 σ_{T7} and either core or β core fragments. Sets of pSC101 and pUC plasmids were selected that had similar levels of activity with the β 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 β core fragment was selected.

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, Umbarger 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, Krippl 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.

Ellefson 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) β-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

Grigorova IL, Phleger NJ, Mutalik VK, Gross CA (2006) Insights into transcriptional regulation and σ 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 σS (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

Iost 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

Johnsson 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

Lempiäinen 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 ςA by sporulation-specific ςF 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* σ 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

Pédelacq 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 α-complementing domain of *E. coli* β-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

Raffaelle 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

Rhodius 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 σs, anti-σs, 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) σ 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, Vassylyev DG, Yokoyama S (2002) Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature* 420: 43–50

Temme K, Hill R, Segall-Shapiro 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

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva 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

Supplementary Information for:

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

Thomas H. Segall-Shapiro, Adam J. Meyer, Andrew D. Ellington, Eduardo D. Sontag, and Christopher A. Voigt

I.   Splitposon method for bisection mapping proteins

I.A.   *Design of the splitposon*

The splitposon is based on a commercial mini-Mu transposon, the HyperMu <KAN-1> transposon (previously available from Epicentre Biotechnologies). Mini-Mu transposons are a commonly used tool in molecular biology, due to their small size and easy *in vitro* transposition protocol (Haapa *et al*, 1999). *In vitro* transposition requires only the addition of a single transposase protein, MuA, along with a linearized mini-Mu transposon. The MuA protein binds specific sequences at the termini of the transposon ('recognition ends') and catalyzes an efficient, mostly sequence-independent transposition event (Mizuuchi & Mizuuchi, 1993; Green *et al*, 2012). In contrast to the native transposon, which contains 6 unique sequences in the recognition ends (L1-L2-L3 at one terminus, R1-R2-R3 at the other), mini-Mu transposons have shorter, palindromic ends consisting of two of the native sequences (R1-R2) (Haapa *et al*, 1999).

While the R1-R2 sequence is required for transposition of a mini-Mu transposon, the sequence does not have to be perfect. The promiscuity of MuA has been studied by mutating the ends of the transposon , and a number of functional transposons with altered ends have been made. To construct the splitposon, we pooled the information from these studies to identify where the transposon could be altered and retain function. We focused on the R1 recognition sequence, since it is closest to the ends of the mini-Mu transposon, and our intention was to split proteins with as little added sequence as possible.

First, we used a consensus alignment of the six recognition sequences from the natural transposon (Goldhaber-Gordon *et al*, 2002) to determine where mutations are generally tolerated. However, it is unclear whether all of these alterations are tolerated specifically in the terminal recognition sites. Next, the R1 sequence was aligned with the L1 sequence, which is at the opposite terminus of the natural transposon. We referenced a mutational study (Lee & Harshey, 2001) to determine tolerated changes to the two bases at the end of the transposon when it is used for *in vitro* transposition reactions. Finally, we collated the mutations in previously built transposon variants. Variants with a NotI cut site insertion and a triple stop codon insertion (Poussu *et al*, 2004, 2005) have been included in commercially available kits (F-701 and F-703 from Thermo Scientific), and have high activity. In addition, transposons with two unique MlyI cut site insertions and two unique AarI cut site insertions are specified in publications (Jones, 2006; Hoeller *et al*, 2008).

A start codon was introduced into the -4 through -2 positions in the transposon. The RBS calculator (thermodynamic model v1.0) (Salis *et al*, 2009) was used to evaluate a number of potential transposon ends for strong RBS activity. One variant proved to retain sufficient transposition efficiency and effectively initiate translation at the start codon. A P_{Tac} IPTG inducible promoter system from pEXT20 (Dykxhoorn *et al*, 1996) mutated to have a symmetric LacO site ("aattgtgagcgctcacaatt") was added to the splitposon to drive expression of the C-terminal protein fragment. The constitutive LacI cassette was included so that the promoter would not drive high levels of expression when in a plasmid lacking LacI expression.
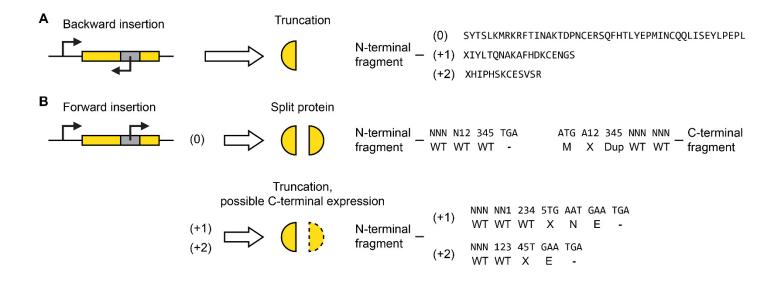
The natural mini-Mu transposon contains a stop codon in-frame with the newly engineered start codon. However, out of frame insertions can lead to many additional amino acids added to the N-terminal fragment of the split protein, potentially complicating the analysis of bisection libraries. For this reason, we mutated the terminus of the splitposon opposite from the start codon to contain three staggered stop codons (one stop codon in each frame). This modification had already been successfully made in a mini-Mu transposon end to create a transposon for generating libraries of truncated proteins (Poussu *et al*, 2005).

I.B. Library generation and characterization

The splitposon can be used to split a protein of interest with two standard cloning steps (Fig 2A). First, MuA is used to transpose the splitposon into a target insertion plasmid (Supplementary Fig S2B), which contains the region of the gene of interest to be targeted for bisection. This library is selected for the Kanamycin resistance gene in the transposon in addition to the resistance gene on the insertion plasmid. A sufficient number of colonies to achieve good coverage are plated, scraped, and harvested to yield an 'insertion library'. Second, the pooled insertion library is digested using Type IIs restriction sites flanking the region of interest. The digested library is run on a gel, and the band with size corresponding to the region of interest plus a single splitposon is excised and purified. Finally, the size-selected fragments are ligated into an expression plasmid (Supplementary Fig S2C) that has also been digested with Type IIs restriction enzymes to produce compatible overhangs. This plasmid contains an inducible expression system, as well as any flanking portions of the gene that were not in the region of interest.

The single transposition yields 6 different outcomes, depending on the orientation and position of the splitposon in the protein that is being split (Supplementary Fig S1). The splitposon can insert in either the forward or reverse direction. If it is in the reverse direction, only the N-terminal fragment of the protein is expressed, and this fragment has a number of additional bases fused to it depending on the exact insertion location. Reverse transpositions therefore, are only seen if the protein of interest can be truncated and retain function. If transposition is targeted to a region of the protein that is not sufficient for function (*i.e.*, by choosing a small enough region for the insertion plasmid), reverse insertions should have no function and will not be seen in a final selected library.

When the transposon is inserted in the forward direction, the frame of insertion determines what protein fragments will be made. MuA transposition duplicates 5 bp, leading to a few added amino acids on the protein fragments and complicating analysis. If the transposon inserts in frame with respect to the protein fragment at the 5' end of the transposon (frame 0), then a split protein will be expressed as desired. The N-terminal fragment contains no added amino acids, and the C-terminal fragment contains 3 added amino acids: M (for the start codon), a variable residue (coded for by A12, where 1 and 2 are the first two duplicated bp), and a duplicated residue (coded for by 345, the last three duplicated bases). If the transposon is inserted in frame +1 or +2, the C-terminal protein fragment is likely not to be expressed, leading to truncations that should not appear in a selected or screened library. Occasionally, the transposon may insert in frame +1 or +2 very close to an in-frame start codon, or it may create a start codon with the terminal A. In this case, out-of frame split proteins can be expressed, where the N-terminal fragment contains 2-3 variable/added residues (before the latter stop codons are encountered), and the C-terminal fragment contains duplications, insertions, or deletions based on the location of the start codon.

A Backward insertion — Truncation

N-terminal fragment —
(0) SYTSLKMRKRFTINAKTDPNCERSQFHTLYEPMINCQQLISEYLPEPL
(+1) XIYLTQNAKAFHDKCENGS
(+2) XHIPHSKCESVSR

B Forward insertion — Split protein

(0)

N-terminal fragment —
| NNN | N12 | 345 | TGA |
| WT | WT | WT | - |

| ATG | A12 | 345 | NNN | NNN |
| M | X | Dup | WT | WT |
— C-terminal fragment

Truncation, possible C-terminal expression

(+1)
(+2)

N-terminal fragment —

(+1)
| NNN | NN1 | 234 | 5TG | AAT | GAA | TGA |
| WT | WT | WT | X | N | E | - |

(+2)
| NNN | 123 | 45T | GAA | TGA |
| WT | WT | X | E | - |

Supplementary Figure S1. ***Outcomes of a splitposon library.*** (***A***) *If the splitposon inserts in the reverse direction, only the N-terminal fragment of the protein is expressed. Additionally a number of amino acids are fused to this fragment depending on the frame of insertion (as judged by protein fragment at the 5' end of the transposon). X indicates a variable residue that depends on the sequence of the insertion site. (**B**) If the splitposon inserts in the forward direction, a split protein or truncation is expressed depending on the frame of insertion. If the splitposon inserts in-frame (0), a split protein is expressed with 3 AAs added to the C-terminal fragment. The DNA sequence and encoded AAs directly flanking the splitposon are shown. For DNA (top row), Ns indicate bases in the original coding sequence of the protein, 1-5 indicates the 5 bps of DNA duplicated during MuA transposition, and other letters indicate the sequence of the splitposon. For AAs (bottom row), WT indicates a residue in the split protein, X indicates a variable residue (i.e. one coded for by bps both from the splitposon and original protein coding sequence), Dup indicates a WT residue that is present in both the N and C-terminal fragments, and other letters represent the appropriate AAs. If the splitposon inserts in the (+1) or (+2) frames, the N-terminal fragment will be expressed with a few added AAs and the C-terminal fragment may be expressed by an in-frame start codon. The residues added to the N-terminal fragment are shown in the same manner as for the (0) frame.*

4

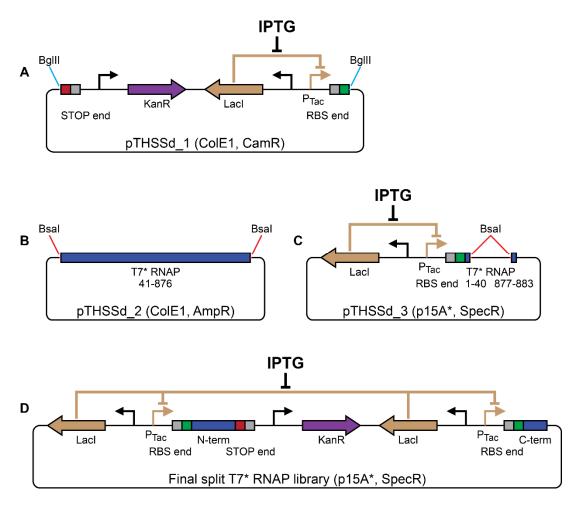## II.   Bisection mapping of T7 RNA polymerase

II.A.   Library design and statistics

To avoid seeing any truncations in the library of bisected T7* RNAP, we chose to target transpositions to a subset of the gene. Previous studies on T7 RNAP have identified the C-terminus of the gene as having a key role in catalysis and function. A version of the polymerase lacking the last two residues has been shown to lack productive polymerase activity (Mookhtiar *et al*, 1991). We excluded the last 7 residues of the gene from our library to ensure that functional truncations would not be generated. In contrast, the N-terminal region of the gene appears less sensitive to alterations. A pilot library indicated that truncations of up to 30-35 residues were tolerated, so we conservatively excluded the first 40 residues from our bisection library. Hence, the insertion plasmid contains only residues 41-876 of T7* RNAP (Supplementary Fig S2B). This section of the gene is flanked by BsaI Type IIs restriction sites for subcloning. We chose a ColEI backbone with Ampicillin resistance for the insertion plasmid. For the expression plasmid, the flanking portions of the polymerase (AAs 1-40 and 877-883) were placed downstream of the same P_{Tac} expression system that is in the splitposon (Supplementary Fig S2C). BsaI restriction sited are located between these fragments to allow seamless subcloning of the T7 RNAP* 41-876 fragment from the insertion plasmid. Based on the size of the insertion plasmid and T7 RNAP* fragment it contains we calculated the library sizes of the insertion and final libraries. Based on the number of colonies harvested for each library, sufficient coverage was achieved at each library step to achieve a high probability of sampling all possible variants (Supplementary Table S1) (Patrick *et al*, 2003).

II.B.   Library characterization

After the final split T7* RNAP library was built and harvested, it was transformed into cells containing the plasmid Nif_489 (Temme *et al*, 2012). This plasmid contains a P_{T7} driven RFP gene. Colonies were plated on selective media and 384 visually red colonies were picked (P_{Tac} is leaky enough on plates that colonies were visibly red without IPTG induction). These colonies were assayed for fluorescence in liquid media and the most active 192 selected for sequencing and further analysis. Each of the 192 selected clones was assayed four times and the mean promoter activity calculated.

The 192 active clones were each sequenced to determine the splitposon insertion location. In 180/192 clones this sequencing read gave enough information to unambiguously determine the insertion site of the splitposon. The other 12/192 clones were double splitposon insertions, other failure modes of the library, or sequencing errors, and were discarded. Of the 180 sequenced clones, 56 unique split sites were identified, with 36 in-frame and 20 out-of frame. The vast majority of the out-of-frame splits inserted in a location predicted to have a close downstream in-frame start codon, leading to a split protein. However, due to high predicted variability in the RBS strength for out-of-frame splitposon insertions, we focused on the in-frame splits for all further analysis. Information on the 192 analyzed clones is given in the source data for Fig 2B.

Supplementary Figure S2. **Plasmids used for bisection mapping of T7 RNA polymerase.** *(A) The splitposon is carried in a high copy ColE1 plasmid with chloramphenicol resistance. It is excised with BglII and purified from an agarose gel to produce the 'cleaved' linear transposon substrate for an* in vitro *transposition reaction. (B) The insertion plasmid carries the coding sequence for residues 41-876 of T7* RNAP flanked by BsaI sites on a high copy ColE1 backbone with ampicillin resistance. (C) The expression plasmid contains an inducible P$_{Tac}$ expression system and the coding sequences for residues 1-40 and 877-883 of T7* RNAP. The P$_{Tac}$ expression system and RBS are identical to those in the splitposon. (D) An example of a clone in the final bisection library. In this case, the splitposon is inserted in the forward direction into the T7* RNAP CDS. Plasmids pTHSSd_4-7, which were used to re-verify the 601 split and test the effect of adding SynZIPs (Fig 2E) look identical (plus the added SynZIPs). Both the expression plasmid and final library contain the p15A* origin and are spectinomycin resistant. Because of the splitposon, the final library is also kanamycin resistant.*
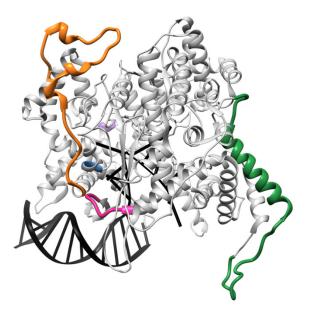
Supplementary Table S1. Statistics of T7 RNA polymerase bisection mapping.

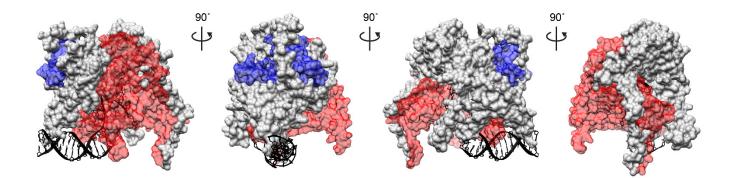| Library	Library variants[a]	Harvested colonies[b]	Library coverage[c]
Transposon Insertion	9026	7.8 x10^5	87
Final	4564	6.0 x10^5	132

a. *The number of possible variants in the insertion and final split T7* RNAP libraries. Equal to 2x the size of the insertion plasmid and 2x the size of the T7* RNAP 41-876 fragment, respectively.*
b. *The approximate number of colonies scraped and pooled for the two libraries, determined by plating dilutions and counting colonies.*
c. *The harvested clones divided by the number of variants in each library.*

II.C.   Split sites shown on the T7 RNAP structure

Supplementary Figure S3. *The five seams identified in Fig 2B are shown on the T7 RNAP transcribing initiation complex structure (*PDB# = 1QLN (Cheetham & Steitz, 1999), visualized using UCSF Chimera (Pettersen et al, 2004)) *using the same color scheme: Purple = 67-74, Orange = 160-206, Blue = 301-302, Green = 564-607, Pink = 763-770. DNA and the nascent RNA strand are shown in black.*

Supplementary Figure S4. Surface model of three-piece T7 RNAP. *A surface model of the T7 RNAP transcribing initiation complex structure (*PDB# = 1QLN, visualized using UCSF Chimera) *is shown, with the α fragment colored blue, the β core fragment colored grey, and the σ fragment colored red. The leftmost view shows transcription from left to right, and each subsequent image is rotated 90° around the y axis. DNA and the nascent RNA strand are shown in black.*

III.   Supporting experiments

III.A.   Directed evolution of the K1F and N4 σ fragments

Error-prone PCR was applied to increase the activity of σ fragments based on the K1F and N4 RNAP variants (Temme *et al*, 2012). After a visual screen for fluorescence, a number of clones with increased activity were identified for each σ fragment (Supplementary Tables S2 and S3). Nearly the full K1F σ fragment (residues 610-871 in the full-length polymerase) was mutated and screened for function. 13 highly active clones from this library were assayed and sequenced, revealing that 100% contained a point mutation affecting the residue corresponding to 750 in the full polymerase sequence. Based on these results, a variant of the K1F σ fragment was created with the M750R mutation (K1FR), which exhibits activity within 4-fold that of the T7 σ fragment and was used in all remaining experiments. The error-prone PCR protocol was applied to a smaller region of the N4 σ fragment (residues 716-789 in the full-length polymerase), and 12 improved clones were sequenced, but no sufficiently active mutations were found.

Supplementary Table S2. Improved activity clones from the K1F σ fragment ePCR library.

Clone #[a]	Mutations[b,c]		
1	**M750K**	Y746H	
2	K721R	**M750K**	
3	E694G	**M750K**	
4	**M750K**		
5	Q669R	**M750K**	
6	**M750K**		
7	Q744R	**M750V**	H772Y
8	**M750V**	H772R	E755K
9	**M750V**	H772R	E855K
10	**M750K**	Q786H	
11	E652K	**M750K**	
12	Q669R	**M750R**	
13	**M750R**	K826R	

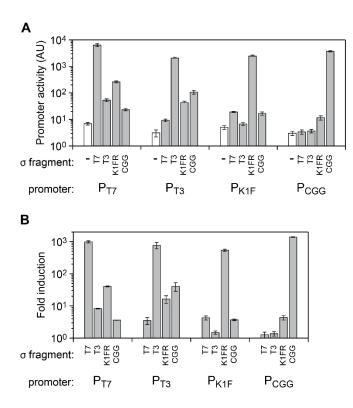a.   The clones are ordered from least to most active.
b.   Residues are numbered by their position in the full-length T7 RNAP sequence.
c.   Mutations affecting residue 750 are shown in bold.

Supplementary Table S3. Improved activity clones from the N4 σ fragment ePCR library.

Clone #[a]	Mutations[b,c]	
1	H755R	
2	H755R	
3	H755R	
4	H755R	
5	H755R	
6	H755R	
7	H755R	
8	H755R	
9	H755R	
10	V725A	H772R
11	H755R	
12	H755R	

a.   The clones are ordered from least to most active.
b.   Residues are numbered by their position in the full-length T7 RNAP sequence.
c.   Additional silent mutations were found in #1 and #4.

III.B.   Means and error underlying the σ fragment orthogonally matrix

The data used to generate the orthogonality heatmap in Fig 3E are shown with error bars (Supplementary Fig S5). The promoter activity was measured for each σ fragment with each promoter, and each promoter in the absence of a σ fragment. Dividing the level of activity with each σ fragment by the level of activity without a σ fragment yields the fold induction. This data is also available in the source data file for Fig 3D-E.

Supplementary Figure S5.   **Detailed σ fragment orthogonality results. (A)** *Each of the σ fragments and a negative control were induced with 10 µM IPTG in the presence of the core fragment and each of the four promoters. Grey bars represent promoter activity with expressed σ fragments, white bars indicate the promoter activity of negative controls with no expressed σ fragment.* **(B)** *The fold induction of each σ fragment in combination with each promoter is shown. Each bar represents the mean value of three independent assays performed on different days, with error bars showing standard deviation.*
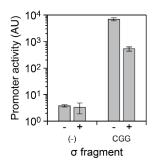
III.C. Identifying the null fragment

To determine the optimal null fragment, three known inactivating mutations (Bonner *et al*, 1992; Mookhtiar *et al*, 1991) were tested in the background of three σ fragments. The intention was to find a mutation that abolishes polymerase function without inhibiting the ability of the null fragment to compete with other σ fragments to bind the core fragment. A deletion of residues 882-3 and two point mutations, Y639A and H811A, were tested. These mutations were made to the T7 σ fragment, the CGG σ fragment, and a σ fragment based on WT T7 RNAP (rather than T7* RNAP as for the T7 σ fragment). The T7 σ fragment was expressed constitutively with the core fragment and a P_{T7} reporter plasmid, and the variant null fragments were induced with IPTG. By comparing the P_{T7} promoter activity with and without induction of the null fragments, a fold repression value was calculated for each variant (Supplementary Table S4). Based on this data, The CGG σ fragment with mutation Y639A was found to be the most active and was chosen as the null fragment. To determine whether the null fragment retains residual activity, it was expressed with the core fragment and a P_{CGG} reporter. Even at high levels of induction, this null fragment shows no P_{CGG} promoter activity when expressed with the core fragment (Supplementary Fig S6).

Supplementary Table S4. Comparison of null fragment variants.

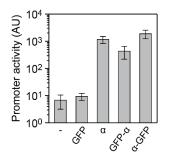| Null variant	Fold repression[a]
σ_{T7} Δ882-3	9
σ_{T7} Y639A	12
σ_{T7} H811A	11
σ_{CGG} Δ882-3	14
σ_{CGG} Y639A	18
σ_{CGG} H811A	16
σ_{T7WT} Δ882-3	12
σ_{T7WT} Y639A	13
σ_{T7WT} H811A	13

a. *Fold repression was calculated as the activity of a P_{T7} promoter with constitutive σ_{T7} expression and no null fragment induction (0 µM IPTG) divided by the activity of the P_{T7} promoter with constitutive σ_{T7} expression and high null fragment induction (1000 µM IPTG). Values are the mean fold repression from three independent assays performed on different days.*

***Supplementary Figure S6.   The null fragment lacks σ fragment activity.** The null fragment is induced from P_{Tac} with 0 (-) or 1000 µM (+) IPTG in the presence of the core fragment, a P_{CGG} reporter and either the CGG σ fragment (CGG) or no σ fragment (-). The mean promoter activity from three independent assays is shown, with error bars showing standard deviation.*

III.D. Activation of the β core fragment with proteins fused to the α fragment

We tested fusions of the α fragment to GFP for their ability to complement the β core fragment. Similar to the RFP-α fusion assays in Fig 5G-H, the GFP- α fusions were induced from P_{Tac} in the presence of constitutively expressed β core fragment and σ_{T7}. A reporter plasmid that produces RFP from a P_{T7} promoter was used to measure polymerase activity (Supplementary Fig S7).
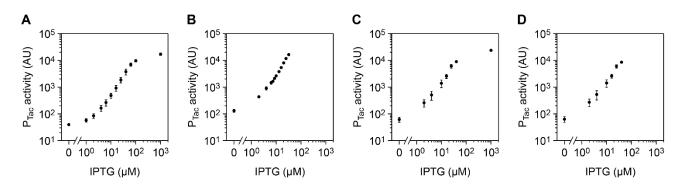

Supplementary Figure S7. *Activity of GFP : α fragment fusions. The ability of α : GFP fusions to complement constitutively expressed β core fragment and σ_{T7} is shown by the activity of a P_{T7} promoter driving RFP. (-): no inducible cassette, GFP: expression of unmodified GFP, α: expression of unmodified α fragment, GFP-α: expression of an N-terminal fusion of GFP to the α fragment, α-GFP: expression of a C-terminal fusion of GFP to the α fragment. Each system was induced with 40 µM IPTG. Each bar shows the mean level of activity from three independent assays, and error bars show the standard deviation.*
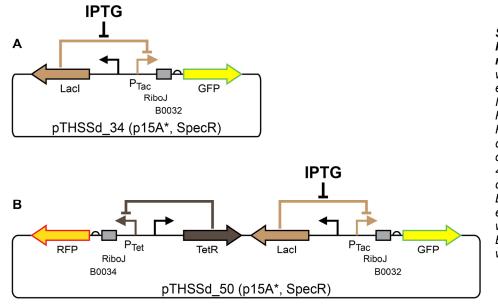
III.E. Measurement of P_{Tac} activity

In order to estimate the amount of RNAP fragments produced by our inducible plasmids, we measured GFP production from similar P_{Tac} expression plasmids (Supplementary Fig S9). The RiboJ insulator removes promoter context issues, leading to linear relationships between the expression levels of two proteins driven by identical promoters (Lou *et al*, 2012). Hence, the measured values for GFP production should linearly correlate with the RNAP fragments produced in each system.

P_{Tac} measurements were taken and plotted on the x-axis for the four the assays presented in Figs 3B, 4B,C,E, 5B, and 5E (Supplementary Fig S8). In each case, the P_{Tac} measurement was taken concurrently with the other measurements, from cells growing in the same conditions.

Supplementary Figure S8. P_{Tac} ***activity measurements.*** *Measurements of GFP production by P_{Tac} were taken under different conditions to determine relative expression levels in a number of assays. (**A**) P_{Tac} measurements for the assay in Fig 3B with plasmid pTHSSd_34. From left to right, expression was induced with 0, 1, 2, 4, 6.3, 10, 16, 25, 40, 63, 100, and 1000 µM IPTG. (**B**) P_{Tac} measurements for the assay in Fig 4B,C,E with plasmid pTHSSd_50. From left to right, expression was induced with 0, 2, 4, 6.3, 7.4, 8.6, 10, 13, 16, 20, 25, and 32 µM IPTG. (**C**) P_{Tac} measurements for the assay in Fig 5B with plasmid pTHSSd_34. From left to right, expression was induced with 0, 2, 4, 10, 16, 25, 40, and 1000 µM µM IPTG. (**D**) P_{Tac} measurements for the assay in Fig 5E with plasmid pTHSSd_34. From left to right, expression was induced with 0, 2, 4, 10, 16, 25, and 40 µM µM IPTG. For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.*
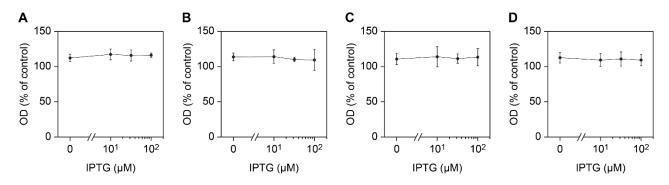

***Supplementary Figure S9. Plasmids used for P_{Tac} activity measurements. (A)** pTHSSd_34 was used to characterize P_{Tac} expression in Figs 3B, 5B, and 5E. It expresses GFP under control of P_{Tac}, with RiboJ and the B0032 RBS. (**B**) pTHSSd_50 was used to characterize P_{Tac} expression in the σ fragment competition assay (Fig 4). It expresses GFP under the control of P_{Tac} with RiboJ and the B0032 RBS. Additionally, RFP is expressed under the control of P_{Tet}, with RiboJ and the B0034 RBS. Both plasmids have a p15A* origin with spectinomycin resistance.*
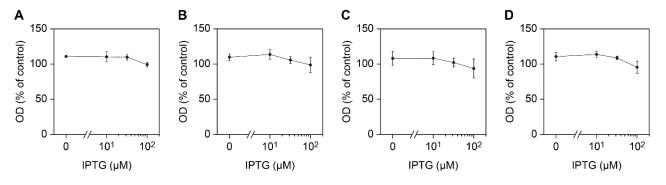
III.F. Growth impact of split polymerase expression

A number of the systems used to test split polymerase activity were measured to determine their impact on cell growth.T7 RNA polymerase is known to be toxic, especially when expressed in the presence of its promoter. Additionally, split proteins can be unstable and misfold, leading to further growth impacts.
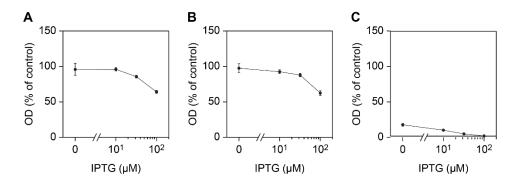
We tested three systems of split polymerase expression for growth impacts: the four orthogonal σ fragment and core fragment expression systems shown in Fig 3A (Supplementary Fig S10), the σ fragment competition systems shown in Fig 4A (Supplementary Fig S11), and the multiply-split polymerase expression systems used in Figs 2D-E (Supplementary Fig S12). The full length T7* RNAP was also tested when expressed from the same system as is used for the multiply-split polymerases (Supplementary Figs S12, S13). Each of these expression systems was induced with varying levels of IPTG and compared to a negative control containing the appropriate plasmid backbones, but not expressing the polymerase fragments or fluorescent proteins.
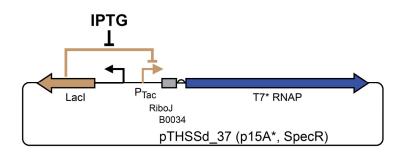

Supplementary Figure S10. **Growth impact of orthogonal split polymerase systems.** *The growth impact of the split polymerase expression systems from Figs 3B-C is shown. The four orthogonal σ fragments were expressed with IPTG (induction from left to right: 0, 10, 32, and 100 μM) in the presence of the core fragment (pTHSSd_38) with the appropriate reporter plasmid and the OD$_{600}$ after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_36, pTHSSd_43, pTHSSd_13). (A) T7 σ fragment and reporter (pTHSSd_23 and pTHSSd_8). (B) T3 σ fragment and reporter (pTHSSd_24 and pTHSSd_9). (C) K1FR σ fragment and reporter (pTHSSd_25 and pTHSSd_10). (D) CGG σ fragment and reporter (pTHSSd_26 and pTHSSd_11). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.*

Supplementary Figure S11. **Growth impact of σ fragment competition systems.** *The growth impact of the competition systems from Fig 4 is shown. The T3 σ fragment was expressed with IPTG (induction from left to right: 0, 10, 32, and 100 μM) in the presence of the K1FR σ fragment (pTHSSd_49) and high or low levels of the core fragment with either the T3 or K1FR reporter plasmid and the OD$_{600}$ after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_36, pTHSSd_43, pTHSSd_13). (A) Higher level of core fragment expression with the T3 reporter (pTHSSd_38, pTHSSd_9). (B) Higher level of core fragment expression with the K1FR reporter (pTHSSd_38, pTHSSd_10). (C) Lower level of core fragment expression with the T3 reporter (pTHSSd_39, pTHSSd_9). (D) Lower level of core fragment expression with the K1FR reporter (pTHSSd_39, pTHSSd_10). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.*

Supplementary Figure S12. ***Growth impact of highly expressed multiply split polymerase.*** *The growth impact of the three and four fragment polymerases from Figs 2D-E is shown with a full-length polymerase control. The split or full T7* polymerases were expressed with IPTG (induction from left to right: 0, 10, 32, and 100 µM) in the presence of a T7 reporter plasmid (pTHSSd_8) and the OD_{600} after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_36, pTHSSd_13). (**A**) Three piece T7* RNA polymerase (pTHSSd_14). (**B**) Four piece T7* RNA polymerase (pTHSSd_18). (**C**) Full-length T7* RNA polymerase (pTHSSd_37). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.*

Supplementary Figure S13. ***Plasmid used for full-length T7* RNAP toxicity measurement.*** *pTHSSd_37 was used to characterize the growth impact of expressing T7* RNAP from the same expression system used to drive the thee- and four-piece polymerases. It contains the full T7* RNAP CDS driven by a P_{Tac} expression system with RiboJ and the B0034 RBS.*

IV. Mathematical models

IV.A. Kinetic model of the resource allocator

We used a kinetic model to examine the contrasting outcomes on total active RNAPs in a system with or without the resource allocator (Figs 1B-C). In each case, four promoters on the controller are modeled driving either four RNAPs or σ fragments. The promoters are switched between fully off and fully on states at different time points.

In the model with expression of full-length RNAPs (Fig 1B), only two reactions per RNAP were considered, yielding one equation per RNAP:

$$\dot{r}_i = u_i - \gamma r_i \qquad i = 1 - 4 \tag{1}$$

where the dot indicates a time derivative, and:

- $r_i = r_i(t) \geq 0$ is the concentration of the ith full-length RNAP,

- u_i is the lumped transcription and translation rate of the ith RNAP, and

- γ is the degradation rate (assumed equal) of the RNAPs.

For the model involving the resource allocator (Fig 1C), a number of additions were made. A core polymerase fragment is produced at a fixed rate equal to RNAP production in the previous model, while the four promoters now drive σ fragments of the polymerase. The σ fragments can bind the core fragment to form full-length RNAP complexes which can dissociate back into σ and core fragments. Again, all degradation rates are assumed to be equal. This yields the following three equations:

$$\dot{\sigma}_i = u_i - \gamma \sigma_i + k_d r_i - k_a \sigma_i c \qquad i = 1 - 4 \tag{2}$$

$$\dot{r}_i = -\gamma r_i - k_d r_i + k_a \sigma_i c \qquad i = 1 - 4 \tag{3}$$

$$\dot{c} = v - \gamma c + k_d(\textstyle\sum_{i=1}^{k} r_i) - k_a(\textstyle\sum_{i=1}^{k} \sigma_i c) \tag{4}$$

where dots indicate time derivatives, and:

- $\sigma_i = \sigma_i(t) \geq 0$ is the concentration of (unbound) σ fragment i,

- $r_i = r_i(t) \geq 0$ is the concentration of the ith full-length RNAP complex,

- $c = c(t) \geq 0$ is the concentration of core fragment,

- u_i, v are the lumped transcription and translation rates of the ith σ fragment and the core fragment, respectively,

- γ is the degradation rate (assumed equal) of the σ fragments, full-length RNAP complexes, and the core fragment,

- k_a is association rate of the σ fragments and the core fragment (assumed equal), and

- k_d is the dissociation rate of full-length RNAP's into σ fragments and the core fragment (again assumed equal)

We simulated time courses of RNAP concentration using these systems of equations and a set of estimated parameters (Supplementary Table S5). The degradation rate, γ, was assumed to be dominated by dilution through cell growth and equal for all species in the system. The lumped transcription and translation rate of the full-length RNAPs was set to yield a steady-state concentration of 0.1 µM when they are expressed, the rate for the core fragment was set to be the same, and the rate for the σ fragments was set to yield 0.2 µM when expressed. Finally, the rates for the σ fragments binding and unbinding the core fragment were based on an *in vitro* analysis of a heterodimeric coiled-coil interaction (Chao *et al*, 1996). Simulations were performed in MATLAB using the ode45 solver.

Supplementary Table S5. Modeling parameters

Parameter	RNAP model	Resource allocator model
γ	$3x10^{-4}s^{-1}$	$3x10^{-4}s^{-1}$
$u_i(off)$	$0Ms^{-1}$	$0Ms^{-1}$
$u_i(on)$	$3x10^{-11}Ms^{-1}$	$6x10^{-11}Ms^{-1}$
v	-	$3x10^{-11}Ms^{-1}$
k_a	-	$4.5x10^{5}Ms^{-1}$
k_d	-	$2x10^{-4}s^{-1}$

IV.B. Uniqueness and stability of steady states in resource allocator model

We study the resource allocator model in (2-4), with the following changes:

- There are k different σ fragments and RNAPs rather than limiting to 4: $\quad i = 1, \ldots, k$

- The lumped transcription and translation rates, u_i and v, are assumed constant and positive.

For simplicity, we also write the system in vector form as

$$\dot{x} = f(x) \qquad\qquad (S)$$

where

$$x(t) \; = \; (\sigma_1(t), \ldots, \sigma_k(t), r_1(t), \ldots, r_k(t), c(t)) \, .$$

In general, a system of nonlinear ODE's (S) might have multiple stable states or persistent oscillations, or even exhibit chaotic behavior. It is thus of interest to show mathematically that our model has none of these, and, as a matter of fact, has the property that all solutions converge to a unique steady state, independently of initial concentrations. This is proved in the following result.

Theorem. There is a unique non-negative steady state of (S), which we will denote as

$$\bar{x} = (\bar{\sigma}_1, \ldots, \bar{\sigma}_k, \bar{r}_1, \ldots, \bar{r}_k, \bar{c}) \, .$$

Moreover, every solution of (S) with $x(t) \geq 0$ satisfies $x(t) \to \bar{x}$ as $t \to \infty$.

Proof. It is convenient to introduce, for any given solution $x(t)$, the following combinations of variables:

$$s_i(t) := \sigma_i(t) + r_i(t) \, , \quad i = 1, \ldots, k \quad \text{(total } \sigma \text{ fragment } i, \text{ bound and unbound)}$$

$$\sigma(t) := \sum_{i=1}^{k} \sigma_i(t) \, , \quad i = 1, \ldots, k \quad \text{(total unbound } \sigma \text{ fragments)}$$

$$s(t) := \sum_{i=1}^{k} s_i(t) \quad \text{(total } \sigma \text{ fragments, bound and unbound)}.$$

$$r(t) := \sum_{i=1}^{k} r_i(t) \quad \text{(total RNAP complexes, without unbound core fragments)}.$$

Observe that

$$\sigma(t) \; = \; s(t) - r(t)$$

for all t, or equivalently $s(t) = \sigma(t) + r(t)$. Since $\sigma(t) \geq 0$, it holds that

$$r(t) \leq s(t) \qquad\qquad (5)$$

for all t. We also introduce

$$R(t) := c(t) + r(t) \quad \text{(total core fragments, bound and unbound)}.$$

Since $c(t) \geq 0$, it holds that

$$r(t) \leq R(t) \qquad\qquad (6)$$

for all t.

17

For each $i \in \{1, \ldots, k\}$, we have:

$$\dot{s}_i = \dot{\sigma}_i + \dot{r}_i = u_i - \gamma\sigma_i - \gamma r_i = u_i - \gamma s_i \,.$$

Therefore, along any solution,

$$\lim_{t \to \infty} s_i(t) = \bar{s}_i := \frac{u_i}{\gamma} \tag{7}$$

and so also

$$\lim_{t \to \infty} s(t) = \bar{s} := \frac{1}{\gamma}\sum_{i=1}^{k} u_i \,. \tag{8}$$

Similarly, for R we have:

$$\dot{R} = \dot{c} + \sum_{i=1}^{k}\dot{r}_i = v - \gamma c - \gamma(\textstyle\sum_{i=1}^{k} r_i) = v - \gamma R$$

and therefore, along any solution,

$$\lim_{t \to \infty} R(t) = \bar{R} := \frac{v}{\gamma} \,. \tag{9}$$

Consider now an arbitrary steady state $\hat{x} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_k, \hat{r}_1, \ldots, \hat{r}_k, \hat{c})$. Let $\hat{s}_i := \hat{\sigma}_i + \hat{r}_i$ $(i = 1, \ldots, k)$, $\hat{\sigma} := \sum_{i=1}^{k}\hat{\sigma}_i$, $\hat{s} := \sum_{i=1}^{k}\hat{s}_i$, $\hat{r} := \sum_{i=1}^{k}\hat{r}_i$, and $\hat{R} := \hat{c} + \hat{r}$. Because of the above remarks, it must hold that $\hat{s}_i = \bar{s}_i$ $(i = 1, \ldots, k)$, $\hat{s} = \bar{s}$, and $\hat{R} = \bar{R}$.

Along any trajectory, r satisfies the following differential equation:

$$\dot{r} = -(\gamma + k_d)r + k_a\sigma c = -(\gamma + k_d)r + k_a(s - r)(R - r) \,. \tag{10}$$

Note that this is a quadratic differential equation with time-dependent coefficients (since R and s are time-dependent functions). We study its stability behavior below, but first note that, at any steady state, since $R = \bar{R}$ and $s = \bar{s}$, the steady state value $\hat{r}$ must satisfy:

$$(\gamma + k_d)\hat{r} = k_a(\bar{s} - \hat{r})(\bar{R} - \hat{r}) \,. \tag{11}$$

It is convenient to introducing the following constant, which can be thought of as an effective dissociation constant for RNAP complexes:

$$K = \frac{\gamma + k_d}{k_a} \,,$$

we can rewrite (11) as

$$K\hat{r} = (\bar{s} - \hat{r})(\bar{R} - \hat{r}) \,. \tag{12}$$

As a function of $\hat{r}$, the left-hand side of (11) is a linear function with positive slope which vanishes at zero, and the right-hand side is a parabola opening up, with roots at $\bar{R}$ and $\bar{s}$. Thus, there is exactly one solution of (11), which we call $\bar{r}$, that is less than $\max\{\bar{R}, \bar{s}\}$, and, in fact, is less than $\min\{\bar{R}, \bar{s}\}$. An explicit formula for $\bar{r}$ (not required for the proof) is:

$$\bar{r} = \frac{1}{2k_a}(B - \sqrt{D}) \quad \text{where} \quad B = \gamma + k_d + k_a\bar{R} + k_a\bar{s}, \; D = B^2 - 4k_a^2\bar{R}\bar{s} \,.$$

By (5) and (6), $r(t) \le s(t)$ and $r(t) \le R(t)$ along all solutions (including constant solutions), so certainly $\hat{r} \le \min\{\bar{R}, \bar{s}\}$, and thus $\hat{r} = \bar{r}$. Therefore

$$\hat{c} = \hat{R} - \hat{r} = \bar{c} := \bar{R} - \bar{r} \,. \tag{13}$$

18

Using that $\sigma_i = s_i - r_i$, we have, for each r_i:

$$\dot{r}_i = -(\gamma + k_d)r_i + k_a(s_i - r_i)c = k_a s_i c - (\gamma + k_d + k_a c)r_i. \tag{14}$$

So, at any steady state, since $s_i = \bar{s}_i$ and $c = \bar{c}$:

$$\hat{r}_i = \bar{r}_i := \frac{k_a \bar{s}_i \bar{c}}{\gamma + k_d + k_a \bar{c}} = \frac{\bar{s}_i \bar{c}}{K + \bar{c}}, \tag{15}$$

which is formally analogous to a Michaelis-Menten product formation law. Notice that, as a consequence of (15), $\hat{r}_i/\hat{r}_j = \bar{s}_i/\bar{s}_j$ for any two i, j, and, in view of (7),

$$\frac{\hat{r}_i}{\hat{r}_j} = \frac{u_i}{u_j} \tag{16}$$

for all $i, j \in \{1, \ldots, k\}$, which means that the RNAP complexes are produced in the same proportion as the proportion between the respective inputs. It also follows that

$$\hat{\sigma}_i = \hat{s}_i - \hat{r}_i = \bar{\sigma}_i := \bar{s}_i - \bar{r}_i. \tag{17}$$

Defining $\bar{x}$ by the formulas in (17), (15), (13), we conclude that $\hat{x} = \bar{x}$, and the steady state is indeed unique.

Next, we show that $x(t) \to \bar{x}$ as $t \to \infty$, for every solution. If we assume that $s(t)$ and $R(t)$ are already at their steady states given by (8) and (9), the differential equation (10) becomes:

$$\dot{r} = -(\gamma + k_d)r + k_a(\bar{s} - r)(\bar{R} - r). \tag{18}$$

(A justification for the assumption that R and s can be assumed to be at steady state will be given later.) The right-hand side of this ODE is the difference between the two sides in (11), and thus is positive on $0 \le r < \bar{r}$ and negative on $\bar{r} < r \le \bar{R}$. Recall that we are only interested in solutions for which $r(t) \le \bar{R}$. Therefore $r(t) \to \bar{r}$ as $t \to \infty$. Since $c(t) = R(t) - r(t)$, it follows that from the definition $\bar{c} = \bar{R} - \bar{r}$ that

$$\lim_{t \to \infty} c(t) = \bar{c}. \tag{19}$$

If we assume (justified later) that $s(t)$ and $c(t)$ are already at their steady states given by (8) and (19), the differential equation (14) becomes:

$$\dot{r}_i = k_a \bar{s}_i \bar{c} - (\gamma + k_d + k_a \bar{c})r_i. \tag{20}$$

for each $i = 1, \ldots, k$. This is a stable linear constant-coefficient differential equation, so

$$\lim_{t \to \infty} r_i(t) = \bar{r}_i \tag{21}$$

for every i. Finally, from $\sigma_i(t) = s_i(t) - r_i(t)$, the definition $\bar{\sigma}_i = \bar{s}_i - \bar{r}_i$, together with (7) and (21), we have that

$$\lim_{t \to \infty} \sigma_i(t) = \bar{\sigma}_i \tag{22}$$

for every i. We have thus proved that $x(t) \to \bar{x}$ as $t \to \infty$.

Since (16) says that $\bar{r}_i/\bar{r}_j = u_i/u_j$ for all $i, j \in \{1, \ldots, k\}$, we have then that, for any arbitrary $j \in \{1, \ldots, k\}$:

$$\bar{r} = \sum_{i=1}^{k} \bar{r}_i = \sum_{i=1}^{k} \frac{u_i}{u_j} \bar{r}_j = \frac{\sum_{i=1}^{k} u_i}{u_j} \bar{r}_j$$

or equivalently:

$$\bar{r}_j = \bar{r} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) \tag{23}$$

which means that the relative expression of the jth RNAP complex is directly proportional to the fraction of its respective control input. For example, suppose that $k = 2$, and u_1 is maintained constant. Then the expression of the second RNAP complex at steady state has the hyperbolic Michaelis-Menten form $\bar{r}_2 = \frac{Vu_2}{u_1+u_2}$, where $V = \bar{r}$.

Justification of quasi-steady state assumption

It only remains to justify the hypotheses made at two points that variables already shown to approach steady state can be replaced by their steady state values in other equations (this is sometimes called the ``CICS'' or ``convergent input to convergent state property''). One way to prove this is to appeal to the theory of asymptotically autonomous systems: we view (10) as a non-autonomous differential equation which, as $t \to \infty$, approaches the autonomous equation (18). Since this latter equation has $\bar{r}$ as a globally asymptotically stable state (for initial conditions in, for example, the interval $[0, \max\{\bar{R}, \bar{s}\}]$), it follows that solutions of (18) also approach $\bar{r}$. (See the last section in (Ryan & Sontag, 2006) for details of this technique and further references.) Similar considerations apply to the linear ODE (14) and its limit equation (20).

Simplifications when $K \ll 1$

For realistic degradation and association and dissociation constants, K is very small, typically $\approx 10^{-9}$M. In that case, the formulas for steady state values can be simplified considerably. We will assume that $v < \sum_{i=1}^{k} u_i$ (the core fragment is the limiting factor), in which case $\bar{R} = v/\gamma < (\sum_{i=1}^{k} u_i)/\gamma = \bar{s}$, and thus $\min\{\bar{R}, \bar{s}\} = \bar{R}$. When $K \approx 0$, the unique steady state value $\bar{r} \le \bar{R}$ that solves $(\bar{s} - \bar{r})(\bar{R} - \bar{r}) = K\bar{r} \approx 0$ is $\bar{r} \approx \bar{R}$. This means that (23) is, more explictly:

$$\bar{r}_j \approx \bar{R} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) = \frac{v}{\gamma} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) \tag{24}$$

It is important to note, however, that informal approximation arguments are not mathematically rigorous, and can easily lead to paradoxical conclusions. For example, (13) implies that $\bar{c} = \bar{R} - \bar{r} \approx 0$ (since we had $\bar{R} \approx \bar{r}$), and this, combined with (15) gives that $\bar{r}_i = \frac{\bar{s}_i \bar{c}}{K + \bar{c}} \approx \frac{\bar{s}_i \times 0}{K + 0} = 0$! (The fallacy in this case comes from the approximation ``$x/(K + x) \approx 0$ when $x \approx 0$'' which is false if $K \ll x$.)

To make the argument mathematically precise, let us think of the unique steady state value $\bar{r} \le \bar{R}$ that solves $K\bar{r} = (\bar{s} - \bar{r})(\bar{R} - \bar{r})$ as a function $\bar{r}(K)$, and take its limit as $K \to 0$ while keeping $\bar{R}$ and the $\bar{s}_i$'s fixed. Keeping these values fixed is valid for example if $k_a \to \infty$, or if $k_d \to 0$ and $\gamma \to 0$ at the same time that the control inputs (v and the u_i's) are proportionally increased. Using implicit differentiation, and primes to indicate derivative with respect to K, we have that $\bar{r} + K\bar{r}' = -\bar{r}'(\bar{R} - \bar{r}) - \bar{r}'(\bar{s} - \bar{r})$. Since $\bar{r} = \bar{R}$ when $K = 0$, the derivative at $K = 0$ is $\bar{r}' = \bar{R}/(\bar{R} - \bar{s})$ and thus we obtain the first-order Taylor expansion

$$\bar{r}(K) = \bar{r}(0) + \bar{r}'(0)K + o(K) = \bar{R} + \frac{\bar{R}}{\bar{R} - \bar{s}}K + o(K).$$

Then, $\bar{c} = \bar{R} - \bar{r} = \frac{\bar{R}}{\bar{s} - \bar{R}}K + o(K)$, and now substituting into $\bar{r}_j = \frac{\bar{s}_j \bar{c}}{K + \bar{c}}$, we conclude that:

$$\bar{r}_j = \frac{\bar{s}_j \bar{c}}{K + \bar{c}} = \bar{s}_j \frac{\bar{R}}{\bar{s}} + O(K) = \frac{v}{\gamma} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) + O(K),$$

which recovers (24) as $K \to 0$.

IV.C. Modeling σ fragment competition data

Using the simplified steady-state equations presented in (24), we can model the σ fragment competition data shown in Fig 4. In the context of the experiments shown in Fig 4, there are only two σ fragments, T3 and K1FR, yielding the equations:

$$\bar{r}_{T3} \approx \bar{R} \left(\frac{u_{T3}}{u_{T3} + u_{K1F}} \right) \tag{25}$$

$$\bar{r}_{K1F} \approx \bar{R} \left(\frac{u_{K1F}}{u_{T3} + u_{K1F}} \right) \tag{26}$$

If the P_{T3} and P_{K1FR} promoter activities are linearly proportional to the concentration of the appropriate RNAP complex, these equations immediately predict the result shown in Fig 4D; changing the resource allocator results in an identical linear scaling of the promoter outputs. Changing the expression of the core fragment from the resource allocator changes the value of $\bar{R}$, which linearly scales $\bar{r}_{T3}$ and $\bar{r}_{K1FR}$ identically for any constant values of u_{T3} and u_{K1FR}.

In Fig 4E, we normalize the promoter activities of P_{T3} and P_{K1FR} by the maximum promoter activities obtained when the appropriate σ fragments are expressed to saturate the core fragment. Assuming that the promoter activities are linearly proportional to the amount of corresponding RNAP present in the system, these normalized values represent the fraction of the core fragment bound by each σ fragment. That is $\bar{r}_{T3}/\bar{R}$, $\bar{r}_{K1F}/\bar{R}$, for the normalized activity values of P_{T3} and P_{K1FR}, respectively. Therefore, we have:

$$N_{T3} = \bar{r}_{T3}/\bar{R} \approx \left(\frac{u_{T3}}{u_{T3} + u_{K1F}} \right) \tag{27}$$

$$N_{K1F} = \bar{r}_{K1F}/\bar{R} \approx \left(\frac{u_{K1F}}{u_{T3} + u_{K1F}} \right) \tag{28}$$

where N_{T3} and N_{K1F} are the normalized P_{T3} and P_{K1FR} promoter activities shown in Fig 4E.

Finally, we have a relative measurement for the expression of the T3 σ fragment: the P_{Tac} expression level with the appropriate amount of inducer. Assuming that this value is linearly proportional to the true expression level of the T3 σ fragment, we can say: $u_{T3} = cP_{Tac}$, where c is a scaling factor to relate the P_{Tac} expression level to the σ_{T3} expression level. Substituting this into the model yields:

$$N_{T3} \approx \left(\frac{P_{Tac}}{P_{Tac} + \frac{u_{K1F}}{c}} \right) \tag{29}$$

$$N_{K1F} \approx \left(\frac{\frac{u_{K1F}}{c}}{P_{Tac} + \frac{u_{K1F}}{c}} \right) \tag{30}$$

As the N_{T3}, N_{K1F}, and P_{Tac} values are all measured, there is only one remaining free variable: $\frac{u_{K1F}}{c}$, which represents the constant expression level of the K1FR σ fragment in the same units as the P_{Tac} expression value. This parameter was determined by simultaneously fitting (29) and (30) to the N_{T3} and N_{K1F} data shown in Fig 4E, using a least-squares algorithm (lsqnonlin) in MATLAB. This yields a value of 617 for $\frac{u_{K1F}}{c}$. Hence, the final models shown in Fig 4E are:
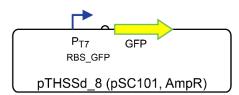
$$N_{T3} \approx \left(\frac{P_{Tac}}{P_{Tac} + 617} \right) \tag{31}$$

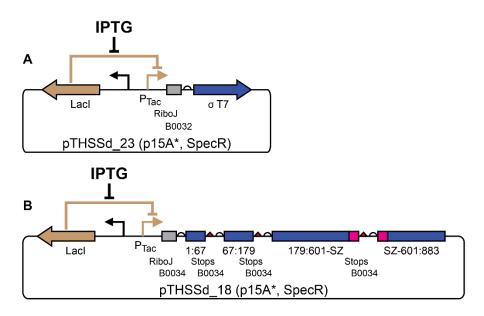$$N_{K1F} \approx \left(\frac{617}{P_{Tac} + 617} \right) \tag{32}$$

And the sum of those two equations:

$$N_{Sum} = N_{T3} + N_{K1F} \approx 1 \tag{33}$$
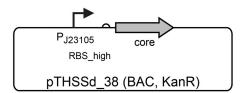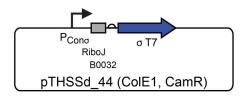
V.   Plasmid details

Supplementary Figure S14.   **Reporter plasmids.** *The reporter constructs used in this work are based on plasmid pUA66 (Zaslaver et al, 2006), which has a pSC101 origin of replication. The GFPmut2 gene is replaced with sfGFP (Pédelacq et al, 2006), and the kanamycin resistance cassette is replaced with an ampicillin resistance cassette. Variants were created with the P_{T7}, P_{T3}, P_{K1F}, and P_{CGG} promoters driving expression of GFP (pTHSSd_8-11). A strong RBS (RBS_GFP: TGTCAATTTCCGCGATAGAGGAGGTAAAG) was generated using the RBS calculator and used to control translation of GFP. For assaying GFP : α fragment fusions, a reporter variant was built with the P_{T7} mRFP1 expression cassette from NiF_489 (Temme et al, 2012) (pTHSSd_12). A negative control plasmid lacking the GFP expression cassette was also generated (pTHSSd_13).*
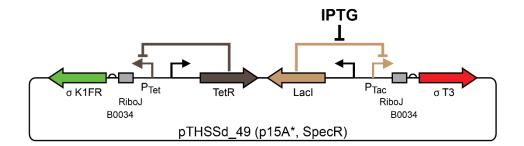

Supplementary Figure S15.   **Inducible expression plasmids.** *Plasmids for the inducible expression of genes from P_{Tac} are built from pSB3C5 (Shetty et al, 2008), which has a p15A origin. This origin appears to maintain at a higher copy number than standard, so we refer to it as p15A*. The chloramphenicol resistance cassette is replaced with a spectinomycin resistance cassette, and a modified section from pEXT20 (Dykxhoorn et al, 1996) containing a LacI expression cassette, a random spacer, and short P_{Tac} promoter is inserted into the plasmid. The lacO binding site in P_{Tac} is mutated to be symmetric (AATTGTGAGCGCTCACAATT), and is followed by RiboJ (Lou et al, 2012). (A) In most systems, only one coding sequence is expressed under the control of P_{Tac} and the B0032 RBS (BBa_B0032) is used. A number of proteins were expressed from plasmids similar to this, including σ fragments (pTHSSd_23-26), the null fragment (pTHSSd_27), the α fragment (pTHSSd_29), α : FP fusions (pTHSSd_30-33), and an RFP only control for the α : FP fusion test (pTHSSd_28). (B) To test T7* RNAP fragmented into three or four fragments, plasmids were constructed that express the fragments or a subset of them on one cistron (pTHSSd_14-22). The B0034 RBS (BBa_B0034) is used for each fragment, and a double stop codon terminates each fragment coding sequence. Two negative control plasmids were made that lack any inducible gene but contain LacI (pTHSSd_35, 36). pTHSSd_35 contains the LacI cassette and P_{Tac} promoter system found in the splitposon and bisection library, while pTHSSd_36 only contains the LacI expression cassette.*
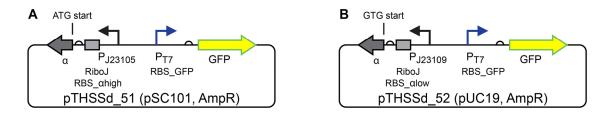
22

Supplementary Figure S16. **Core fragment expression plasmids.** *The core and β core fragments are expressed from plasmids based on pBACr-Mgr940* (Anderson *et al*, 2007) *(BBa_J61039), which carries kanamycin resistance and an F plasmid derived origin. The constitutive P_{J23105} promoter (BBa_J23105) is used to drive expression of the core fragment, core fragment variants, the full T7* RNAP, or β core fragment (pTHSSd_38-42), using different ribosome binding sites to control the strength of expression. The main RBSs used were derived from a degenerate library based on B0032: RBS_high (TACTAGAGTCATTTATGAAAGTACTAG) is used for most constructs, RBS_low: (TACTAGAGTCAGCCAAGAAAGTACTAG) is used for the lower level of core fragment expression. B0032 is used in the β core expression plasmid. A negative control of this plasmid was constructed that lacks an RBS and coding sequence (pTHSSd_43).*

Supplementary Figure S17. **Constitutive σ fragment expression plasmids.** *For the null fragment and α fragment assays, σ fragments were constitutively expressed from plasmids based on pSB1C3* (Shetty *et al*, 2008)*, which has a ColE1 origin and chloramphenicol resistance. A variant of the constitutive promoter P_{J23105} ($P_{Conσ}$: TTGACAGCTAGCTCAGTCCTAGGCTATAGGCTAG), RiboJ, and the B0032 RBS are used to drive expression of each of the four σ fragments (pTHSSd_44-47). A negative control was made that lacks any piece of the expression cassette (pTHSSd_48).*

Supplementary Figure S18. **σ fragment competition plasmids.** *A variant of the T3 σ fragment inducible expression plasmid was built to test σ fragment competition. A modified P_{Tet} expression system* (Moon *et al*, 2012) *is added behind the P_{Tac} expression system facing in the reverse direction. The P_{Tet} promoter is followed by RiboJ and drives expression of the K1FR σ fragment. Both $σ_{T3}$ and $σ_{K1FR}$ use the B0034 RBS.*

A ATG start

α P_{J23105} RiboJ RBS_αhigh P_{T7} RBS_GFP GFP

pTHSSd_51 (pSC101, AmpR)

B GTG start

α P_{J23109} RiboJ RBS_αlow P_{T7} RBS_GFP GFP

pTHSSd_52 (pUC19, AmpR)

Supplementary Figure S19. ***Reporter plasmids with α fragment compensation. (A)*** *A constitutive α fragment expression cassette is added in the reverse direction to the P_{T7} reporter plasmid before the P_{T7} promoter to make pTHSSd_51. This cassette drives production of the α fragment with P_{J23105}, RiboJ, and a RBS derived from B0032 (RBS_αhigh: TCAACCACGAAAGTACTAG). **(B)** pTHSSd_52 has the same two cassettes as pTHSSd_51, inserted into a pUC19 (Yanisch-Perron et al, 1985, 19) ampicillin resistant backbone. The α fragment cassette is changed to lower its expression level: the promoter is switched to P_{J23109} (BBa_J23109), a different RBS is used (RBS_αlow: CTAGTACTTTCGTTCATGA), and the α fragment start codon is changed to a GTG from ATG.*

Supplementary Table S6. New plasmids used in this work.

Name	Origin[a]	Marker[b]	Description
pTHSSd_1	ColE1	K/C	Splitposon in KanR ColE1 backbone
pTHSSd_2	ColE1	A	T7* RNAP 41-876 transposition target
pTHSSd_3	p15A*	S	T7* RNAP expression plasmid
pTHSSd_4	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601
pTHSSd_5	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with SZ17
pTHSSd_6	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with SZ18
pTHSSd_7	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with both SynZIPS
pTHSSd_8	pSC101	A	P_{T7} GFP reporter
pTHSSd_9	pSC101	A	P_{T3} GFP reporter
pTHSSd_10	pSC101	A	P_{K1F} GFP reporter
pTHSSd_11	pSC101	A	P_{CGG} GFP reporter
pTHSSd_12	pSC101	A	P_{T7} RFP reporter
pTHSSd_13	pSC101	A	reporter negative control
pTHSSd_14	p15A*	S	Triple split (at 67, 601-SZ)
pTHSSd_15	p15A*	S	Triple split no fragment 1:67
pTHSSd_16	p15A*	S	Triple split no fragment 67:601-SZ
pTHSSd_17	p15A*	S	Triple split no fragment SZ-601:883
pTHSSd_18	p15A*	S	Quad split (at 61, 179, 601-SZ)
pTHSSd_19	p15A*	S	Quad split no fragment 1:67
pTHSSd_20	p15A*	S	Quad split no fragment 67:179
pTHSSd_21	p15A*	S	Quad split no fragment 179:601-SZ
pTHSSd_22	p15A*	S	Quad split no fragment SZ-601:883
pTHSSd_23	p15A*	S	P_{Tac} T7 σ fragment expression
pTHSSd_24	p15A*	S	P_{Tac} T3 σ fragment expression
pTHSSd_25	p15A*	S	P_{Tac} K1FR σ fragment expression
pTHSSd_26	p15A*	S	P_{Tac} CGG σ fragment expression
pTHSSd_27	p15A*	S	P_{Tac} null fragment ($σ_{CGG}$ Y639A) expression
pTHSSd_28	p15A*	S	P_{Tac} RFP expression
pTHSSd_29	p15A*	S	P_{Tac} α fragment expression
pTHSSd_30	p15A*	S	P_{Tac} GFP-α expression
pTHSSd_31	p15A*	S	P_{Tac} α-GFP expression
pTHSSd_32	p15A*	S	P_{Tac} RFP-α expression
pTHSSd_33	p15A*	S	P_{Tac} α-RFP expression
pTHSSd_34	p15A*	S	P_{Tac} GFP expression
pTHSSd_35	p15A*	S	inducible expression negative control v1
pTHSSd_36	p15A*	S	inducible expression negative control v2
pTHSSd_37	p15A*	S	Inducible full length T7* RNAP control
pTHSSd_38	BAC	K	High core fragment expression (high resource allocator)
pTHSSd_39	BAC	K	Low core fragment expression (low resource allocator)
pTHSSd_40	BAC	K	High core fragment expression without SynZIP
pTHSSd_41	BAC	K	High full length T7 RNAP* expression
pTHSSd_42	BAC	K	β core fragment expression
pTHSSd_43	BAC	K	core fragment expression negative control
pTHSSd_44	ColE1	C	constitutive expression of T7 σ fragment
pTHSSd_45	ColE1	C	constitutive expression of T3 σ fragment
pTHSSd_46	ColE1	C	constitutive expression of K1FR σ fragment
pTHSSd_47	ColE1	C	constitutive expression of CGG σ fragment
pTHSSd_48	ColE1	C	constitutive expression negative control
pTHSSd_49	p15A*	S	P_{Tac} T3 σ fragment, pTet K1FR σ fragment expression
pTHSSd_50	p15A*	S	P_{Tac} GFP, pTet RFP expression
pTHSSd_51	pSC101	A	pSC101 α fragment compensated reporter
pTHSSd_52	pUC19	A	pUC19 α fragment compensated reporter

a. ColE1: derived from pSB1C3, p15A*: derived from pSB3C5, appears to maintain at a higher copy number than p15A, pSC101: derived from pUA66, BAC: derived from pBACr-Mgr940, pUC19: derived from pUC19.
b. A: ampicillin, K: kanamycin, C: chloramphenicol, S: spectinomycin.

VII.   References

Anderson JC, Voigt CA & Arkin AP (2007) Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3:** 133

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

Chao H, Houston Michael E., Grothe S, Kay CM, O'Connor-McCourt M, Irvin RT & Hodges RS (1996) Kinetic Study on the Formation of a de Novo Designed Heterodimeric Coiled-Coil: Use of Surface Plasmon Resonance To Monitor the Association and Dissociation of Polypeptide Chains. *Biochemistry* **35:** 12175–12185

Cheetham GM & Steitz TA (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286:** 2305–2309

Dykxhoorn DM, St. Pierre R & Linn T (1996) A set of compatible tac promoter expression vectors. *Gene* **177:** 133–136

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

Green B, Bouchier C, Fairhead C, Craig NL & Cormack BP (2012) Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob DNA* **3:** 3

Haapa S, Taira S, Heikkinen E & Savilahti H (1999) An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res* **27:** 2777–2784

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

Jones D (2006) European Patent: EP1838851 - Polypeptide mutagenesis method.

Lee I & Harshey RM (2001) Importance of the conserved CA dinucleotide at mu termini. *J Mol Biol* **314:** 433–444

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

Mizuuchi M & Mizuuchi K (1993) Target Site Selection in Transposition of Phage Mu. *Cold Spring Harb Symp Quant Biol* **58:** 515–523

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, Lou C, Tamsir A, Stanton BC & Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. *Nature* **491:** 249–253

Patrick WM, Firth AE & Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng* **16:** 451–457

Pédelacq 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

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–e104

Poussu E, Vihinen M, Paulin L & Savilahti H (2004) Probing the α-complementing domain of E. coli β-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu in vitro DNA transposition. *Proteins* **54:** 681–692

Ryan EP & Sontag ED (2006) Well-defined steady-state response does not imply CICS. *Syst Control Lett* **55:** 707–710

Salis HM, Mirsky EA & Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27:** 946–950

Shetty RP, Endy D & Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* **2:** 5

Temme K, Hill R, Segall-Shapiro TH, Moser F & Voigt CA (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res* **40:** 8773–8781

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