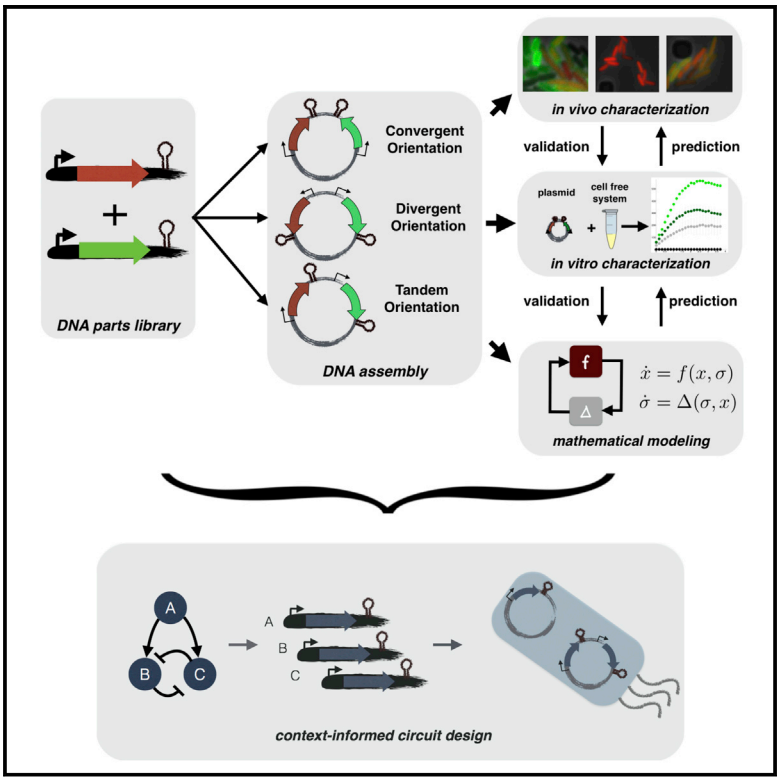
Article

Cell Systems

Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks

Graphical Abstract



Authors

Enoch Yeung, Aaron J. Dy, Kyle B. Martin, ..., James L. Beck, James J. Collins, Richard M. Murray

Correspondence

enoch.yeung@pnnl.gov

In Brief

The composition or spatial arrangement of entire genes in synthetic gene networks can introduce intergenic feedback. We use this insight to design an improved toggle switch.

Highlights

- Compositional context affects gene-induction response in synthetic gene networks
- Supercoiling models can explain the experimental effects of compositional context
- Relaxing positive supercoiling with gyrase reduces compositional context effects
- We use compositional context effects to design and build an improved toggle switch

Data Resources

KX682204	KX682217	KX682230
KX682205	KX682218	KX682231
KX682206	KX682219	KX682232
KX682207	KX682220	KX682233
KX682208	KX682221	KX682234
KX682209	KX682222	KX682235
KX682210	KX682223	KX682236
KX682211	KX682224	KX682237
KX682212	KX682225	KX682238
KX682213	KX682226	KX682239
KX682214	KX682227	
KX682215	KX682228	
KX682216	KX682229	

Yeung et al., 2017, Cell Systems 5, 11–24
July 26, 2017
http://dx.doi.org/10.1016/j.cels.2017.06.001

Cell Systems
Article

Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks

Enoch Yeung,[1,11,*] Aaron J. Dy,[3,10] Kyle B. Martin,[4] Andrew H. Ng,[5] Domitilla Del Vecchio,[6] James L. Beck,[7] James J. Collins,[3,8,9,10] and Richard M. Murray[2,4]
[1]Data Science and Analytics Group, Computational and Statistical Analytics Division, Pacific Northwest National Laboratory, Richland, WA, USA
[2]Control & Dynamical Systems, Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA
[3]Department of Biological Engineering, Institute for Medical Engineering & Science, Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA
[4]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[5]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA
[6]Department of Mechanical Engineering, Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA
[7]Department of Civil and Mechanical Engineering, California Institute of Technology, Pasadena, CA, USA
[8]Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA, USA
[9]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA
[10]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[11]Lead Contact
*Correspondence: enoch.yeung@pnnl.gov
http://dx.doi.org/10.1016/j.cels.2017.06.001

SUMMARY

Synthetic gene expression is highly sensitive to intragenic compositional context (promoter structure, spacing regions between promoter and coding sequences, and ribosome binding sites). However, much less is known about the effects of intergenic compositional context (spatial arrangement and orientation of entire genes on DNA) on expression levels in synthetic gene networks. We compare expression of induced genes arranged in convergent, divergent, or tandem orientations. Induction of convergent genes yielded up to 400% higher expression, greater ultrasensitivity, and dynamic range than divergent- or tandem-oriented genes. Orientation affects gene expression whether one or both genes are induced. We postulate that transcriptional interference in divergent and tandem genes, mediated by supercoiling, can explain differences in expression and validate this hypothesis through modeling and *in vitro* supercoiling relaxation experiments. Treatment with gyrase abrogated intergenic context effects, bringing expression levels within 30% of each other. We rebuilt the toggle switch with convergent genes, taking advantage of supercoiling effects to improve threshold detection and switch stability.

INTRODUCTION

Synthetic biological parts can be highly sensitive to context effects originating from the external environment, biological processes within the host organism, or the arrangement of genetic elements that constitute the part itself (Cardinale and Arkin, 2012). Context effects make it difficult to predict how synthetic biological parts behave when assembled into complex biological circuits. Context effects from physical composition and spatial arrangement of regulatory and genetic elements on DNA molecules are especially challenging for engineering of biological circuits, since the compositional design space is combinatorially complex. The ''compositional context'' of genes can stem from the composition of internal genetic elements, such as promoter structure, spacing regions between promoter and coding sequences, or ribosome binding sites (intragenic context), or from the spatial arrangement of entire genes within and between operons on plasmids or genomes (intergenic context).

In natural gene networks, the effects of intergenic compositional context have long been studied. Rhee et al. (1999) analyzed the expression of genes with two overlapping divergently oriented promoters, ilvY and ilvC, and found that ilvC activity is strongly activated by ilvY activity but not vice versa. Similarly, Shearwin et al. (2005) showed that gene expression from overlapping convergently oriented coliphage 186 promoters had an asymmetric effect on each other's expression. These results leave no doubt that overlapping promoters can interfere with each other's expression. But what about genes with non-overlapping promoters? This question is especially of interest since many natural operons feature divergently oriented or convergently oriented closely spaced gene pairs that do not overlap at all but exhibit correlated steady-state expression (Korbel et al., 2004).

There is evidence for a possible biophysical basis for intergenic context effects. The torsional stress introduced by unwinding DNA during transcription introduces supercoiling, which is generally accepted as a mechanism for global gene regulation in natural biological systems (Hatfield and Benham, 2002). Chong et al. (2014) showed that transient accumulation of localized positive supercoiling leads to reduction in single-gene

expression; they showed through *in vitro* transcription experiments that supercoiling could be the physical mechanism behind transcriptional bursting. Their results also suggested that the presence of nearby topological barriers such as DNA-bound proteins or transcriptional activity of neighboring genes may affect local gene expression. Thus, there is a fundamental relationship between gene regulation and compositional context in natural systems.

In contrast to work in natural systems, prior results in synthetic biology focus on the effects of intragenic compositional context. This can be considered a form of unwanted interference, prompting development of design rules for attenuating context effects (Davis et al., 2011; Kosuri et al., 2013; Mutalik et al., 2013a), or can be exploited to tune gene expression, e.g., by inducing anti-sense promoters within the expression cassette of a sense promoter to modulate sense expression levels (Brophy and Voigt, 2016). Intragenic compositional context has been used to inform synthetic biological circuits design, but the effects of intergenic compositional context are less widely explored in synthetic biology.

We investigate the impact of intergenic context on circuit design and performance. We perform a range of control experiments to identify possible sources of these effects. Specifically, we control for resource loading effects, plasmid copy-number variability, plasmid replication, terminator leakage, identity and nature of coding sequences, transcript lengths, and the stage of gene expression. We show that intergenic context creates significant differences in expression consistently across all these design variables. We then use gyrase to relax the supercoiling state of the reporter plasmids, *in vitro*, to test the hypothesis that supercoiling is responsible for the observed intergenic context effects and recapitulate the phenomenon in a mathematical model of gene expression incorporating supercoiling. We then leverage intergenic compositional context effects to design and build a new improved version of the toggle switch synthetic gene circuit.

RESULTS

Compositional Context Significantly Affects Synthetic Gene Induction

Context interference is only relevant to synthetic gene network design to the extent it alters critical processes, e.g., gene regulation or gene expression. Does intergenic compositional context affect how one gene regulates another in a synthetic gene network? If there is a significant effect, we seek to quantify it and elucidate the underlying biophysical mechanism, since gene regulation is the baseline of synthetic biocircuit design.

To systematically examine the effects of intergenic compositional context, we constructed a set of plasmids, varying gene orientation with respect to plasmid genes, gene orientation with respect to other synthetic genes on the plasmid, coding sequence identity, and the length of spacing between two non-overlapping genes. There are three relative orientations that two genes can assume: (1) convergent orientation, where transcription of both genes proceeds in opposite directions and toward each other, (2) divergent orientation, where transcription of both genes proceeds in opposite directions, away from each other and toward genetic elements on the plasmid backbone, and (3) tandem orientation, where transcription of both genes proceeds in the same direction (Liu and Wang, 1987; Shearwin et al., 2005). We constructed plasmids of each orientation to examine their effect on gene expression *in vivo* and *in vitro*.

Each plasmid incorporated two reporter genes, assembled and inserted in the same locus of a consistent vector backbone. Each gene consisted of an inducible promoter, the Lac or Tet promoter, and a corresponding fluorescent reporter, CFP (Veening et al., 2004) and RFP (Bba E1010; Zhang et al., 2002), respectively. We deliberately used a weak RBS from Mutalik et al. (2013a), for CFP, and a strong RBS (Lee et al., 2011) to ensure that any ribosome competition effects would be unidirectional (RFP loading on CFP and not vice versa) (Gyorgy et al., 2015). Thus, if both genes are induced, any differences in RFP expression would elucidate compositional context effects and not competition for translational resources. Each plasmid was transformed into MG1655Z1 *Escherichia coli*, which expresses LacI and TetR constitutively from the genome. We chose LacI and TetR since they provide independently inducible systems (Ceroni et al., 2015).

To see how compositional context altered the induction response over a range of inducer concentrations, we titrated both ispropyl-β-D-1-thiogalactopyroside (IPTG) and anhydrotetracycline (ATC), and quantified CFP and RFP expression in bulk culture plate reader experiments (Figure 1). As predicted by our choice of RBSs (using a strong RBS for RFP and a weak RBS for CFP), increases in RFP expression consistently resulted in decreased CFP expression independent of orientation. As expected, increasing CFP expression did not decrease RFP expression. We did observe that relative gene orientation affected the induction response of RFP expression to varying amounts of ATC inducer.

In the convergent orientation, we saw that the transfer curve of RFP expression exhibited strong ultrasensitivity, increasing by 120-fold in response to only an 8-fold change in ATC. At 100–200 ng/mL of ATC, RFP expression plateaued in an on state of expression, and below 25 ng/mL, RFP expression plateaued in an off state of expression. Thus, diluting ATC with an 8× dilution factor had the effect of completely switching RFP from an on to an off state.

In contrast, the tandem orientation required a 64-fold change in ATC concentration to achieve a comparable (100×) fold change in RFP expression. At 100–200 ng/mL, we also saw that RFP expression plateaued in an on state of expression (for all concentrations of IPTG tested). However, RFP reached an off state of expression only when ATC was diluted down to 3 ng/mL or lower. Thus, to achieve the same dynamic range as convergent RFP required an 8× increase in dilution factor.

Divergent-oriented RFP exhibited the smallest dynamic range. Varying ATC concentration by 200-fold produced at most a 2.7-fold change in RFP expression. Even without ATC, RFP expressed at much higher levels than background. We also saw leaky expression at the single-cell level, in both the divergent-oriented RFP and CFP MG1655Z1 strain (Figure 2).

We used a model to fit the induction response of each fluorescent protein while maximally inducing the other gene (Figures 1B and 1C). Our fits characterized induction response in terms of four parameters, leaky expression l, effective cooperativity n, maximum expression V_{max}, and half-maximal induction K_m. We

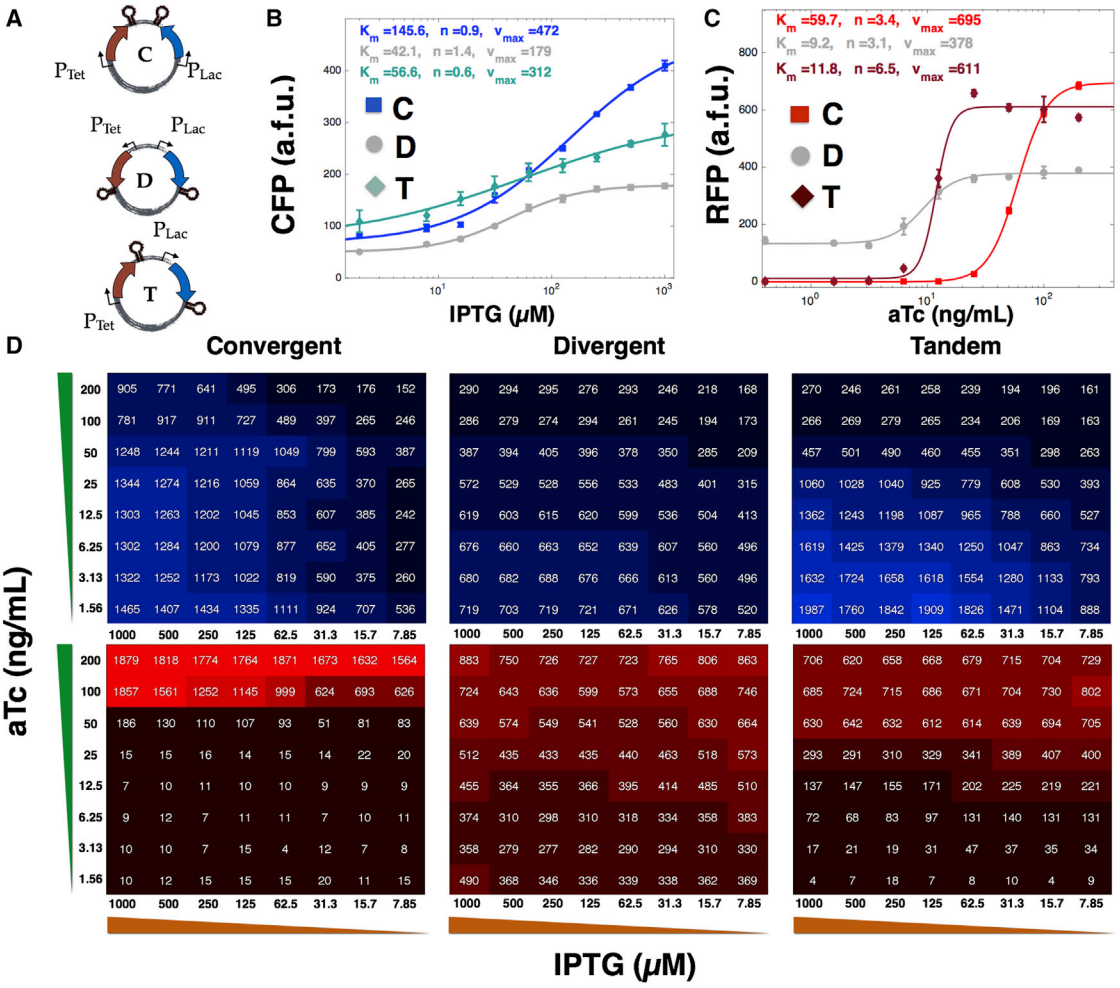


Figure 1. Ultrasensitivity, Basal Expression, and Amplitude of Gene-Induction Responses Are Significantly Affected by Compositional Context

(A) Plasmid maps of convergent-, divergent-, and tandem-oriented CFP and RFP on the ColE1 plasmid backbone.

(B) Induction curves for convergent-, divergent-, and tandem-oriented CFP fitted to a Hill function at ATC = 200 ng/mL and varying concentrations of IPTG.

(C) Induction curves for convergent-, divergent-, and tandem-oriented RFP fitted to a Hill function at IPTG = 1,000 nM and varying concentrations of ATC.

(D) Quantitative heatmaps of CFP and RFP expression in two variable titration assays of IPTG and ATC for convergent-, divergent-, and tandem-oriented ColE1 plasmids. IPTG is titrated left to right with 2× dilutions starting from 1 mM IPTG (far left), while ATC is titrated top to bottom with 2× dilutions starting from 200 ng/mL.

noticed that convergent-oriented RFP had a larger cooperativity coefficient, nearly 4-fold more than the tandem orientation and 8-fold more than the divergent orientation. Also, convergent orientation consistently fitted with the highest K_m value in both RFP and CFP induction curves, suggesting that convergent-oriented genes raises the threshold of activation.

Our experimental data thus suggest that intergenic compositional context alters key features of gene induction, such as dynamic range, maximum expression, and the activation threshold. We noted that two of these features could be explained by the nearly 4-fold increase in gene expression in convergent-oriented genes over divergent- and tandem-oriented genes. Were these compositional context effects observable with transcriptional reporters? More broadly, what was the biophysical driver of these context effects?

Compositional Context Effects Observed in Transcriptional Reporters

To understand how compositional context affects transcription, we exchanged the coding sequences for CFP and RFP with an mSpinach RNA aptamer and MG RNA aptamer, respectively. Since the mSpinach RNA aptamer is not cytotoxic, it can be used in live-cell fluorescence microscopy to explore how the temporal dynamics of gene induction varies with intergenic compositional context. After equilibrating background levels of fluorescence in mSpinach, we induced the Lac promoter with 1 mM of IPTG, thus activating expression of the mSpinach RNA aptamer. We observed that the induction response of the Lac promoter varied significantly depending on its relative gene orientation, even though the neighboring gene was never activated by ATC (Figure 3).

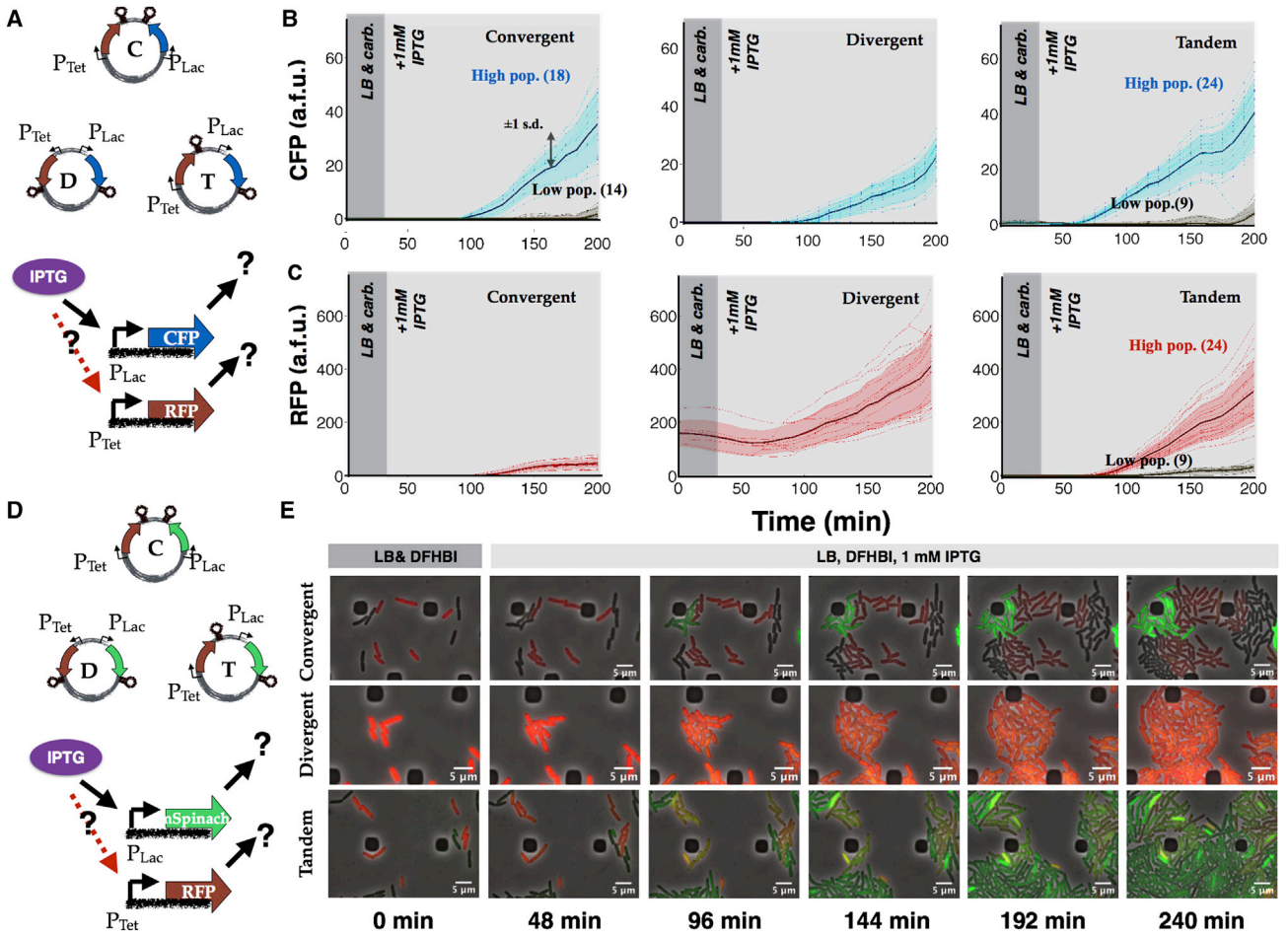


Figure 2. Consistent Compositional Context Effects Are Observed in Translated Reporters

(A) Top: Plasmid maps for convergent-, divergent-, and tandem-oriented CFP and RFP on the ColE1 plasmid backbone. Bottom: Schematic of experimental logic: a single gene is induced and the response of both genes is quantitated and compared across varying compositional constructs.

(B and C) Single-cell traces of CFP (B) and RFP (C) response to IPTG induction in the convergent, divergent, and tandem orientation. Note that both convergent orientation and tandem orientation exhibit bimodal expression phenotypes. High-expressing CFP cells also correspond to high-expressing RFP cells, while low-expressing CFP cells (gray traces) correspond to low-expressing RFP cells. Single-cell traces of RFP response to IPTG induction in the convergent, divergent, and tandem orientation. Note that both divergent orientation and tandem orientation respond to IPTG induction with significant leaky RFP expression.

(D) Top: Plasmid maps for convergent-, divergent-, and tandem-oriented mSpinach and RFP on the ColE1 plasmid backbone. Bottom: Schematic of experimental logic: a single gene is induced and the response of both genes is quantitated and compared across varying compositional constructs.

(E) Single-cell microscopy images of convergent-oriented (top) mSpinach and RFP expression, cells responded with a strong bimodal phenotype, (middle) divergent-oriented RFP and mSpinach, and (bottom) tandem-oriented RFP and mSpinach.

Convergent-oriented mSpinach expression produced a ramp-like response to IPTG induction, rising gradually over the course of 3 hr to reach a steady-state level of expression coinciding with saturation in the microfluidic chamber (Figure 3). Convergent-oriented mSpinach also gave a strong bimodal response to IPTG induction, with one group of cells achieving high levels of expression and another with low expression (Figure 3B).

In contrast, divergent-oriented mSpinach had a unimodal weak response to induction. Tandem-oriented mSpinach had bimodal expression as well, with its brightest population of cells expressing at steady-state levels comparable with the weak population in convergent orientation. The remainder of tandem-oriented mSpinach *E. coli* cells showed minimal fluorescence.

Cells with tandem-oriented mSpinach exhibited pulsatile expression, in contrast to the ramp-like response shown by convergent-oriented mSpinach. A few outlier cell traces achieved levels of mSpinach expression comparable with the bright convergent mSpinach population, but only at the peak of their transient pulses. Overall, tandem-oriented mSpinach exhibited bursty yet weaker gene expression than convergent-oriented mSpinach. Convergent mSpinach expression was approximately 4-fold brighter than divergent- and 2-fold brighter than tandem-oriented mSpinach expression, with divergent mSpinach expression being the weakest. These context effects were consistent with the trends observed in the induction study, confirming their independence from transcript identity and length.

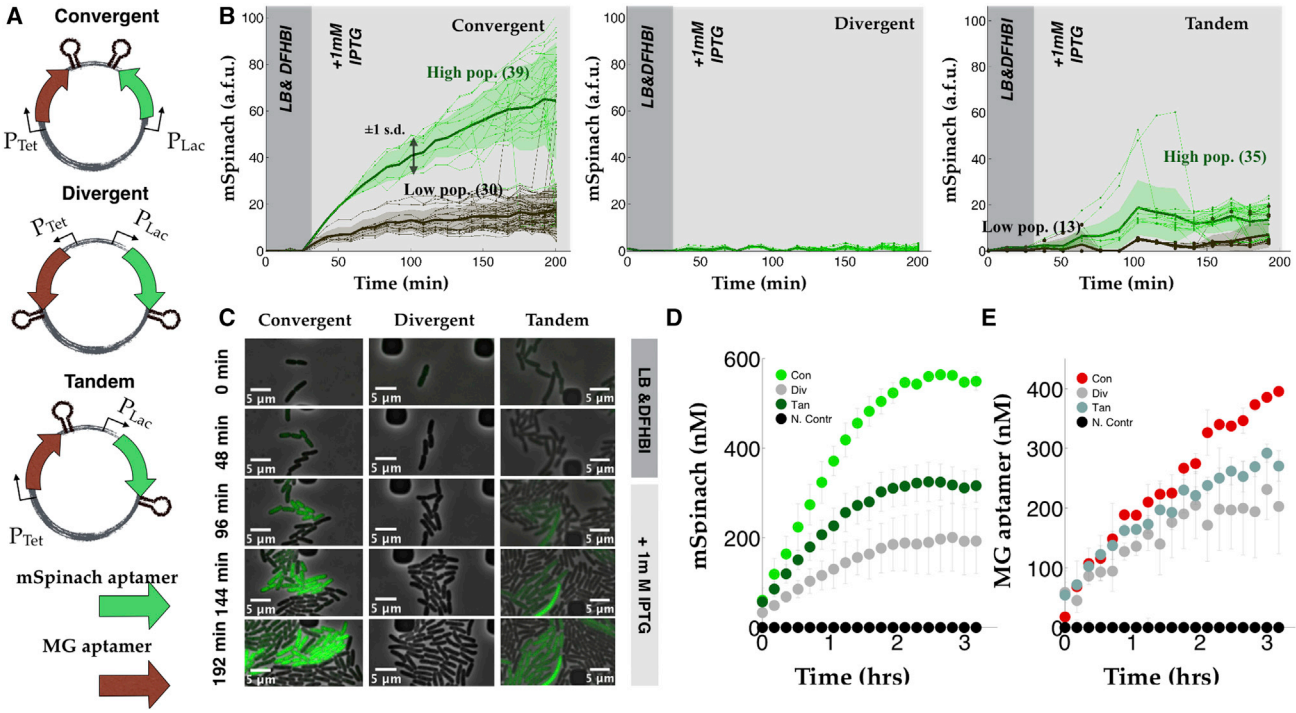


Figure 3. Compositional Context Alters Single-Cell RNA Expression Profiles

(A) Convergent-, divergent-, and tandem-oriented mSpinach and MG aptamer reporters on a ColE1 backbone.

(B) Time-lapse mSpinach expression curves for individual cell traces in response to 1 mM IPTG induction. Solid central lines within a shaded region denote the mean expression across cell lineages within a population, while the shaded area shows 1 SD from the mean.

(C) Single-cell microscopy images of convergent-, divergent-, and tandem-oriented mSpinach expression.

(D and E) Convergent-, divergent-, and tandem-oriented mSpinach and MG RNA aptamer expression in an *E. coli* cell-free expression system.

Compositional Context Introduces Regulatory Effects between Synthetic Genes

To examine whether compositional context introduced regulatory effects in the absence of translational loading (Gyorgy et al., 2015), we swapped the MG aptamer sequence with RFP. Further, by only inducing mSpinach expression, we minimize any competition for transcriptional proteins. We also swapped out a different spacer sequence between mSpinach and RFP, to see if our results were dependent on the sequence of the spacer. We then ran an identical experiment, as in Figure 3, to see how induction of mSpinach affected and correlated with RFP expression in single cells (Figure 2).

We observed the same context effects on mSpinach transcription as in Figure 3. Even with RFP in place of the MG aptamer, mSpinach expression was highest in the convergent orientation and weakest in the divergent orientation. RFP expression was extremely leaky in the divergent orientation.

In contrast, convergent-oriented mSpinach and RFP showed strong XOR logic; any cells that expressed appreciable amounts of RFP did not respond to IPTG induction with mSpinach expression, while cells that did not express any RFP showed strong mSpinach expression (Figure 2E). These results suggested that intergenic compositional context confers a form of XOR regulatory logic on nearby promoters, even when they are not targeting each other for transcriptional regulation.

We also performed the same IPTG induction experiment, analogous to experiments run for Figures 2E and 3, but using the RFP and CFP constructs. Induction of CFP with IPTG showed that mean CFP expression was again strongest in the convergent (slightly) weaker in the tandem orientation, and weakest in the divergent orientation (Figures 2A–2C). We also noted that RFP repression was tightest in convergent orientation, consistent with the higher K_m values observed in the RFP and CFP induction assay. Again, convergent-oriented CFP showed bimodal expression.

Moreover, the time-lapse traces of RFP indicated that induction of the tandem-oriented Lac promoter somehow activated the upstream RFP gene.

Removal of Neighboring Gene Abolishes Compositional Context Effects

If two synthetic genes placed in close proximity of each other experience strong intergenic context effects, what is the effect from maintenance genes on the plasmid backbone? The replication origin and resistance marker play important roles in plasmid stability and propagation; does their expression introduce additional context effects? A simple way to see if intergenic context effects were a consequence of the plasmid backbone, as opposed to neighboring synthetic genes, is to clone a single-reporter gene onto a plasmid backbone in sense or anti-sense orientation. If the orientation of the reporter gene relative to the

backbone genes impacts its expression, then the plasmid genes also are important to compositional context.

Accordingly, we extracted and cloned the single-gene CFP expression cassette (Figures 2A–2C) into the exact same plasmid locus on the ColE1 backbone, varying its relative orientation (to the backbone) as sense or anti-sense. In both sense and anti-sense plasmids, 100 bp flanking upstream and downstream sequences were retained, to eliminate any promoter sensitivity to upstream sequence perturbation. Sense and anti-sense expression of CFP showed only a small change (at most 10% difference in expression). We also constructed RFP single-gene control plasmids and observed almost exactly a 10% difference in expression. Transcribing the gene in the anti-sense direction had a small effect on gene expression, but the fold change was 40 times smaller than the fold change from varying the intergenic context of synthetic genes. In other words, the plasmid genes did not seem to confer significant intergenic context effects.

We saw a 5-fold increase in single-gene CFP expression over CFP expressed in divergent orientation with RFP (compare Figures 2A–2C and S1). This is still 25% more expression than convergent-oriented CFP, even with RFP repressed. This suggested that the very presence of another promoter with DNA binding proteins, such as a repressor, imposes a form of biophysical interference that reduces overall transcriptional activity. This may also explain the discrepancy between dramatic intergenic context effects imposed by our inducible synthetic genes and the small effects observed from varying the genetic context defined by plasmid genes.

As a final control, we transferred all convergent-, divergent-, and tandem-oriented CFP and RFP gene cassettes onto a different plasmid backbone, with a plasmid origin (from ColE1 to p15A) and resistance marker (AmpR to CmR) (see Figures S1A and S2C). The new backbone featured a different relative orientation and ordering of the replication origin and resistance marker (as well as entirely different backbone sequences). It was difficult to make an exact quantitative comparison in this case, since the variation in plasmid copy number has a direct effect on reporter expression levels. Nonetheless, we identified qualitatively similar intergenic context effects, especially in log phase. Taken as a whole, these control experiments confirmed that the plasmid backbone identity could not sufficiently explain the observed context effects.

Compositional Context Effects Are Present in the Absence of DNA Replication

To what extent was plasmid DNA replication responsible for the observed intergenic context effects? To address this question, we ran control experiments of each plasmid in a cell-free *E. coli*-derived expression system (Noireaux et al., 2003). All deoxynucleotide triphosphates are removed during preparation of cell extract, thus eliminating any confounding effects of plasmid replication. In addition, the effects of single-cell variability are eliminated, e.g., variations in LacI and TetR repressor concentration, polymerase, ribosome, and tRNA pools, allowing for a more controlled study of intergenic context effects. We prepared separate cell-free reactions for each orientation, assaying mSpinach and MG aptamer expression in a plate reader, using equimolar concentrations for each reaction (Figures 3D and 3E). Because all cell-free reactions were derived from a single batch of well-mixed extract, the variability in LacI repressor concentration was minimal.

In our transcriptional reporter plasmids, we observed that mSpinach was brightest in the convergent orientation, weakest in the divergent orientation and achieved intermediate expression in the tandem orientation. Likewise, the MG aptamer expressed strongest in the convergent orientation, weaker in the tandem orientation, and weakest in the divergent orientation. These *in vitro* outcomes were all consistent with the data from *in vivo* single-cell experiments.

We also tested RFP and CFP expression of each context variant in a cell-free expression system (Shin and Noireaux, 2012). CFP and RFP expressed strongest in convergent orientation, weaker in tandem orientation, and weakest in divergent orientation (Figure 4B). These results were consistent with results of our prior *in vitro* tests with mSpinach-MG aptamer plasmids.

A Transcriptional Model with Supercoiling States Reproduces Compositional Context Effects in Simulation

Building on the work of Chong et al. (2014), Liu and Wang (1987), and Meyer and Beslon (2014), we investigated whether supercoiling can explain the compositional context effects seen in our data. We constructed an ordinary differential equation model describing transcription and translation of both genes. To describe the interplay between gene expression and accumulation of supercoiling for each gene, we introduced separate states to keep track of promoter supercoiling and coding sequence supercoiling. This model structure allowed us to study how supercoiling buildup affected both the processes of transcription initiation and elongation (Drolet, 2006).

The kinetic rates of transcriptional initiation and transcriptional elongation are significantly affected by supercoiling buildup (Drolet, 2006). Negative supercoiling relaxes or melts the DNA double helix, facilitating transcription initiation and elongation benefits. However, excessive negative supercoiling can lead to the formation of R-loops, structural complexes that involve DNA binding to nascently produced RNA still attached to RNA polymerase. These R-loop complexes have been shown to cause transcriptional stalling (Drolet, 2006).

Conversely, positive supercoiling of DNA introduces torsional stress since positive supercoils naturally oppose the left-handed twist of DNA. Such stress leads to localized regions of tightly wound DNA that is less likely to be transcribed; positive supercoils downstream of a transcription bubble can also impose torsional resistance against further unwinding of the DNA, thereby stalling transcription. When a gene expresses and produces positive supercoiling downstream of the transcription bubble, the accumulation of positive supercoiling is especially exacerbated by the presence of a topological barrier, e.g., the binding of a DNA binding protein such as a transcription factor, or even the presence of another active gene in negatively supercoiled state. Buildup in positive supercoiling reduces the initial rate of gene transcription Chong et al. (2014). Thus, excessive supercoiling in the DNA double helix in either direction can decrease transcription rates.

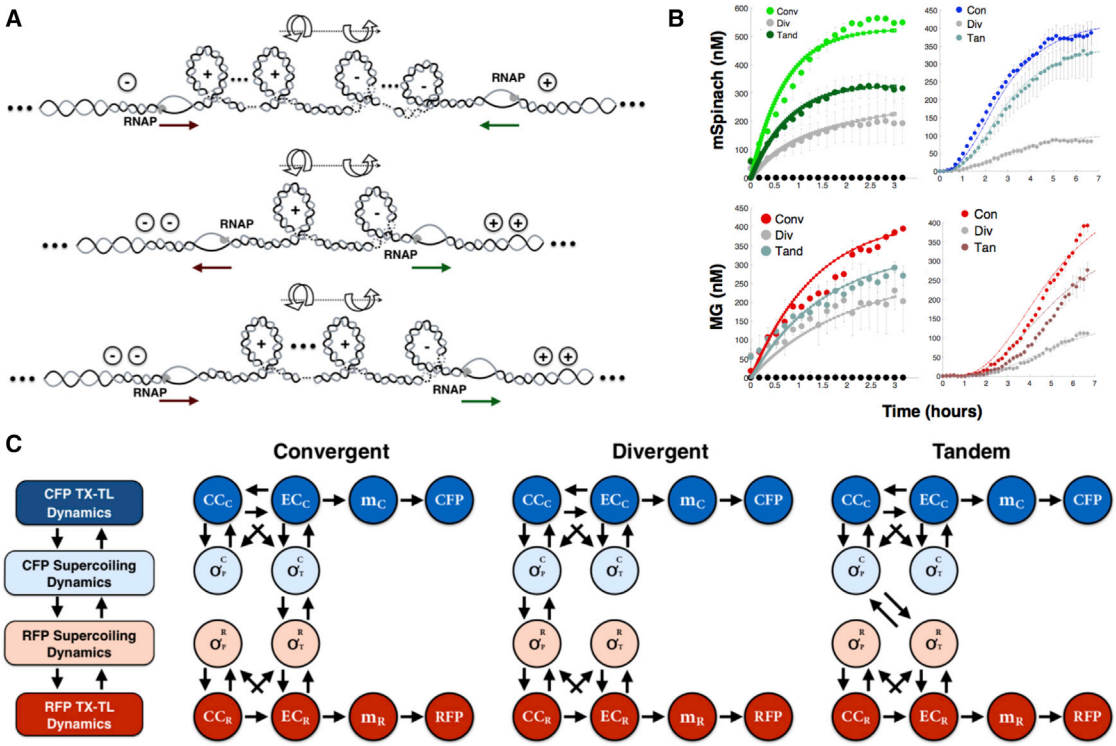


Figure 4. Mathematical Models Incorporating Supercoiling Dynamics Are Able to Recapitulate Experimental Data

(A) A diagram showing how positive supercoiling builds up downstream of transcription bubbles, and negative supercoiling builds up upstream of transcription bubbles. When two genes are adjacently placed, the intermediate region is exposed to opposing forces of torsional stress from positive (left-handed twist) and negative (right-handed twist) supercoiling. These forces do not cancel each other out, but rather oppose each other to achieve a dynamic equilibrium dependent on the transcriptional activity of nearby genes.

(B) Expression curves of a mathematical model, integrating supercoiling dynamics of promoter and transcript states with gene expression, with supercoiling parameters fit to experimental data from the TX-TL cell-free expression system (Shin and Noireaux, 2012).

(C) Schematic illustrating the state dependencies between traditional transcriptional states and supercoiling states in convergent-, divergent-, and tandem-oriented RFP and CFP.

In our model we account for the above considerations by encoding a dependency of transcription rate parameters on local supercoiling density. We model the buildup in supercoiling density as a consequence of the presence of DNA binding proteins, e.g., repressor proteins, or torsional stress from transcription of nearby genes. We build on the analysis of Meyer and Beslon (2014) and consider transcription initiation rates to be dynamically dependent on supercoiling density. We model them as Hill functions of the absolute deviation of the promoter supercoiling state from a natural supercoiling state (Rhee et al., 1999). In other words, as DNA becomes too twisted in either the positive or negative direction, transcription initiation rates and elongation rates decrease. Similarly, we suppose that the elongation rate of the gene of interest can be modeled as a Hill function of the supercoiling state over the transcript region. Thus, by modeling the dependency of transcription rates on supercoiling, we can model context-specific coupling between neighboring genes (Figure 4).

After incorporating these supercoiling hypotheses, our model was able to recapitulate compositional context trends observed in our experimental data (Figure 4B). Our simulations showed that convergent-oriented mSpinach (and CFP) is able to achieve higher levels of expression than its divergent and tandem counterparts, due to differences in supercoiling levels. These differences arise in our supercoiling model from topological barriers imposed by transcription bubbles and DNA binding proteins. Since mSpinach expresses in the anti-sense direction, its expression introduces negative supercoils upstream according to the twin-domain model by Liu and Wang (1987). In moderate amounts, negative supercoiling facilitates the unwinding of DNA and thus enhances the amount of transcriptional initiation and elongation occurring over the Lac promoter and mSpinach transcript. In the divergent and tandem orientation, mSpinach is expressed in the sense direction, which results in positive supercoiling buildup downstream of the promoter (Figure 4A). The outcome is that divergent and tandem mSpinach expression is reduced (compared with convergent), since the buildup of positive supercoiling in the presence of adjacently positioned DNA binding proteins or an active transcription bubble inhibits initiation and elongation. This effect is more severe in the divergent orientation, since excessive positive and negative supercoiling generated by initiation of the Tet and Lac promoter can interfere with each other's initiation (Figure 4C).

In exploring the parameter space of our model, we also found that gyrase (an enzyme that relaxes positive supercoiling) and

topoisomerase (an enzyme that relaxes negative supercoiling) activity are not sufficiently high to counteract the coils introduced by rapid repeated transcription events on DNA with multiple genes. These findings were consistent with the analysis of Chong et al. (2014), Liu and Wang (1987), and Meyer and Beslon (2014), which argued that buildup of transcription-induced supercoiling far outpaces the activity of supercoiling maintenance enzymes in *E. coli*. This explains why we are able to see compositional context effects both *in vivo* and *in vitro*, where gyrase and topoisomerase enzymes are presumably present and active.

Relaxing Positive Supercoiling in Plasmids Significantly Reduces Compositional Context Effects

To determine experimentally the degree to which supercoiling contributes to intergenic context, we measured the expression levels of each plasmid with or without gyrase pre-treatment in the cell-free TX-TL system (Shin and Noireaux, 2012). In the absence of gyrase, convergent orientation expressed higher than divergent and tandem orientation in both RFP and CFP channels. After gyrase treatment, tandem-oriented CFP and RFP expressed brighter than their convergent and divergent counterparts. Treating with gyrase changed the relative ordering of expression levels, as opposed to unilaterally shifting all orientations simultaneously. This suggested supercoiling as an intrinsic driver of context interference, rather than an extrinsic global factor. Also, the disparity in protein expression between the two orientations farthest apart in expression, convergent and divergent, shrunk from 300 to 100 nM (66%) for CFP, and from 500 to 180 nM (64%) for RFP. Since gyrase serves only to relax plasmids of positive supercoiling, this confirmed that supercoiling is the mechanism underlying compositional context effects.

We anticipated that treatment with gyrase would release positively supercoiled domains in the downstream region of tandem-oriented CFP and RFP, and release positive supercoiling buildup from divergently (leaky) expressed RFP (Figure 4A), and thereby reduce torsional stress in the promoters of divergently oriented CFP and RFP. Our experimental results confirmed these hypotheses, with divergent-oriented CFP and RFP increasing by more than 2-fold and tandem-oriented CFP and RFP increasing by 1.4-fold.

Gyrase treatment of convergent-oriented CFP and RFP appeared to reduce the signal slightly, by approximately 10%. This may be because convergent-oriented CFP and RFP exhibited little or no leak when uninduced (in contrast to divergent and tandem orientation); thus, the purified plasmid for convergent orientation did not have as much positive supercoiling for gyrase to mitigate. Treatments with gyrase may actually have introduced too much negative supercoiling, leading to the small drop in expression observed.

These experimental outcomes are consistent with our model of supercoiling and its impact on compositional context. Gyrase relaxes positively supercoiled domains downstream of convergent- and tandem-oriented RFP, while in the divergent orientation gyrase relaxes any positive supercoiling buildup near the promoter region. Once these positive supercoils are removed, the genes are able to express at much higher levels than prior to treatment.

Linear DNA Experiments

Since supercoiling buildup is only possible in certain scenarios, e.g., in the presence of chromosomal binding proteins, topologically constrained plasmids or linear DNA tethered to a scaffold (Chong et al., 2014), we used linear DNA to explore how orientation affects gene transcription in the absence of topological barriers. On linear DNA, divergently oriented mSpinach and MG aptamer have relatively little supercoiling buildup, since the linear ends of the DNA enable free rotation of the DNA about the helical axis as transcription occurs. On tandem-oriented DNA, MG aptamer expression can be adversely affected if mSpinach is actively transcribed, since the presence of the open complex on the mSpinach sequence may physically impede unwinding of the double helix over the MG transcript region. Expression of convergent-oriented genes on linear DNA can result in buildup of local torsional stress in between the two genes, with positive supercoiling downstream of the sense cassette and negative supercoiling downstream the anti-sense cassette (see Figure 4). In theory, the negative supercoiling could facilitate expression of the anti-sense gene, while the positive supercoiling would interfere with expression of the sense gene.

To explore these biophysical hypotheses in detail, we amplified linear DNA fragments of mSpinach and MG aptamers in convergent, divergent, and tandem orientation, and gel purified each sample. We also amplified single-gene linear DNA controls containing either mSpinach or MG aptamers. After an additional PCR purification step (to wash out any salt content from gel purification), we expressed convergent, divergent, and tandem mSpinach and MG aptamers from equimolar concentrations of linear DNA (see Figures 5D–5F).

We observed that convergent-oriented mSpinach expressed significantly higher than divergent-, tandem-oriented, or single-gene control linear DNA. The expression of both divergent- and tandem-oriented mSpinach was comparable with the mSpinach control and to each other, suggesting that, in the absence of topological barriers, differences in expression between tandem and divergently oriented mSpinach were significantly attenuated.

In contrast, the convergent-oriented MG aptamer expressed at levels comparable with the control, while the divergent-oriented MG aptamer was expressed at slightly higher concentrations. MG aptamer expression was completely shut off in the tandem orientation. This outcome further confirmed the importance of supercoiling, since no hypothesis based on transcriptional interference could explain the loss of MG aptamer expression.

Compositional Context Improves Memory and Threshold Detection in Toggle Switch

Synthetic gene networks, for the most part, have been designed primarily to avoid one type of compositional context effect: terminator leakage. Terminator leakage can cause positive correlation between a downstream gene with an upstream gene. While this is a noteworthy consideration in designing synthetic gene networks, we can actually utilize compositional context to improve or reinforce the feedback architecture of synthetic gene networks.

The toggle switch provides an excellent case study of how an informed understanding of compositional context can improve design. Being one of the first synthetic biocircuits ever made, it

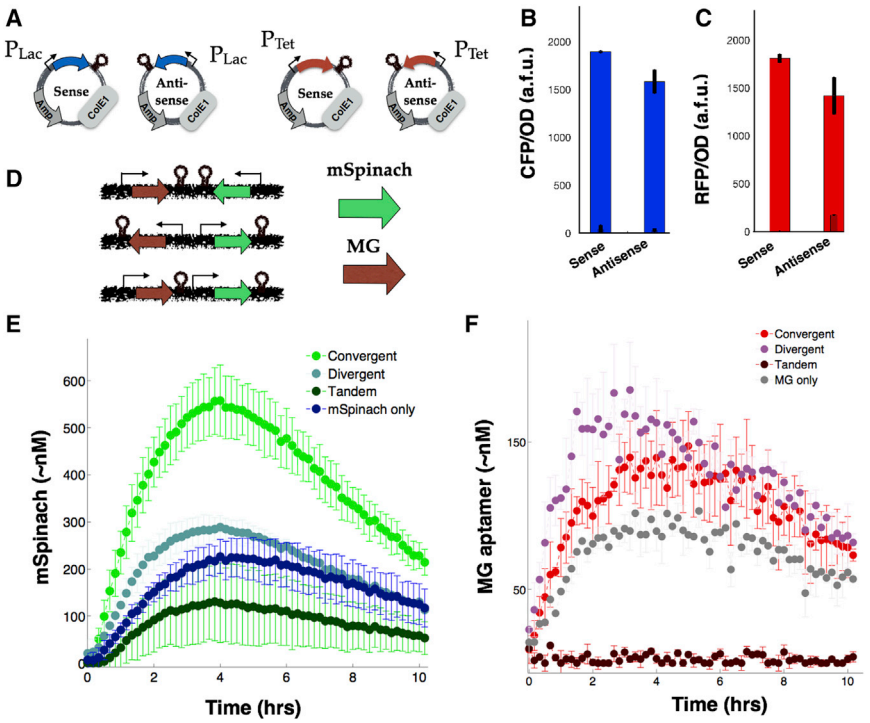


Figure 5. Compositional Context Effects Are Abrogated on Single-Gene Control Plasmids yet Persist on Linear DNA

(A) A diagram showing the sense and anti-sense CFP and RFP single-gene cassette controls, expressed on the ColE1 backbone.

(B) Expression at t = 550 min for CFP expressed from sense and anti-sense control ColE1 plasmids.

(C) Expression at t = 550 min for RFP expressed from sense and anti-sense control ColE1 plasmids.

(D) Diagram of linear DNA fragments with mSpinach and MG RNA aptamers in convergent, divergent, and tandem orientation.

(E) Cell-free *in vitro* expression of equimolar concentrations of linear mSpinach in convergent, divergent, and tandem orientation, and as a single gene on a linear DNA.

(F) Cell-free *in vitro* expression of equimolar concentrations of a linear MG RNA aptamer in convergent, divergent, and tandem orientation, and as a single gene on linear DNA.

was constructed in divergent orientation to avoid terminator leakage effects between two mutually repressing genes, LacI and TetR (Gardner et al., 2000; Kobayashi et al., 2004). From the perspective of protein regulation, two proteins, LacI and TetR, enforce mutual repression by transcriptional repression.

However, we can also build the toggle switch in convergent or tandem orientation. The convergent toggle switch is most appealing, based on several experimental insights: (1) the competing dynamics of positive and negative supercoils between the two genes encodes an additional layer of mutual negative feedback (Figure 6A), (2) the co-expression profiles of RFP and CFP in the convergent orientation (Figure 1D) and mSpinach and RFP in the convergent orientation (Figure 2E) were strongly anti-correlated. All of these properties of compositional context have the potential to enhance or strengthen the existing mutual negative feedback in the toggle switch.

Since our previous controls of sense and anti-sense encoded single genes showed that changing orientation of a single gene on a backbone did not affect expression more than 15%, we thus constructed a two-plasmid version of the toggle switch, with LacI and TetR expressed on separate plasmids. This "context-free" version of the toggle acted as a reference for how a toggle switch should function independent of genetic context.

In both versions of the toggle switch, each gene cassette in the toggle switch was bicistronic, with LacI reported by translation of RFP and TetR reported by GFP. We used stronger ribosome binding sites to express LacI and TetR and weak BCDs to express the downstream reporters.

This was done to minimize any ribosomal loading effects from reporter translation and, again, to show that even a toggle switch built *de novo* from existing synthetic biological parts with different CDSs, promoters, and RBSs could utilize compositional context (Figure 6). We also tested the original Gardner-Collins toggle switch, comparing performance in the original orientation to a convergent variant (see Figure S2).

We first tested the ability of the toggle switch to act as a threshold detector. In theory, the phase portrait of a toggle switch consists of two locally asymptotically stable equilibrium points and a separatrix, which drives state trajectories into the basin of attraction of one of the equilibrium points (Gardner et al., 2000). As a proxy for varying the amount of actively repressing TetR and LacI, we simultaneously varied the concentration of inducers ATC and IPTG, thereby allowing us to attenuate the activity of LacI and TetR repression independently. Most notably, when the toggle switch was configured in the convergent orientation, it exhibited much sharper XOR logic and separation between high-GFP-low-RFP states and high-RFP-low-GFP states compared with its two-plasmid (and divergent) counterpart (Figures 6, S2A, and S2B).

Our second experimental test was a stability test of the memory properties of the toggle (Figures S2C–S2E). We found that in both strains tested, MG1655ΔLacI and MG1655 *E. coli,* there were significant differences between the convergent and divergent toggles. In particular, we observed cells transformed with the divergent toggle tended to drift from its high-GFP state into a lower-GFP state over time, while the convergent toggle tended to maintain a high-GFP state throughout the course of the entire 53 hr experiment (48 hr post-latching and 5 hr of latching).

From our data, we see that the two-plasmid toggle exhibits weaker thresholding in two specific parameter regimes: when IPTG and ATC are both present in high concentrations and when IPTG and ATC are both present in low concentrations. When both inducers are present in high concentrations, the majority of Lac and Tet promoters are unrepressed because most repressor proteins are sequestered by inducers, leading

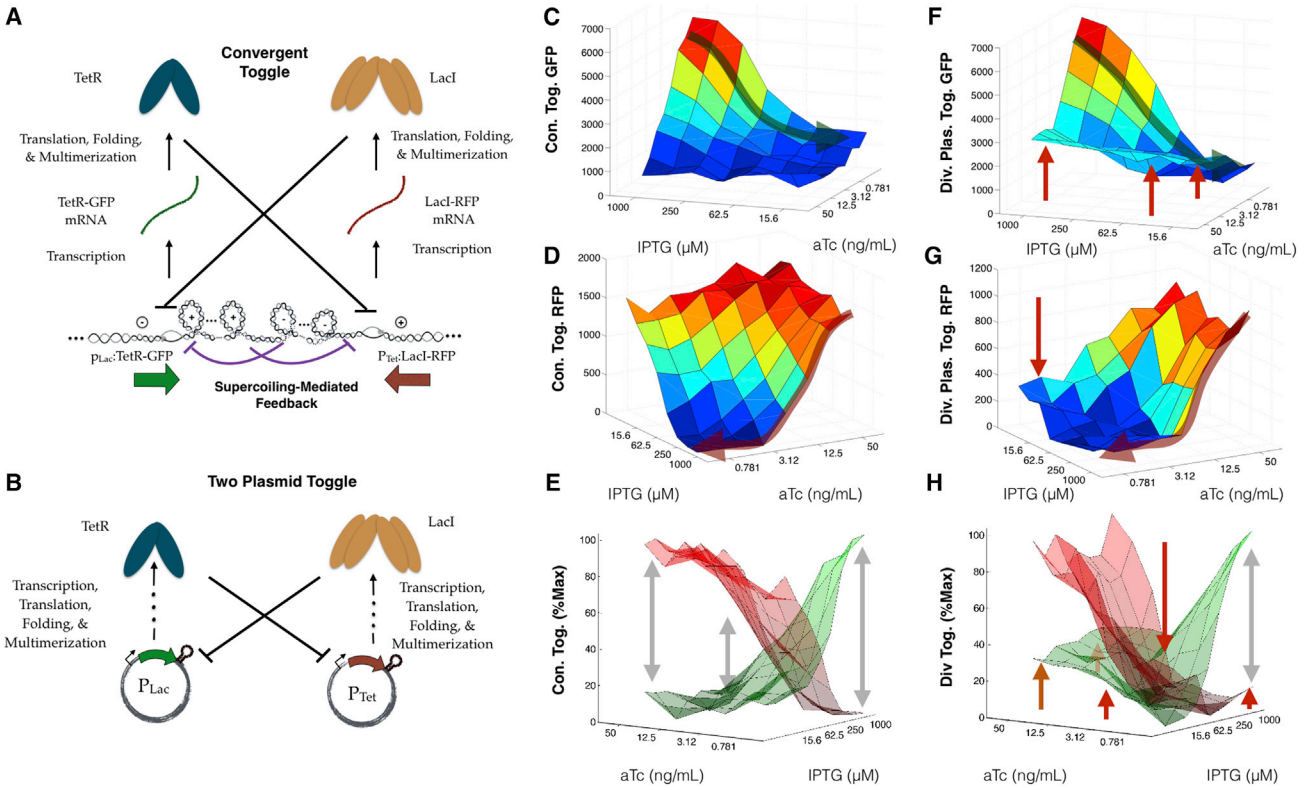


Figure 6. Compositional Context Can Be Used to Introduce Supercoiling-Mediated Feedback, Improving the Sharpness of the Threshold in the Toggle Switch

(A) Diagram of feedback architecture in a convergent toggle switch.

(B) Diagram of feedback architecture in a convergent toggle switch TetR-GFP (ColE1) and LacI-RFP (p15A).

(C–E) Experimental data of convergent toggle GFP expression in response to titrating IPTG and ATC concentrations. Single-headed arrows (red and green) highlight the sensitivity of the induction response, while gray double-headed arrows highlight the separation between on and off states in the convergent toggle.

(F–H) Experimental data of the two-plasmid toggle GFP expression in response to titrating IPTG and ATC concentrations. The green and red curved arrows indicates the sensitivity of the induction response. Red straight arrows indicate points on the response surface that deviate from ideal toggle behavior, e.g., leaky expression in an off-state. The gray double-headed arrow highlight the separation between on and off states in the divergent toggle.

to weak feedback. The weak feedback makes it difficult to differentiate which inducer is higher, since all promoters are essentially expressing constitutively (Figures 6F–6H). When both inducers are present in low concentrations, both promoters are strongly repressed making it difficult for one promoter to gain a dominant foothold over the other sufficient to produce fluorescent signal. Thus, in the low-inducer concentration regime, even if one inducer is higher in concentration than the other, neither gene is strong enough to repress the other to the point of producing detectable fluorescence (Figures 6F–6H).

On the other hand, the convergent toggle shows clear separation between high-GFP-low-RFP states and low-GFP-high-RFP states in both of these parameter regimes.

DISCUSSION

Our findings show that compositional context significantly alters gene expression in synthetic gene networks. When appropriately harnessed, compositional context can be used to strengthen or enhance existing feedback loops in the intended biocircuit design.

The results of our experimental studies emphasize the importance of understanding intergenic compositional context effects, i.e., composition of entire genes. We have seen that compositional context effects can cause variations of 3- to 4-fold of the same gene (promoter, RBS, coding sequence, etc.) simply by rearranging its orientation and the orientation of other neighboring genes.

Our analysis considered supercoiling as the physical basis for generating expression differences. In past work, the primary context effects considered in designing synthetic biocircuits are the effects of terminator leakage and transcriptional interference from overlapping promoter and RBS elements. We claim that supercoiling is the dominant source of compositional context effects observed in our data.

We see consistent differences in gene expression, even when only one gene is induced and the other remains repressed. If terminator leakage and transcriptional interference were the source of compositional context effects, we would not expect to see any effects in the case of single-gene induction.

However, there is more than a 2-fold difference in expression between divergent-expressed CFP and its single-reporter

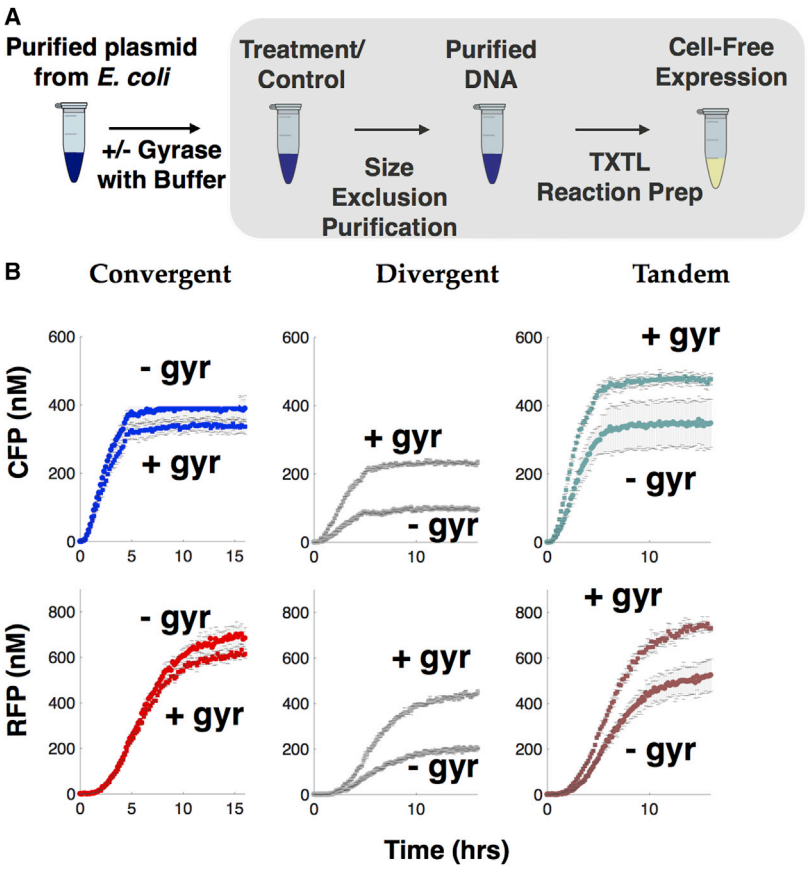


Figure 7. Relaxation of Positive Supercoiling in Plasmids with Gyrase Enzyme Significantly Reduces Compositional Context Effects on Gene Expression

(A) Workflow for gyrase treatment experiments.

(B) Expression of CFP and RFP for convergent-, divergent-, and tandem-oriented ColE1 plasmids pre- (small dots) and post-treatment (large dots) with gyrase.

counterpart (for sense or anti-sense, compare Figures 5A–5C and 1C). The physical presence of a neighboring gene has an effect, even if it is not transcriptionally active. Thus, transcriptional interference via terminator leakage does not explain the data.

If transcriptional interference were the primary driver for context effects, we would expect convergent-oriented genes to achieve far weaker levels of expression than divergent or tandem orientation. In theory, transcribing polymerases that managed to leak through two terminators—Larson et al. (2008) characterized the termination efficiency of our terminators at 98%—would collide in the convergent orientation, leading to an increase in abortive transcription events or transcriptional stalling.

Admittedly, we see from our *in vivo* characterizations that early log phase CFP and RFP expression is weaker in the convergent orientation than the divergent and tandem orientation. The fly in the ointment for this argument is that both CFP and RFP expression are higher (see Figures 1, S1A, S1C, and S1D) in the doubly induced case than in the singly induced case in early log phase, which contradicts the predictions of transcriptional interference theory.

Transcriptional interference also does not account for the sudden rise in convergent-oriented expression relative to divergent orientation as cells approach the end of their exponential growth phase. Supercoiling theory, on the other hand, predicts that, as gyrase activity wanes, the promoter regions of divergently oriented genes become more positively supercoiled, which inhibits their activity. This positive supercoiling originates from the RFP promoter as it transcribes in the anti-sense direction, thus asymmetrically inhibiting CFP expression and favoring RFP expression (see Figures 1D and 2B).

Consider the differences in expression *in vitro* of convergent-, divergent-, and tandem-transcribed RFP and CFP (Figure 7). Our experimental characterizations *in vitro* control for variations for plasmid copy number (as a function of orientation), since plasmid replication does not occur in the TX-TL cell-free system (Noireaux et al., 2003). Nonetheless, we see that levels of CFP (and RFP) expression in the convergent and divergent orientation differ by nearly 300 nM (and 500 nM) when purified directly from cells in their natural supercoiled state, whereas treating with gyrase to eliminate positive supercoiling decreases the difference by nearly 70% in both genes. Relaxing positive supercoiling in divergently oriented RFP and CFP with gyrase also allows expression levels comparable with tandem orientation prior to gyrase treatment, while treating tandem-oriented RFP and CFP enables expression levels higher than both post- and pre-treatment convergent-oriented CFP and RFP. Taken as a whole, these observations confirm that the dominant physical process driving the effects of compositional context is supercoiling.

The improved performance of the toggle switch (Figures 6A and 6B) can also be explained by examining the effects of supercoiling and compositional context. Suppose, for illustration, that LacI-RFP is slightly more induced than TetR-GFP. The positive supercoiling from TetR-GFP expression propagates downstream to meet the negative supercoils generated from transcription of LacI-mRFP. As more and more LacI-mRFP expresses, it forces the positive supercoils back into the TetR-GFP coding sequence. When this happens, TetR-GFP is no longer able to express, and its transcript region is thus available to LacI-RFP as a downstream region for dissipating its internal torsional stress. Thus, by propagating supercoils into its neighboring gene, LacI-RFP exerts a form of negative feedback independent of transcription factor-mediated repression.

This explains why the convergent toggle is able to function in regimes where IPTG and ATC are simultaneously high or low. When IPTG and ATC are both present in high concentrations, the attenuation of transcription factor repression is compensated by the presence of supercoiling-mediated repression. Thus, even though LacI and TetR are not as effective in repressing their respective promoters, the extra layer of feedback

allows the convergent toggle to decide on a dominant state (LacI-RFP). Similarly, in the low-parameter regime, even though both repressors are strong, the additional feedback from supercoiling favors one gene or the other (an enhancement of the winner-takes-all or XOR logic) and evidently improves the ability of the toggle switch to again allow LacI-RFP to dominate over TetR-GFP. Thus, there is a multi-layer feedback effect introduced by supercoiling in the convergent orientation, conformal with the intended feedback architecture of the toggle switch.

In the divergent orientation, the supercoiling propagated by proximal promoters generally results in decreased expression of the repressors. This results in weaker repression, since both promoters are affected by the presence of supercoiling (reference the results in Figures 1, 2, and 3). This leads to overall reduction in reporter signal, but also leaky repression. As time transpires post-induction (growing in medium without inducer), this allows the population of cells to drift from the high-GFP state.

In the convergent orientation, once TetR-GFP is expressed in the high state, it continuously dominates by propagating supercoils into the LacI-RFP transcript region. These supercoils may impose a higher activation threshold for the LacI-RFP state, thus keeping it effectively off throughout the course of the experiment. We observed that, even in the presence of constitutively expressed genomic LacI repressor in MG1655 *E. coli*, the convergent toggle did not drift significantly from its initial state. This result can be interpreted as enhanced disturbance rejection capabilities of the convergent toggle; it requires a significant amount of LacI to flip the toggle switch to a high-LacI-RFP state. Small amounts of LacI are not sufficient to overcome the combined repression barrier imposed by supercoiling and TetR repression. In this way, the convergent orientation of the toggle reinforces the feedback architecture of the toggle switch, resulting in improved memory, disturbance rejection, and better thresholding performance. Clearly, compositional context can be a powerful tool for encoding feedback and improving performance of synthetic gene networks.

These experimental outcomes raise an important question. As intragenic context, e.g., choice of BCD, promoter design, or polycistronic design, is optimized to produce a functional gene cassette with model-predicted gene expression levels (Kosuri et al., 2013; Mutalik et al., 2013b), how do we ensure these predictions are not confounded by intergenic context as genes are composed?

One solution is to separate genes that need to have precise regulated expression levels on to different plasmids. However, the drawback of this approach is that separating genes on different plasmids introduces imbalances in gene copy number, which in turn can lead to additional design-build-test cycles to rebalance circuit dynamics. Also, it is often the case that there are too many genes in a biocircuit to be isolated individually on separate plasmids. In such settings, the findings of this work are important to consider, as they can be used to inform how to optimally compose adjacent genes.

The effects of adding spacing sequences between genes are complex. Specifically, we varied the amount of spacing between mSpinach and MG aptamers in convergent, divergent, and tandem orientation by adding increments of 100 bps between genes, and found that spacing did not have a monotonic effect on decreasing the fold change across orientations (see Figures S1B, S1E, and S1F). Most unusual was the sudden drop in signal observed in the divergent and tandem orientation, but not in the convergent orientation, with 450 bp of spacing between the two genes. It is possible that, since the persistence length of DNA is 150 bps, 450 bps of spacing facilitates formation of plectonomes (with DNA loops consisting of three 150 base pair domains) in the spacing region, which induce torsional stress and inhibit formation or movement of the transcription bubble.

In general, genes responded well to induction when induced one by one, although their raw expression levels varied depending on compositional context. This may explain why some circuits in the past have been successfully engineered, with little consideration given to the effects of supercoiling and compositional context. For example, the original toggle switch (oriented divergently) was designed to respond and latch to the presence of a single inducer (Gardner et al., 2000); it did this well and latched to LacI or TetR dominant states. In contrast, the threshold detection abilities of the toggle were not explored.

Likewise, the fold change in "off" versus "on" states of three input and four input AND gates developed by Moon et al. (2012) was strongest when comparing singly induced expression levels against the corresponding fully induced state. The four layer and three logic gates in these biocircuits were compositionally composed so that no two genes involved in any constituent layer of logic were placed adjacent to each other. Pairs of genes involved in logic gates were always separated by an auxiliary backbone gene or placed on separate plasmids. Overall, the success in this work suggests that genes can be insulated by inserting short "junk" transcriptional units in between each other. Engineering approaches for attenuating compositional context effects are a subject of future research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS
  - Plasmid Assembly & Strain Development
- METHOD DETAILS
  - Single Cell Fluorescence Microscopy
  - Plate Reader Experiments
  - Deriving Supercoiling Dynamics in an ODE Model of mSpinach and MG Aptamer Expression
  - Convergent Orientation Model
  - Divergent Orientation Model
  - Tandem Orientation Model
  - RFP CFP Reporter Models
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Software
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2017.06.001.

AUTHOR CONTRIBUTIONS

E.Y. wrote the paper. E.Y. and A.J.D. designed the experiments. K.B.M., A.H.N., A.J.D., J.L.B., and D.D.V. reviewed and edited the paper. E.Y. and A.H.N. developed models and performed formal analysis. E.Y., A.J.D., K.B.M., and A.H.N. conducted the experiments. R.M.M., D.D.V., and J.J.C. secured funding. R.M.M., D.D.V., J.L.B., and J.J.C. supervised modeling and experimental research design.

ACKNOWLEDGMENTS

We thank Ophelia Venturelli for her invaluable inspiration and guidance in this project; Victoria Hsiao, Jin Park, Anu Thubagere, and Adam Rosenthal for great advice on imaging; David Younger, Ania Baetica, Vincent Noireaux, Clarmyra Hayes, and Zachary Z. Sun for guidance and assistance with TX-TL experiments; and Lea Goentoro, Johann Pauls-son, Long Cai, Jennifer Brophy, John Doyle, Eric Klavins, and Julius Lucks for insightful conversations. This work was supported in part by a Charles Lee Powell Foundation Fellowship, a Kanel Foundation Fellowship, a National Science Foundation Graduate Fellowship, a National Defense Science and Engineering Graduate Fellowship, Air Force Office of Scientific Research grant (AFOSR) FA9550-14-1-0060, Defense Threat Reduction Agency grant HDTRA1-14-1-0006, and Defense Advanced Research Projects Agency grant HR0011-12-C-0065.

Received: December 24, 2015
Revised: February 8, 2017
Accepted: June 5, 2017
Published: July 19, 2017

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Babendure et al. (2003); Berg et al. (2002); Bintu et al. (2005); Bremer and Dennis (2008); Buc and McClure (1985); Kalisky et al. (2007); Lutz and Bujard (1997); Maier et al. (2011); Nelson and Sauer (1985).

REFERENCES

Babendure, J.R., Adams, S.R., and Tsien, R.Y. (2003). Aptamers switch on fluorescence of triphenylmethane dyes. J. Am. Chem. Soc. *125*, 14716–14717.

Balke, V.L., and Gralla, J. (1987). Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. J. Bacteriol. *169*, 4499–4506.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). Biochemistry (Freeman and Company).

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

Bremer, H., and Dennis, P.P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. EcoSal Plus *3*, http://dx.doi.org/10.1128/ecosal.5.2.3.

Brophy, J.A., and Voigt, C.A. (2016). Antisense transcription as a tool to tune gene expression. Mol. Syst. Biol. *12*, 854.

Buc, H., and McClure, W.R. (1985). Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry *24*, 2712–2723.

Cardinale, S., and Arkin, A.P. (2012). Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. Biotechnol. J. *7*, 856–866.

Ceroni, F., Algar, R., Stan, G.B., and Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat. Methods *12*, 415–418.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in bacteria. Cell *158*, 314–326.

Davis, J.H., Rubin, A.J., and Sauer, R.T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res. *39*, 1131–1141.

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol. Microbiol. *59*, 723–730.

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1*, http://dx.doi.org/10.14440/jbm.2014.36.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

Filonov, G.S., Kam, C.W., Song, W., and Jaffrey, S.R. (2015). In-gel imaging of RNA processing using Broccoli reveals optimal aptamer expression strategies. Chem. Biol. *22*, 649–660.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. Nature *403*, 339–342.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gyorgy, A., Jiménez, J.I., Yazbek, J., Huang, H.H., Chung, H., Weiss, R., and Del Vecchio, D. (2015). Isocost lines describe the cellular economy of genetic circuits. Biophys. J. *109*, 639–646.

Han, K.Y., Leslie, B.J., Fei, J., Zhang, J., and Ha, T. (2013). Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. J. Am. Chem. Soc. *135*, 19033–19038.

Hatfield, G.W., and Benham, C.J. (2002). DNA topology-mediated control of global gene expression in *Escherichia coli*. Annu. Rev. Genet. *36*, 175–203.

Kalisky, T., Dekel, E., and Alon, U. (2007). Cost-benefit theory and optimal design of gene regulation functions. Phys. Biol. *4*, 229.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R., and Collins, J.J. (2004). Programmable cells: interfacing natural and engineered gene networks. Proc. Natl. Acad. Sci. USA *101*, 8414–8419.

Korbel, J.O., Jensen, L.J., von Mering, C., and Bork, P. (2004). Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. Nat. Biotechnol. *22*, 911–917.

Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *110*, 14024–14029.

Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. Cell *132*, 971–982.

Lee, T.S., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., and Keasling, J.D. (2011). BglBrick vectors and datasheets: a synthetic biology platform for gene expression. J. Biol. Eng. *5*, 1–14.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA *84*, 7024–7027.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. *25*, 1203–1210.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.C., Aebersold, R., and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol. Syst. Biol. *7*, 511.

Meyer, S., and Beslon, G. (2014). Torsion-mediated interaction between adjacent genes. PLoS Comput. Biol. *10*, e1003785.

Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., and Voigt, C.A. (2012). Genetic programs constructed from layered logic gates in single cells. Nature *491*, 249–253.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P., et al. (2013a). Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods *10*, 354–360.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Mai, Q.A., Christoffersen, M.J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G., et al. (2013b). Quantitative estimation of activity and quality for collections of functional genetic elements. Nat. Methods *10*, 347–353.

Nelson, H.C., and Sauer, R.T. (1985). Lambda repressor mutations that increase the affinity and specificity of operator binding. Cell *42*, 549–558.

Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003). Principles of cell-free genetic circuit assembly. Proc. Natl. Acad. Sci. USA *100*, 12672–12677.

Opel, M.L., and Hatfield, G. (2001). DNA supercoiling-dependent transcriptional coupling between the divergently transcribed promoters of the ilvYC operon of *Escherichia coli* is proportional to promoter strengths and transcript lengths. Mol. Microbiol. *39*, 191–198.

Ouafa, Z.A., Reverchon, S., Lautier, T., Muskhelishvili, G., and Nasser, W. (2012). The nucleoid-associated proteins H-NS and FIS modulate the DNA supercoiling response of the pel genes, the major virulence factors in the plant pathogen bacterium *Dickeya dadantii*. Nucleic Acids Res. *40*, 4306–4319.

Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011). RNA mimics of green fluorescent protein. Science *333*, 642–646.

Rahmouni, A.R., and Wells, R.D. (1992). Direct evidence for the effect of transcription on local DNA supercoiling in vivo. J. Mol. Biol. *223*, 131–144.

Rhee, K.Y., Opel, M., Ito, E., Hung, Sp., Arfin, S.M., and Hatfield, G.W. (1999). Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the ilvYC operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *96*, 14294–14299.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference – a crash course. Trends Genet. *21*, 339–345.

Shin, J., and Noireaux, V. (2012). An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. ACS Synth. Biol. *1*, 29–41.

Siegal-Gaskins, D., Tuza, Z.A., Kim, J., Noireaux, V., and Murray, R.M. (2014). Gene circuit performance characterization and resource usage in a cell-free bread- board. ACS Synth. Biol. *3*, 416–425.

Tuza, Z., Singhal, V., Kim, J., and Murray, R.M. (2013). An in silico modeling toolbox for rapid prototyping of circuits in a biomolecular breadboard system. In Proceedings of the 2013 IEEE Conference on Decision and Control, A. Astolfi, ed. (IEEE), pp. 1404–1410.

Veening, J.W., Smits, W.K., Hamoen, L.W., Jongbloed, J.D., and Kuipers, O.P. (2004). Visualization of differential gene expression by improved cyan fluorescent protein and yellow fluorescent protein production in *Bacillus subtilis*. Appl. Environ. Microbiol. *70*, 6809–6815.

Young, J.W., Locke, J.C., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P.S., Mjolsness, E., and Elowitz, M.B. (2012). Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. Nat. Protoc. *7*, 80–88.

Zhang, J., Campbell, R.E., Ting, A.Y., and Tsien, R.Y. (2002). Creating new fluorescent probes for cell biology. Nat. Rev. Mol. Cell Biol. *3*, 906–918.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Amicon 0.5 mL PCR Purification Kit (3 kDa)	EMD Millipore	UFC500324
DNA Gyrase *(E. coli)*	NEB	M0306L
Deposited Data		
mSpinach-timelapse	www.github.com/YeungRepo/context_models	x4.41/x4.36
mSpinach-RFP-timelapse	www.github.com/YeungRepo/context_models	x4.37
RFP and CFP-timelapse	www.github.com/YeungRepo/context_models	x4.54
Gyrase treatment of MG mSpinach plasmid	www.github.com/YeungRepo/context_models	x4.44-x4.45
Gyrase treatment of RFP CFP plasmid	www.github.com/YeungRepo/context_models	x4.50-x4.51
Single gene controls	www.github.com/YeungRepo/context_models	x4.89
Convergent Toggle experiments	www.github.com/YeungRepo/context_models	x5.03
Two Plasmid Toggle experiments	www.github.com/YeungRepo/context_models	x5.04
Experimental Models: Organisms/Strains		
MG1655Z1 (Experimental Strain)	S. Mayo Caltech Laboratory	NA
MG1655 ΔLacI (Experimental Strain)	J. J. Collins Laboratory	NA
JM109 (Cloning strain)	Zymogen	T3005
DH5-alpha Z1 (Cloning strain)	Murray Laboratory	NA
Recombinant DNA		
1. Convergent MG mSpinach (ColE1)	This Paper	pEYC1A
2. Divergent MG mSpinach (ColE1)	This Paper	pEYD1A
3. Tandem MG mSpinach (ColE1)	This Paper	pEYT1A
4. Conv. RFP, CFP 100 bp spacing (ColE1)	This Paper	pEYRC.C1A
5. Divergent RFP, CFP, 100 bp spacing (ColE1)	This Paper	pEYRC.D1A
6. Tandem RFP, CFP, 100 bp spacing (ColE1)	This Paper	pEYRC.T1A
7. Convergent CFP, RFP 100 bp sp (p15A)	This Paper	pEYRC.C1C
8. Divergent CFP, RFP, 100 bp sp) (p15A)	This Paper	pEYRC.D1C
9. Tandem CFP, RFP, 100 bp sp (p15A)	This Paper	pEYRC.T1C
10. Convergent MG mSpinach 200 bp spacing (ColE1)	This Paper	pEYC2A
11. Divergent MG mSpinach 200 bp spacing (ColE1)	This Paper	pEYD2A
12. Tandem MG mSpinach 200 bp spacing (ColE1)	This Paper	pEYT2A
13. Convergent MG mSpinach 300 bp spacing (ColE1)	This Paper	pEYC3A
14. Divergent MG mSpinach 300 bp spacing (ColE1)	This Paper	pEYD3A
15. Tandem MG mSpinach 300 bp spacing (ColE1)	This Paper	pEYT3A
16. Convergent MG mSpinach 400 bp spacing (ColE1)	This Paper	pEYC4A
17. Divergent MG mSpinach 400 bp spacing (ColE1)	This Paper	pEYD4A
18. Tandem MG mSpinach 400 bp spacing (ColE1)	This Paper	pEYT4A
19. Convergent MG mSpinach 500 bp spacing (ColE1)	This Paper	pEYC5A
20. Divergent MG mSpinach 500 bp spacing (ColE1)	This Paper	pEYD5A
21. Tandem MG mSpinach 500 bp spacing (ColE1)	This Paper	pEYT5A
22. Convergent MG mSpinach 100 bp spacing linear DNA	PCR Purified	C1-Linear
23. Divergent MG mSpinach 100 bp spacing linear DNA	PCR Purified	D1-Linear
24. Tandem MG mSpinach 100 bp spacing linear DNA	PCR Purified	T1-Linear
25. Convergent. RFP, mSpinach (ColE1)	This Paper	pEYmR.C3A
26. Divergent RFP, mSpinach (ColE1)	This Paper	pEYmR.D3A
27. Tandem RFP, mSpinach (ColE1)	This Paper	pEYmR.T3A

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
28. Single reporter - CFP sense (ColE1)	This Paper	pEYsCFPA
29. Single reporter - CFP antisense (ColE1)	This Paper	pEYasCFPA
30. Single reporter - RFP sense (ColE1)	This Paper	pEYsRFPA
31. Single reporter - RFP antisense (ColE1)	This Paper	pEYasRFPA
32. Convergent Toggle Plasmid	This Paper	pEY31-CBiT
33. Two Plasmid Toggle	This Paper	pEY29A, pEY35C
34. Divergent Toggle	This Paper	pGC011
35. Convergent Toggle	This Paper	pGC014
Software and Algorithms		
ConvModelRFPCFP_Journal.m	www.github.com/YeungRepo/context_models	ConvModel
DivModelRFPCFP_Journal.m	www.github.com/YeungRepo/context_models	DivModel
TandModelRFPCFP_Journal.m	www.github.com/YeungRepo/context_models	TandModel

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Enoch Yeung (enoch.yeung@pnnl.gov).

EXPERIMENTAL MODELS

Plasmid Assembly & Strain Development

Initial efforts to characterize orientation effects involved cloning plasmids with no spacing DNA between genes. We used Gibson assembly to build these plasmids and naturally found that in the convergent and divergent orientation, the primers used to amplify overhangs had strong secondary structure, which reduced cloning efficiency. Thus, we inserted a minimum of 150 base pairs of randomly generated DNA. DNA sequences were randomly generated in MATLAB, using a custom script and the function randi() and subsequently screened for secondary structure in Geneious, a gene designer software. All spacer sequences between genes were determined to have no hairpins or any predicted secondary structure at 37° C before use in cloning workflows.

To construct mSpinach and MG RNA aptamer reporter plasmids, we ordered 500 bp Integrated DNA Technologies gBlocks containing the mSpinach and MG RNA aptamer coding sequences in convergent, divergent, and tandem orientation. Back-bones and DNA inserts were amplified and prepared at equimolar concentrations in an isothermal Gibson Assembly, incubated for one hour at 50 C, following the methods of Gibson et. al (Gibson et al., 2009). Gibson products were subsequently transformed into JM109 Zymogen E. coli using a quick-transform protocol, plated at 29° C on LB agar plates with 100 μg/mL of carbencillin. Colonies were screened using standard colony PCR techniques, sequence verified using Operon Sequencing's overnight sequencing service (both Standard and Power Read products). All strains were sequence verified both in JM109 and experimental strains of MG1655Z1 and MG1655Δ LacI.

To construct mSpinach and RFP plasmids, we used a similar approach as described above, except that we used an RFP coding sequence derived from BglBrick plasmid (pBbE5k-RFP), amplified as a linear double stranded DNA molecule compatible with Gibson assembly. We used an analogous approach to construct CFP and RFP reporter plasmids on the ColE1 backbone. To switch backbones (p15A with chloremphenicol resistance marker) and construct CFP and RFP sense and anti-sense plasmids, we used Golden Gate assembly with BsaI-HF enzyme from New England Biolabs (NEB R3535L). All Golden Gate parts were constructed using an internal protocol with standardized four base pair overhangs. Colonies were screened and sequence verified following the same techniques used for plasmids built by Gibson Assembly. Finally, all plasmids developed for this paper were sequence verified both from Qiagen purified plasmid and as glycerol stock (using Operon's DNA prep service). Sequence verified strains were stored in 17% glycerol stocks at -80 C with LB and either 100 μg/mL of carbencillin or 34 μg/mL of chloremphenicol.

Plasmids were designed and constructed using either the Gibson isothermal DNA assembly technique (Gibson et al., 2009) or Golden Gate DNA assembly approach (Engler et al., 2008) using BsaI type II restriction enzyme. All plasmids were cloned into JM109 E. coli (Zymo Research T3005) or NEB Turbo E. coli (NEB C2984H) strains and sequence verified. Sequence verified plasmids were transformed into MG1655Z1 and MG1655ΔLacI (also lacking TetR) strains of E. coli. All plasmids with ColE1 replication origin were transformed and cloned at 29 C to maintain low copy number of the ColE1 replication origin. Sequence verified colonies were grown in LB and the appropriate antibiotic and stored as glycerol stocks (17 % glycerol) at -80° C.

METHOD DETAILS

Single Cell Fluorescence Microscopy

Based on the principles elucidated by Han et al. (Han et al., 2013), we ran all our experiments at 29° C when imaging mSpinach. Cells were revived from glycerol stock overnight at 29° C in LB, diluted to an OD of 0.1 and recovered for 2 hours in log-phase. Cells were then diluted to an density of approximately 5×10^6 cells/mL of LB and loaded into a CellASIC plate. Separate solutions for flowing LB with 200 μM DFHBI and LB with 200 μM DFHBI and 1 mM IPTG were prepared and loaded into reagent wells in the CellASIC ONIX B04A plate for imaging.

Fluorescence and bright field images from time-lapse microscopy were cropped using ImageJ and analyzed in MATLAB with Schnitzcell (Young et al., 2012). For characterizing coexpression of mSpinach and MG RNA aptamer, we used single cell agar pad microscopy, with all cells grown shaking at 29° C in a 96-well plate96-well plate from overnight recovery until they reached log-phase (~4 hours). Induction occurred by transferring 10 μL of cultures into another 96-well plate96-well plate into 350 μL of LB with 1mM IPTG and 200 ng/mL aTc.

Plate Reader Experiments

For plate reader experiments, all cultures were revived from glycerol stock at 37° C in LB and the appropriate antibiotic, followed by redilution to OD 0.05-0.1, recovered at log-phase for 2 hours at 37° C, and then pipetted into a 96 square well glass bottom plate (Brooks Life Sciences MGB096-1-2- LG-L) with the appropriate media, antibiotic and inducer. All measurements were taken on Biotek Synergy H1 plate readers, using the internal monochromomator with excitation (and emission) wavelengths for mSpinach, MG aptamer, CFP, and RFP at 469 nm (and 501 nm) at gain 100, 625 (and 655 nm) at gain 150, CFP at 430 nm (and 470 nm) at gain 61 and 100, RFP at 580 nm (and 610 nm) at gain 61 and 100. For RNA aptamer imaging, all in vitro and in vivo experiments were performed at 29° C with 200 μM DFHBI (for mSpinach) and 50 μM of malachite green dye.

Analysis of Plate Reader Data

To generate the data plotted in Figures 5, S1B, and S1C, we extracted data from the Biotek H1 Synergy plate readers using the Gen5 software package, exported to MATLAB matrices for optical density (OD) and fluorescence intensity in either mSpinach (469 excitation, 501 emission, gain 100), GFP (485 excitation, 525 emission, gain 61) CFP (430 excitation, 470 emission, gain 61) and RFP (580 excitation, 610 emission, gain 61 or 150) channels with inverted (bottom-up) fluorescence acquisition. Each sample was background subtracted, normalized by OD, plotted either as a single time point t = 9.2 hours corresponding to the tail-end of exponential growth phase or as complete time traces from t = 0 to t = 11.7 hours (t = 700 minutes). Each strain was grown in duplicate in MatriPlate (Brooks Life Science Systems MGB096-1-2-LG-L) 96 well square well glass bottom plates at 500 μ L volumes.

Similarly, for toggle switch data analysis we followed the approach outlined above. We note that given our choice of ribosome binding sites for GFP and RFP (BCD1 and BCD9 respectively), expression of GFP and RFP was weaker to avoid ribosomal loading effects. Thus, we did not see significant signal until 8 hours after initial induction. Signal increased monotonically throughout the experiment, varying depending on the balance of IPTG and aTc induction. Data plotted in Figure 6 were background subtracted and normalized by OD.

To estimate RNA aptamer and protein expression in the TX-TL system, we used data from prior calibration experiments, titrating purified fluorescent protein or RNA aptamer and quantitating expression in the Biotek. Each Biotek was calibrated independently; the results of the calibration were used to back out fluorescent protein from raw AFUs, after background subtraction.

Imaging RNA Aptamers: Quantitating mSpinach Expression using Single Cell Time-Lapse Fluorescence Microscopy and Plate Readers

In our preliminary tests, we quickly found that mSpinach RNA aptamer is not particularly bright, compared to GFP, RFP, and other standard fluorescent proteins. Moreover, its brightness depends on the operating temperature of the experiment (Han et al., 2013), since the steady state folding configuration of the mSpinach RNA aptamer depends on temperature. We found that mSpinach signal at 200 μM DFHBI (Lucerna Technologies) was nearly undetectable at 37° C. To minimize photobleaching of mSpinach, we developed a custom Python script to interface with MicroManager (Edelstein et al., 2014), employing the fast shutter of the XFO-citep 120 PC (8 ms resolution) to time exposure of the mSpinach expressing cells to light. To maintain an operating temperature of 29° C we used a custom-built microscopy incubation chamber with a World Precision Instruments Heater controller.

Once the microfluidic plate (EMD Millipore Cell ASIC ONIX B04A) was thermally equilibrated, cells were loaded into the imaging chamber and trapped using a loading protocol provided by EMD Millipore to a density of about 3 cells per field of view. Fluorescence microscopy imaging was performed on an Olympus IX81 inverted fluorescence microscope using a Chroma wtGFP filter cube (450/50 BP excitation filter, 480 LP dichroic beamsplitter, and 510/50 BP emission filter), with an XFO-Cite^{p} 120 PC light source at 100% intensity and a Hamamatsu ORCA-03G camera. Following the recommendations of (Han et al., 2013), we limited imaging frequency and exposure to every 10 minutes and for 200 ms, respectively. All experiments were conducted with an untransformed control strain of MG1655Z1 E. coli in a parallel microfluidic chamber, to quantify background cell fluorescence in DFHBI. For Figures 2C, 2D, 3B, and S2B we segmented and tracked single cell traces of mSpinach (or RFP and CFP) fluorescence using Schnitzcell (Young et al., 2012) and subtracted background fluorescence from each experimental strain. For each point in time, background fluorescence was defined as the maximum of background chamber fluorescence (quantified using ImageJ as

mean fluorescence in a nearby non-occupied area of the microfluidic chamber housing the experimental strain) and background cell fluorescence of the control strain for each frame. The majority of background fluorescence was defined by the background fluorescence in cells, with infrequent fluctuations in background fluorescence due to slight perturbations to the autofocus plane.

We also found that fixing cells with paraformaledehyde lead to inconsistent RNA aptamer fluorescence, with unusually high levels of fluorescence in the negative control (especially in the MG aptamer channel). While fixing cells traditionally allows fixation of protein dynamics, this is not true for imaging mSpinach and MG aptamer. It is possible that fixation alters the permeability of the membrane and enables excessive buildup of MG oxalate dye, which at high concentrations non-specifically binds to RNA molecules in the fixed cell. For this reason, our experimental technique involved imaging of live cells on agar pads, moving as quickly as possible from agar pad to agar pad, and well after the dynamics of induction had reached steady-state.

In contrast, imaging mSpinach in the cell-free expression system developed by Shin and Noireaux (Shin & Noireaux, 2012) required relatively little effort. Fluorescence quantification was performed on a Synergy H1 Biotek plate reader with 469 nm wavelength excitation, 501 nm wavelength emission. Since TX-TL reactions are typically run at 29 C, this further facilitated formation of the mSpinach RNA aptamer in the 32-2 configuration, see the Supplemental Online Material for Paige et. al. (Paige et al., 2011). We did notice that imaging more frequently than 15 minutes had an effect on the dynamics of mSpinach (presumably due to photobleaching), hence we ran all experiments with 15 minute imaging frequencies.

It is important to note that production of mSpinach in 10 μL bulk volume in vitro reactions allows for approximately 109 more copies of mSpinach than produced in a cell, we speculate this greatly increases the detectability of mSpinach signal over in vivo assays. We found imaging mSpinach in dense cultures (OD $\approx$ 1) also produced significant signal above background. Thus, the primary challenges of working with mSpinach is its relatively weak signal per single cell. We anticipate that using the latest version of mSpinach (mSpinach2) or dBroccoli in future tests will greatly improve signal (Filonov et al., 2015).

Flow Cytometry Experiments

Flow cytometer experiments were performed using a BD Biosciences Flow Assisted Cell Sorter (FACS) Aria II Flow Cytometer to quantify GFP and RFP fluorescence. GFP fluorescence was detected using a 488 nm laser and 530/30 nm internal band-pass filter while RFP fluorescence was detected using 561 nm laser and a 610/20 nm internal band-pass filter. Each plasmid strain (featuring convergent or divergent orientation of the modified pIKE107 Gardner Collins toggle switch) in either MG1655 E. coli or MG1655 Δ LacI E. coli was plated on cells from clonal glycerol stocks, grown at 37° C on selective media agar plates overnight. Three colonies were picked from each plate to seed replicate cultures for the experiment. All cell cultures were grown in LB media with carbencillin at 100 μg/mL at 37° C. Cultures were induced with either 50 ng/mL of aTc or 1000 μM of IPTG for 5 hours (defined as the latching period from t = −5 to t = 0 hours). After latching, cells were diluted with a dilution factor of 1000x, in approximate 8 hour intervals, in selective LB media from t = 0 to t = 48 hours in the experiment. At t = 0, 24, and 48 hours, cell cultures were rediluted and grown for two hours to reach exponential growth phase and rediluted 1:10 in 1x phosphate buffer saline solution. As a negative control, we quantified GFP, RFP, and CFP fluorescence of an untransformed strain of MG1655 E. coli as well as cell-free 1x PBS stock to determine forward and side-scatter gating parameters for background particulate matter.

Flow Cytometry Data Analysis

All flow cytometry data was processed using the FlowJo Software. Cells were gated using an ellipsoidal gate of forward and side-scatter values. We utilized live-gating during data acquisition to obtain approximately 20,000 events. All distributions were plotted as modal percentage versus GFP intensity (in arbitrary fluorescent units). Modal percentage for a given GFP intensity is defined as the ratio of cell count for the given GFP intensity bin normalized by the cell count for the modal GFP intensity bin, multiplied by 100. This method of plotting eliminates the variability of total counts in sub-populations after gating, while still portraying important features of the distribution such as mode, modal variance and modal kurtosis.

Fitting Hill Functions for Different Gene Orientations

We modify the standard Hill Equation to include a term for promoter leakiness that is independent from the dynamic range due to inducer concentration.

The equation for expression due to a promoter with some leakiness and Hill function-type response to an inducer chemical is given as:

$$f([I]) = l + \alpha \frac{[I]^n}{K_m^n + [I]^n},$$

where *[I]* is inducer concentration, *l* is leaky expression, α is the amplitude of expression due to inducer, n is the apparent cooperativity of the response to inducer, and *Km* is the concentration at which induction is half maximal. Thus, the maximum total expression upon full induction is given by:

$$V_{max} = l + \alpha$$

All four parameters were fit using RFP/CFP expression data shown in Figure 4. Both RFP/CFP induction functions were fit for the case in which the other gene is fully induced using the MATLAB function nlinfit. RFP was fit to the data that varies aTc (1.56 ng/mL to 200 ng/mL) while keeping IPTG at 1000 nM (Figure 1). Similarly, CFP was fit to the data that varies IPTG (7.85 nM to 1000 nM) while keeping aTc at 200 ng/mL (Figure 1). Fits along with experimental data points were plotted for all three orientations for both RFP and CFP.

Deriving Supercoiling Dynamics in an ODE Model of mSpinach and MG Aptamer Expression

In this subsection we explore a detailed model for describing the interplay of supercoiling and gene expression. The motivation to do this arises from 1) experimental results which strongly suggest that supercoiling and not transcriptional interference is the primary cause of differences observed in mSpinach, CFP, RFP, and MG aptamer expression across different gene orientations and 2) the need for a mathematical modeling framework that describes how the temporal dynamics of gene expression vary as a function of supercoiling state and neighboring gene activity.

We consider three structural phenomena that arise in supercoiled DNA: positively supercoiled DNA, negatively supercoiled DNA, and R-loop formation (Drolet, 2006) of the RNAP-DNA elongation complex in negatively supercoiled DNA. We begin with the basic premises of the twin-domain supercoiling models (Liu & Wang, 1987), namely that when a gene is transcribed, negative supercoiling is introduced upstream of the open complex and positive supercoiling is introduced downstream of the open complex. We introduce several concepts from the supercoiling literature (Drolet, 2006; Korbel et al., 2004; Liu and Wang, 1987; Opel and Hatfield, 2001; Rahmouni and Wells, 1992).

Definition 1 We define the constant $h_0 = 10.5$ to be the number of DNA base pairs involved in a single turn of a B-form DNA molecule in its natural state.

Definition 2 We define the linking number α_{LN} of a region of DNA to be the number of supercoiling turns in that region.

Definition 3 We define the supercoiling density σ_X of a region of DNA X of N base pairs length as $\sigma = \alpha_{LN}/N$.

Thus, we will assume that the plasmid DNA in our experiments is in its natural B-form configuration. Of course, by simply defining $h_0 = 11$ or $h_0 = 12$, it is possible extend our results to consider DNA in its A and Z form respectively.

It is important to note the notions of positive and negative supercoiling correspond to the notions of left-handed twist and right-handed twist, respectively, and are well defined as long as the direction along which gene expression occurs is specified and fixed. For example, a gene expressing in the sense direction (as considered in the model by Wang and Liu (Liu & Wang, 1987) and the recent analysis of Chong and Xie (Chong et al., 2014)) creates right-handed twist or negative supercoiling, conformal with the natural twist or direction of turn in a DNA double helix, downstream of the transcription bubble and left-handed twist, or positive supercoiling, upstream of the transcription bubble. Thus, the convention that negative- supercoiling builds upstream of gene expression and positive supercoiling downstream, is sensible only when considering 'sense' transcription.

When a gene expresses in the anti-sense direction, then using the reference frame defined by sense transcription and the right-handed twist of DNA, we note that unless we rotate the axis of the reference frame 180 degrees, the buildup of supercoiling downstream of anti-sense transcription is still right-handed (i.e. negative) and the buildup of supercoiling upstream of anti-sense transcription is still left-handed (i.e. positive) (see Figure 4 for a visual example).

A simple way to prove this is to construct a physical model of a supercoiled double-helix. Take two ropes, twisted into a double helix with right-handed twist. Note that defining the twist of the double helix as right-handed inherently imposes directionality in your rope (e.g. left to right or bottom to top, your thumb pointing in the direction of right or top). Tie both ends to a topological barrier, e.g. by connecting them to form a loop (like a plasmid) or fused to two separate posts, so that the twist internal to the double helix cannot dissipate past these barriers. Simulate a transcription bubble by pulling the two ropes apart and notice that preceding the bubble (opposing the direction that your thumb pointed) you will see the generation of additional right-handed twist and succeeding the bubble you generate left-handed twist (conformal with the direction of your thumb). Notice that the bubble could have been formed by unwinding the double helix left to right (sense transcription) or right to left (anti-sense transcription). However, it does not matter what direction we unwound the DNA to form the bubble; the end result is the same — negative supercoiling or right-handed twist preceding the bubble and positive supercoiling (or left-handed twist) succeeding the bubble. Thus, the original twist of the DNA, not the direction of bubble propagation, defines what type of supercoiling builds up preceding and succeeding a transcription bubble.

It is important to clarify that we are not declaring the default supercoiling state of of DNA in vivo as generally negatively supercoiled. Rather, we are referencing the classical convention that states that the double helix inherently has right-handed curl or twist (Drolet, 2006). Moreover, we make no assertions about the exact amount of additional negative or positive supercoiling introduced surrounding a sense transcription bubble or an anti-sense transcription bubble. Various aspects of the nature of supercoiling build-up and propagation have yet to be characterized fully via experiments, such as the rate of propagation of supercoils, the spatial distribution of supercoils succeeding or preceding a transcription bubble, and how DNA promoter and transcript sequence pertain to the rate at which supercoils are introduced. While our model is thus constructed with the capacity for quantitative prediction, until it is supported by robust estimates of physiological parameters, it is meant provide a mechanistic hypothesis for explaining the effects observed in our in vivo and in vitro experiments as opposed to exact predictions.

When two genes are present, e.g. in the convergent orientation, the intergenic region between the two genes becomes exposed to both left-handed (positive) and right-handed (negative) twist. It is important to note that left-handed (positive) and right-handed twist (negative) do not simply cancel out — the arbitrary nomenclature of positive or negative twist does not confer the same algebraic consequences of adding positive and negative numbers. Rather, when the a right-handed DNA double helix experiences torsional stress from simultaneously introducing both left-handed and right-handed twist from two opposing point sources (e.g. transcription bubbles in convergent orientation), the two twists define opposing forces that meet each other at some kink point between the two point sources. The outcome is not annihilation of positive and negative supercoiling but rather the transition

of the kink along the longitudinal axis of DNA until an equilibrium is achieved, i.e. the forces driving left-handed twist through the kink are equally balanced by forces driving right-handed twist through the kink. At equilibrium the net force is zero but this does not implicate in any way that the presence of right-handed or left-handed coils have been annihilated. With each transcriptional event or binding of a gyrase or topoisomerase to modulate the surrounding DNA's supercoiling state, the equilibrium position of the kink is correspondingly adjusted.

With these observations in order, we now consider the scenario when two non-overlapping genes are adjacent to each other in varying orientations. For the purposes of our model, three regions of DNA for each gene will be of interest, the promoter of a transcriptional unit, the coding sequence of a transcriptional unit, and the intergenic spacing region between adjacent genes in our constructs. Supercoiling has been experimentally demonstrated to affect both the processes of transcription initiation and transcription elongation. Thus, we make a point to distinguish and keep track of the supercoiling states of both the promoter and coding transcript. For simplicity of exposition, we do not explicitly model the supercoiling density of the intergenic spacing region, however, our models will implicitly assume that the spacing region is able to absorb supercoils propagated from upstream or downstream transcription events up to the kink (if present).

For notation, when modeling the RNA aptamer plasmids, we will use TL_X where $X = G$ or S to denote the length of the MG and mSpinach RNA aptamer transcript respectively, EC_X to denote the elongation complex formed while transcribing gene X, R to denote RNA polymerase, PL_X to denote the length of the p_{Lac} and p_{Tet} promoters, and N_S the length of the intergenic spacing region of non-coding DNA between genes. Similarly, we will use the subscript $X = RF$ or CF to indicate the parameter of interest pertains to the coding sequence for RFP or CFP respectively.

Convergent Orientation Model

In the convergent orientation, promoters face each other and as both genes express,

$$\begin{aligned}
\varnothing &\xrightarrow{\rho_L} LacI & \varnothing &\xrightarrow{\rho_T} TetR\\
LacI &\xrightarrow{\delta_p} \varnothing & TetR &\xrightarrow{\delta_p} \varnothing\\
LacI + IPTG &\underset{k_{ua,\,L}}{\overset{k_{a,\,L}}{\rightleftharpoons}} aLacI & TetR + aTc &\underset{k_{ua,\,T}}{\overset{k_{a,\,T}}{\rightleftharpoons}} aTetR\\
p_{Lac} + LacI &\underset{k_{u,\,L}}{\overset{k_{seq,\,L}}{\rightleftharpoons}} pLacC & p_{Lac} + TetR &\underset{k_{ua,\,T}}{\overset{k_{seq,\,T}}{\rightleftharpoons}} pTetC\\
pLacC + IPTG &\xrightarrow{k_{a,L}} pLac + aLacI & pTetC + aTc &\xrightarrow{k_{a,T}} pTet + aTetR\\
R + p_{Lac} &\underset{k_r}{\overset{k_f(\sigma_{p,\,S})}{\rightleftharpoons}} CCs & R + p_{Tet} &\underset{k_r}{\overset{k_f(\sigma_{p,\,G})}{\rightleftharpoons}} CC_G\\
CC_S &\xrightarrow{k_{open}} EC_S & CC_G &\xrightarrow{k_{open}} EC_G\\
EC_S &\xrightarrow{k_{cat}(\sigma_{t,s})} mS + R + p_{Lac} & EC_G &\xrightarrow{k_{cat}(\sigma_{t,G})} MG + R + p_{Tet}\\
mS &\xrightarrow{\delta_m} \varnothing & mG &\xrightarrow{\delta_m} \varnothing
\end{aligned}$$

positive supercoiling propagates from the sense transcription bubble into the intergenic spacing region and negative supercoiling propagates from the anti-sense transcription bubble (see Figure 4) to form a kink in between the two genes. Where standard transcription translation models of gene expression assume constant rates of transcription initiation and transcription elongation, we now make explicit the dependencies of these rates on supercoiling state. The chemical reaction network for this orientation is given as:

We note here, that for simplicity, we model Lac_I and Tet_R as naturally occurring in their tetrameric and dimeric forms. The complex $aLac_I$ and $aTet_R$ denote the inducer-bound forms of Lac_I and Tet_R that are unable to bind to their target promoter. When Lac_I and Tet_R bind their respective promoter, we denote them with p_{LacC} and p_{TetC} to indicate the promoter is sequestered from transcriptional processes.

We now derive an expression for $\sigma_{p,S}$ (t) by first considering the effects of transcription on the supercoiling density of the transcript. Consider the collection of plasmids present in the cell or volume of cell-free extract. Consider a small time interval $[t, t+\varepsilon]$ for small $\varepsilon > 0$. Suppose that $\Delta_{LN,t}$ turns are introduced with the production of each mSpinach transcript and that $x_{\Delta LN,\,t}$ number of turns are introduced into the transcript as x mSpinach molecules are produced. Simultaneously, we suppose that if y additional open complexes have been formed, then correspondingly $y_{\Delta LN,t}$ turns have been removed from the transcript region (in order to facilitate open complex formation). Also, although the promoter region is short, each time a transcription initiation event occurs, the promoter region is unwound and propagates supercoiling. However, many transcription initiation events stall or reversibly dissociate. We suppose such events do not introduce significant amounts of supercoiling — rather, only when an elongation complex is formed do we suppose that the promoter region has been unwound and introduced supercoils in the proximal regions. Thus, we suppose that if there are y new elongation complexes, then there are $y_{\Delta LN,p}$ turns. Moreover, once the elongation complex departs, it is not necessarily true the promoter will resume its normal B-form DNA state. However, we suppose that the reaction event of a new initiation complex finally forming is indicative of a supercoiling state being removed. In this way, turns are gained and lost by the incoming and outgoing of holoenzyme complexes on the promoter and transcript regions. We assume that any transcriptional pausing,

abortive initiation, and aborted elongation events are effectively modeled by their respective transcriptional parameters. The dynamics of $\sigma_{t,S}$ can then be expressed as:

$$\sigma_{t,S}(t+\epsilon) = \sigma_{t,S}(t) + \frac{\Delta_{LN,t}}{n_{f,S}}(x-y) + \frac{\Delta_{LN,p}}{n_{f,S}}(y-z),$$

$$\sigma_{t,S}(t+\epsilon) = \sigma_{t,S}(t) + \frac{\Delta_{LN,t}}{n_{f,S}}\left((mS^c(t+\epsilon) - mS^c(t)) - \left(EC_S^c(t+\epsilon) - EC_S^c(t)\right)\right)$$

$$+ \frac{\Delta_{LN,p}}{n_{f,S}}\left(EC_S^c(t+\epsilon) - EC_S^c(t) - \left(CC_S^c(t+\epsilon) - CC_S^c(t)\right)\right),$$

where $\sigma_{t,S}(t)$ denotes the supercoiling density at time t, $mS^c(t)$ denotes the integer molecular count of total mSpinach molecules produced by time t, Δ_{LN} denotes the change in the linking number of the mSpinach coding region per mSpinach transcript expressed, and $n_{f,S}$ is the combined length of free mSpinach transcript and spacer that is able to absorb the residual twist introduced by transcription. The amount of free spacer DNA between negatively supercoiled and positively supercoiled DNA available to absorb additional supercoiling depends on the dynamic equilibrium between negative twist and positive twist from mSpinach and MG aptamer transcription, respectively. We suppose that the length

$$n_{f,S} \equiv \max(PL_S + TL_S + N_S/2(p_{Tet}/(p_{Tet} + pTetC)) + (N_S/2 + TL_G)pTetC/(p_{Tet} + pTetC) - \Delta_{kink}, 0)$$

where

$$\Delta_{kink,0} \equiv (\sigma_{t,S} + \sigma_{t,G})h_0)$$

and the third and fourth terms on the right-hand side constitute a weighted average of the length of DNA spacing available for either 1) the scenario where transcription is active in the adjacent gene or 2) the transcription factor TetR is bound to the pTet promoter. Similarly, we write

$$n_{f,G} \equiv \max(PL_G + TL_G + N_S/2(p_{Lac}/(p_{Lac} + pLacC)) + (NS/2 + TL_S)pLacC/(pLac + pLacC) + \Delta_{kink}, 0)$$

When $\sigma_{t,S} = \sigma_{t,G}$ note that the point of transition between negative and positive supercoiling is exactly centered. When $\sigma_{t,S} < \sigma_{t,G}$, i.e. the force from negative twist exceeds the force of positive twist, the kink is forced in the direction of the MG aptamer coding transcript and $n_{f,S} > TL_S + N_S/2$. Conversely, if mSpinach transcription does not produce additional negative supercoils to counteract the positive twist from MG aptamer expression, then $n_{f,S} < TL_S + N_S/2$.

The above equation states that the supercoiling density at time t + ε is the supercoiling density at time t with an additive perturbation term, corresponding to the change in supercoiling density from transcription of $x = mS^c(t+\varepsilon) - mS^c(t)$ transcripts. Normalizing by the reaction volume Ω on both sides, dividing by ε, and taking $\varepsilon \to 0$, we obtain an expression in terms of the derivative of mSpinach *concentration*:

$$\frac{d(\sigma_{t,S})}{dt} = \left(\frac{d(mS)}{dt} + \delta_m mS - \frac{d(EC_S)}{dt}\right)\frac{\Delta_{LN,t}}{n_{f,S}} + \left(\frac{d(EC_S)}{dt} - \frac{d(CC_S)}{dt}\right)\frac{\Delta_{LN,p}}{n_{f,S}}.$$

Notice that the quantity $d(mS)/dt + \delta_m mS$ represents the rate at which total mSpinach RNA aptamer is produced in the system, since it is the state dynamics of mSpinach without mRNA degradation. However, the supercoiling state of DNA is continuously regulated by gyrase, an enzyme that relieves positive supercoiling, and topoisomerase, an enzyme that relieves negative supercoiling. We estimate that gyrase relieves positive supercoiling of the transcript region at roughly $\gamma = 0.5$ turns per second per plasmid, while topoisomerase relieves negative supercoiling of the transcript region at roughly $\tau = 0.25$ turns per second per plasmid (Liu & Wang, 1987). Both enzymes act to maintain the natural physiological (negative) supercoiling density of σ_0. We suppose that in the absence of any transcriptional activity, the balance of these rates tends toward gyrase activity and a steady state of σ_0. For simplicity we suppose that gyrase and topoisomerase binding does not interfere with the transcriptional binding dynamics of polymerase. We incorporate these maintenance dynamics as follows:

$$\frac{d(\sigma_{t,S})}{dt} = \left(\frac{d(mS)}{dt} + \delta_m mS - \frac{d(EC_S)}{dt}\right)\frac{\Delta_{LN,t}}{n_{f,S}} + \left(\frac{d(EC_S)}{dt} - \frac{d(CC_S)}{dt}\right)\frac{\Delta_{LN,p}}{n_{f,S}} + m(\sigma_{t,S})$$

where

$$m(\sigma) \equiv T_0\tau\frac{[\sigma-\sigma_0]^-/k_{M,\tau}}{\sigma_0 + (\sigma-\sigma_0)^2/k_{M,\tau}} - G_0\gamma\frac{[\sigma-\sigma_0]^+/k_{M,g}}{\sigma_0 + (\sigma-\sigma_0)^2/k_{M,g}}$$

where ν is the total length of DNA, $x \equiv [x]^- + [x]^+$ denotes an additive decomposition of x into its strictly negative and nonnegative parts, and T_0 and G_0 are the topoisomerase and gyrase concentrations present in vivo or in vitro cell-free extract.

Next, to obtain an expression for $\Delta_{LN} < 0$, i.e. the number of negative supercoiling turns introduced by expression of one mSpinach transcript, we argue as follows. As the open complex proceeds along the anti-sense DNA template of mSpinach, it unwinds and displaces the supercoiling of a 17 base pair region (Liu & Wang, 1987), corresponding to the DNA footprint of a transcription bubble (i.e.

DNA-RNAP open complex). The transcription bubble requires an uncoiled region of DNA to transcribe. Thus, an additional $17/h_o$ turns are introduced into the upstream and downstream regions. We suppose that half of these turns are introduced as negative supercoiling and the other half as positive. Thus, in the wake of the transcription bubble passing through the entire transcript, there are

$$-\frac{17}{h_o}\frac{TL_S}{17}\frac{1}{2}=-\frac{TL_S}{(2h_o)}$$

negative supercoiling turns introduced into intergenic spacer downstream. When transcription termination occurs, the bubble is no longer held open by the open complex and the negative supercoils travel back into the unwound DNA of the mSpinach transcript and spacer, while the positive supercoils dissipate upstream of the promoter. Similarly, as the promoter expresses it also introduces negative supercoils downstream into the transcript region. The expression for $\sigma_{t,S}$ *(t)* then simplifies to

$$\dot{\sigma}_{t,S}=-\left(\dot{m}S-\delta_m mS-\dot{E}C_S\right)\frac{TL_S}{2h_0n_{f,S}}-\left(\dot{E}C_S-\dot{C}C_S\right)\frac{PL_S}{2h_0n_{f,S}}+m(\sigma_{t,S}).$$

Here we use dot notation over θ to denote the derivative of θ. Following similar arguments, we can write the dynamics of $\sigma_{p,S}$ *(t)* as

$$\dot{\sigma}_{p,S}=-\left(\dot{E}C_S-\dot{C}C_S\right)\frac{PL_S}{2h_0n_{f,S}}+m(\sigma_{p,S}).$$

Similarly, the supercoiling density dynamics for the MG RNA aptamer gene are given as:

$$\dot{\sigma}_{t,G}=-\left(\dot{E}C_G-\dot{C}C_G\right)\frac{PL_G}{2h_0n_{f,G}}+\left(\dot{M}G-\delta_m MG-\dot{E}C_G\right)\frac{TL_G}{2h_0n_{f,G}}m(\sigma_{t,G}),$$

$$\dot{\sigma}_{p,G}=-\left(\dot{E}C_G-\dot{C}C_G\right)\frac{PL_G}{2h_0n_{f,G}}+m(\sigma_{p,G}).$$

Notice the change in sign in the MG RNA aptamer dynamics. In this way, the directionality of sense transcription, relative to the right-handed twist of DNA, is encoded. If MG RNA aptamer was expressed in the anti-sense direction (which is the case in our divergently orientated construct), then the supercoiling introduced would be negative.

An important question is how transcription initiation rate $k_f(\cdot)$ and elongation rate $k_{cat}(\cdot)$ depends on supercoiling density. In Meyer and Beslon (2014)) it was argued that the reaction rate of transcription initiation could be modeled with a Hill function type curve, based on experimental data characterizing the pel_A and pel_E promoters (Ouafa et al., 2012). Although these results are specific to the bacterium *Dickeya dadantii*, it has been generally postulated that supercoiling density acts as a form of global gene regulation (Drolet, 2006; Rahmouni and Wells, 1992) both in prokaryotic and eukaryotic organisms. A study of the ilvY and ilvC promoters (Rhee et al., 1999) in E. coli suggest that promoter activity is optimal around around a certain value of σ^* and that activity tapers as σ diverges toward positive or negative infinity. Balke and Gralla (Balke & Gralla, 1987) argued that global supercoiling state forms the basis of a feedback loop for a system of genes in an organism, in response to environmental cues regarding metabolite and resource availability.

Broadly speaking, it is difficult to draw general conclusions regarding the relation of supercoiling state and promoter activity — all experimental measurements in the studies described above were of the global supercoiling density. In these studies, the common approach was to treat a purified plasmid with topoisomerase to introduce additional twist. Whether this twist was introduced uniformly across the plasmid or non-uniformly is unclear. However, we can suppose that when a topoisomerase was used to treat plasmid, it introduced a monotone amount of twist (gyrase introduced only negative coils and TopoI introduced only positive coils). We thus proceed supposing that incubation and treatment with a topoisomerase had a monotonic effect on promoter supercoiling state and that the overall qualitative trends observed regarding the ilvY, ilvC, and pelE, and pelA promoters can be used to inform the qualitative or phenomenological model of how local supercoiling density and promoter activity are related. Drawing from physical intuition, we argue that a promoter cannot initiate transcription if it is excessively wound with positive or negative twist. We suppose that transcription initiation is thus optimal at a particular value of local supercoiling density σ^*. Moreover, we suppose that for a given promoter of length PL_X, $X = S, G, RF$, or CF the optimal local supercoiling density optimum roughly is related to the optimal global supercoiling density σ_0 via the following approximation:

$$\sigma^*\approx\sigma_0 P_L/PL_X,$$

where P_L is the length of the plasmid. We model the rate of transcription initiation as a second-order symmetric Hill function,

$$k_{cat,X}(t)=\frac{\beta}{1+(\sigma_{t,X}(t)-\sigma*)^2\Big/k_{M,\sigma}},$$

with an optimum centered around σ^*, where $X = G$ or S for MG and mSpinach transcription respectively and ζ is the optimal putative forward reaction rate of transcription initiation assuming the supercoiling state $\sigma_{p,X}$ is optimal for transcription initiation. Similarly, we suppose in the case of transcriptional elongation that the optimum $\sigma^* = \sigma_0 P_L/TL_X$ and the elongation rate is defined by the functions

$$k_{f,X}(t) = \frac{\varsigma}{1 + (\sigma_{p,X}(t) - \sigma *)^2 \Big/ k_{M,\sigma}},$$

where $X = G$ or S for MG and mSpinach respectively and β is the putative transcription elongation rate when the supercoiling state $\sigma_{t,X}$ is optimal for transcription. Finally, we note the following conservation laws hold since DNA and RNAP concentration are constant in our in vitro system

$$\begin{aligned}
R^{tot} &= R + EC_S + EC_G + CC_S + CC_G,\\
p_{\text{Lac}}^{tot} &= p_{\text{Lac}} + CC_S + EC_S + pLacC,\\
p_{\text{Tet}}^{tot} &= p_{\text{Tet}} + CC_G + EC_G + pTetC.\\
LacI^{tot} &= LacI + aLacI + pLacC\\
TetR^{tot} &= TetR + aTetR + pTetC\\
IPTG^{tot} &= IPTG + aLacI\\
aTc^{tot} &= aTc + aTetR
\end{aligned}$$

Using these laws, we can write a reduced order dynamical system model for the convergent biocircuit:

$$\begin{aligned}
\dot{m}S &= k_{cat,S}(\sigma_{t,S})EC_S - \delta_m mS,\\
\dot{m}G &= k_{cat,G}(\sigma_{t,G})EC_G - \delta_m MG,\\
\dot{EC}_S &= k_{open}CC_S - k_{cat}(\sigma_{t,S})EC_S,\\
\dot{EC}_G &= k_{open}CC_G - k_{cat}(\sigma_{t,G})EC_G,\\
\dot{CC}_S &= k_f(\sigma_{p,S})\left(R^{tot} - EC_S - EC_G - CC_S - CC_G\right)\left(p_{\text{Lac}}^{tot} - CC_S - EC_S - pLacC\right) - (k_r + k_{open})CC_S\\
\dot{CC}_G &= k_f(\sigma_{p,G})\left(R^{tot} - EC_S - EC_G - CC_S - CC_G\right)\left(p_{\text{Tet}}^{tot} - CC_G - EC_G - pTetC\right) - (k_r + k_{open})CC_G\\
\dot{LacI} &= \rho_l + k_{ua,L}\left(IPTG^{tot} - IPTG\right) + k_{u,L}\left(LacI^{tot} - LacI - IPTG^{tot} + IPTG\right)\\
&\quad - k_{aL}LacI\ IPTG - k_{seq,Lp\text{Lac}}LacI - \delta_p LacI\\
\dot{TetR} &= \rho_t + k_{ua,T}\left(aTc^{tot} - aTc\right) + k_{u,T}\left(TetR^{tot} - TetR - aTc^{tot} + aTc\right)\\
&\quad - k_{aL}Tet\ RaTc - k_{seq,Tp\text{Tet}}TetR - \delta_p TetR\\
\dot{IPTG} &= -k_{a,L}(LacI + pLacC)IPTG + k_{ua,L}\left(LacI^{tot} - LacI - pLacC\right)\\
\dot{aTc} &= -k_{a,T}(TetR + pTetC)aTc + k_{ua,T}\left(TetR^{tot} - TetR - pTetC\right)\\
\dot{\sigma}_{t,S} &= -\left(\dot{m}S - \delta_m mS - \dot{EC}_S\right)\frac{TL_S}{2h_0 n_{f,S}} - \left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 n_{f,S}} + m(\sigma_{t,S})\\
\dot{\sigma}_{t,G} &= -\left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{2h_0 n_{f,G}} - \left(\dot{MG} - \delta_m MG - \dot{EC}_G\right)\frac{TL_G}{2h_0 n_{f,G}} + m(\sigma_{t,G}),\\
\dot{\sigma}_{p,S} &= -\left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 n_{f,S}} + m(\sigma_{p,S})\\
\dot{\sigma}_{p,G} &= -\left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{2h_0 n_{f,G}} + m(\sigma_{p,G})
\end{aligned}$$

In simulating the supercoiling dynamics we noticed that the magnitude of our local supercoiling states settle around steady- state values much higher than the traditionally accepted range of global supercoiling density. In practice, experiments have determined that DNA is negatively supercoiled with a global supercoiling density of -0.065 and can drop to as low as -0.1. This parameter does not reflect the local supercoiling density of the regions of interest in our model, namely the supercoiling density of the transcript and the promoter.

For example, a small region of DNA can maintain a positively coiled plectonome while the rest of the DNA is relatively relaxed. The global supercoiling density will reflect the overall twist, as opposed to the high density in either of the two regions. Wang and Liu estimated that expression of a single transcript can introduce supercoils into DNA at a rate of 4 supercoils per second per transcript. Assuming gyrase introduces $\gamma = 1$ negative supercoils per second and $\tau = 0.5$ positive supercoils per second on a given plasmid, if half the supercoils introduced propagate upstream and the over half downstream, over the course of just five minutes (Liu & Wang, 1987) the region downstream (such as the 150 bp spacer sequence in our plasmids) of the transcript could achieve a local supercoiling density of $\sigma = 2.0$. A measurement of the global supercoiling state of the plasmid, say 3.5 kbp in length, would yield a global estimate of only $\sigma = 0.08$! Therefore, it is important to note the distinction between local and global supercoiling density; the local supercoiling density of a region of DNA can reach much higher magnitudes despite a relatively low (and conventionally acceptable) global supercoiling density.

Divergent Orientation Model

In our divergently oriented plasmid, the Tet promoter and Lac promoter express in opposing directions, but the transcription bubbles diverge or move away from each other. Thus, the only torsional stress introduced comes from backward propagation of coils from unwinding the regions of DNA encoding promoters into the intergenic spacing region between the two genes. The Tet promoter back propagates positive supercoils into the intergenic spacing region while the Lac promoter back propagates negative supercoils. The position of dynamic equilibrium between the positively supercoiled region upstream of the Tet promoter and the negatively supercoiled region upstream of the Lac promoter is determined by the balance of forces arising from positive and negative twist (diametrically opposing each other) in the promoter supercoiling states $\sigma_{p,S}$ and $\sigma_{p,G}$. If $\sigma_{p,S}$ is much larger than $\sigma_{p,G}$ then the equilibrium shifts in favor of the Lac promoter and the positive coils are pushed closer to the actual Tet promoter (or vice-versa). We model the amount of spacer available to the promoters as $n_{f,S}$ and $n_{f,G}$ where

$$n_{f,S} = \max\{PL_S + N_S/2 - \Delta_K, 0\} \equiv \max\{PL_S + N_S/2 - (\sigma_{p,S} + \sigma_{p,G})h_0, 0\}$$
$$n_{f,G} = \max\{PL_G + N_S/2 + \Delta_K, 0\} \equiv \max\{PL_G + N_S/2 - (\sigma_{p,S} + \sigma_{p,G})h_0, 0\}$$

again noting that $n_{f,S} + n_{f,G} = PL_S + PL_G + NS$ base pairs defines the total length of DNA in which localized supercoiling buildup can propagate. We suppose that all other supercoils arising from transcription elongation are dissipated within regions downstream of the promoters.

The supercoiling dynamics of the divergently oriented construct for mSpinach and MG RNA aptamer are thus given as

$$\dot{\sigma}_{t,S} = \left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 TL_S} + m(\sigma_{t,S}),$$
$$\dot{\sigma}_{t,G} = \left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{TL_G} + m(\sigma_{t,G}),$$
$$\dot{\sigma}_{p,S} = -\left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 n_{f,S}} + m(\sigma_{p,S}),$$
$$\dot{\sigma}_{p,G} = \left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{2h_0 n_{f,G}} + m(\sigma_{p,G}).$$

while the rest of the system dynamics are as presented in the convergent model. Any differences in expression are thus a function of the supercoiling dynamics above, the initial conditions of these four states, and their effects on k_f, k_{cat} and the topoisomer maintenance function $m(\cdot)$.

Tandem Orientation Model

In the tandem orientation, negative supercoiling backpropagates from the p_{Lac} promoter into the intergenic spacing region between the MG aptamer coding sequence and the mSpinach promoter. The torsional stress introduced by downstream propagation of positive supercoils from MG aptamer elongation and upstream propagation of negative supercoils from mSpinach transcription initiation again defines a dynamic equilibrium that is determined by the balance of $\sigma_{t,G}$ and $\sigma_{p,S}$. When the Lac promoter for mSpinach is much more active relative to the transcriptional activity of the MG aptamer coding sequence, $\sigma_{p,S}$ can dominate $\sigma_{t,G}$ such that any residual positive supercoils from MG aptamer transcription are pushed back into the coding sequence for MG aptamer. Excessive negative supercoiling from the mSpinach promoter, likewise, can make their way into the MG aptamer coding sequence. This is especially likely if the transcript region of MG aptamer is short (since it generates less positive supercoils to counteract the twisting force of negative supercoiling from mSpinach expression). The presence of excessive negative supercoiling in a transcript region can result in the formation of a R-loop complex, a hybrid of the RNAP-DNA open complex and the nascent mRNA chain with upstream DNA (Drolet, 2006); this complex stalls the elongation process indefinitely and impedes subsequent transcription events. These effects are accounted for in the function $k_{cat}(\sigma)$, which tapers off toward 0 if $\sigma \to -\infty$. The sensitivity of k_{cat} to $-\sigma$ is determined by the parameter $k_{M,\sigma}$.

Alternatively, if MG aptamer expression is high or leaky, it can likewise propagate positive supercoils downstream into the spacer region, which subsequently shutoff promoter activity of mSpinach. The decrease in promoter activity in mSpinach only further enables MG aptamer expression, which leads to MG aptamer dominant expression. This is particularly relevant if the coding sequence for MG aptamer is long, or replaced with a long coding sequence for a protein, e.g. RFP. In such a scenario, the expression of RFP can repress future expression of mSpinach.

It is thus important to consider the dynamic equilibrium of $\sigma_{t,G}$ and $\sigma_{p,S}$ and how the balance of these forces impact the positioning of positive and negative supercoils in the transcript region of MG aptamer and the Lac promoter. We suppose that the length of DNA available for positive supercoiling buildup (from MG aptamer expression) is given as:

$$n_{f,G} = \max\{TL_G + N_S/2 + \Delta_K, 0\}$$

and the length of DNA available for negative supercoiling buildup (from mSpinach expression) is given as:

$$n_{f,S} = \max\{PL_S + p_{Tet}/(pTetC + p_{Tet})N_S/2 + pTetC/pTetC + p_{Tet}(TL_G + N_S/2) - \Delta_K, 0\}$$

where $\Delta_K \equiv (\sigma_{p,S} + \sigma_{t,G})h_0$. The supercoiling dynamics are thus given by

$$\dot{\sigma}_{t,S} = \left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 n_{f,S}} + m(\sigma_{t,S}),$$

$$\dot{\sigma}_{t,G} = \left(\dot{m}_G + \delta_m m_G - \dot{EC}_G\right)\frac{TL_G}{2h_0 n_{f,G}}\left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{2h_0 TL_G} + m(\sigma_{t,G}),$$

$$\dot{\sigma}_{p,S} = -\left(\dot{EC}_S - \dot{CC}_S\right)\frac{PL_S}{2h_0 n_{f,S}} + m(\sigma_{p,S}),$$

$$\dot{\sigma}_{p,G} = \left(\dot{EC}_G - \dot{CC}_G\right)\frac{PL_G}{2h_0 n_{f,G}} + m(\sigma_{p,G}).$$

RFP CFP Reporter Models

Modeling the expression of RFP and CFP instead of mSpinach and RFP does not change any of the preceding arguments for deriving dynamics of supercoiling states. The only difference at the transcriptional level is that we use different length parameters, TL_R and TL_C (see Table S1) to define the length of transcript regions and denote the transcriptional products of each transcription elongation reaction as m_C (for the CFP mRNA transcript) and m_R (for the RFP mRNA transcript).

The primary source of genetic context effects in our model is supercoiling at the DNA level. Therefore, in this work we do not consider the effects of secondary structure in mRNA or superhelicity of mRNA-DNA hybrids. Thus, we deliberately model translation reactions simplistically, with the following chemical reactions:

$$m_{CFP} \xrightarrow{k_{B16}k_{TL}R^{tot}} P_{C,u}$$
$$m_{RFP} \xrightarrow{k_{Bgl}k_{TL}R^{tot}} P_{R,u}$$
$$P_{C,u} \xrightarrow{k_{f,C}} P_{C,f}$$
$$P_{R,u} \xrightarrow{k_{f1,R}} P_{R,f}$$
$$P_{R,f} \xrightarrow{k_{f2,R}} P_R$$
$$CFP \xrightarrow{\delta_p} \varnothing$$
$$RFP \xrightarrow{\delta_p} \varnothing$$

We suppose these translation reactions are the same inclusion of maturation reactions for CFP and RFP. We suppose that CFP matures through a one-step process while RFP matures through a two-step process (Zhang et al., 2002). Here we do not necessarily assume that RFP is dimeric, since the variant of dsRed1 that we used in our experiments is actually monomeric. However, we suppose that there is an intermediate stage between unfolded RFP and the final folded RFP. We found that including this intermediate stage recapitulated the significant delay observed in RFP expression in the cell-free expression system, that was not seen in CFP.

Moreover, the cell-free expression system is typically run in a bulk reaction setting, as a closed biochemical reaction system with a finite and limited amount of ATP, NTPs, and energy molecules to carry out transcription and translation. It has been observed empirically and shown through experiments that as ADP levels build up relative to ATP, enzymatic reactions become increasingly unfavorable (otherwise fluorescent reporters not subject to degradation would express in unbounded and increasing concentrations). Throughout our experiments, we observed these effects of resource depletion, beginning at $t_0 \approx 2$ hours onwards. To be consistent with the modeling approaches of Tuza and Singhal, (Siegal-Gaskins et al., 2014; Tuza et al., 2013), we suppose that the translation rate $k_{T\,L}\,(t)$ decays with time as a first order process, beginning at time t_0, and with decay parameter $\alpha_d = log(2)/(480)$

$$k_{TL}(t) \equiv k_{TL}^0 e^{\left(-\alpha_d (t-t_0)\mathbf{1}_{t>t_0}\right)}$$

where k_{TL}^0 is the nominal translation rate assuming an open system with limitless ATP and energy. The additional reaction dynamics in the state-space model are thus specified as follows:

$$\dot{P}_{C,u} = k_{B16}k_{f,C}k_{TL}R^{tot}m_{CFP} - k_{f,C}P_{C,u} - \delta_p P_C$$
$$\dot{P}_{C,f} = k_{f,C}P_{C,u}$$
$$\dot{P}_{R,u} = k_{Bgl}k_{f,R}k_{TL}R^{tot}m_{RFP} - k_{f,R}P_{R,u} - \delta_p P_R$$
$$\dot{P}_{R,f} = k_{f1,R}P_{R,u} - k_{f2,R}P_{R,f}$$
$$\dot{P}_R = k_{f,2R}P_{R,f}$$

The outcomes of our simulations are plotted in Figure 4 using parameters from Table S1. We see that RFP and CFP expression varies depending on orientation, initial condition of supercoiling states, and that the model is able to recapitulate the trends observed in the data.

While our model is able to describe the effects observed, it is the gyrase experiments that definitively confirm the validity of supercoiling as a working hypothesis for the physical mechanism driving compositional context effects. Our model serves to validate supercoiling as a hypothesis for compositional context, but not necessarily to prove it.

In conclusion, we have constructed three versions of a simple biocircuit to motivate the need to model compositional context in biocircuit assembly. Our initial data suggests that promoter orientation between pairs of promoters has a salient effect on gene expression. We developed a nonlinear model incorporating various phenomena resulting from compositional context and show it captures the patterns seen in experiments. We emphasize that these results are wholly the consequences of compositional context. There is no designed interaction in the biocircuit, yet different expression biases arise depending on how genes are arranged. Therefore, with any biocircuit comprised of multiple parts, modeling the effects of compositional context should be a chief consideration during the design and prototyping process.

QUANTIFICATION AND STATISTICAL ANALYSIS

Figures 1B, 1C, 3D, 3E, 4, 5, and 7 all utilize statistical analysis, specifically the error bars around mean traces are computed as the SD of experimental replicates. In all such plots, an experimental replicate refers to a separate experiment quantitated and executed in batch mode, using identical experimental parameters. For example, a replicate of an in vivo bulk culture expression assay for RFP and CFP expression corresponds to a separate batch reaction/growth assay inoculated and executed in a separate well and quantitated in parallel with all other replicates. All imaging parameters, growth temperatures, growth media, selection pressure, and shaking conditions for these replicates are identical.

Figures 2B and 3B utilize statistical analysis to estimate the SD of the mean fluorescent response to induction, where variability is defined across a population of cells growing in a microfluidic chamber. The SD over time was calculated by computing the SD across all segmented cell traces at each time point.

DATA AND SOFTWARE AVAILABILITY

Software

Models were simulated in MATLAB and.m script files for each model are available on GitHub at www.github.com/YeungRepo/context_models.

Data Resources

The accession numbers for all deposited plasmid sequences reported in this paper are GEO: KX682204 (pEY29A), GEO: KX682205 (pEY31-CBiT), GEO: KX682206 (pEY35C), GEO: KX682207 (pEYasCFPA), GEO: KX682208 (pEYasRFPA), GEO: KX682209 (pEYC1A), GEO: KX682210 (C1-Linear), GEO: KX682211 (pEYC2A), GEO: KX682212 (pEYC3A), GEO: KX682213 (pEYC4A), GEO: KX682214 (pEYC5A), GEO: KX682215 (pEYD1A), GEO: KX682216 (D1-Linear), GEO: KX682217 (pEYD2A), GEO: KX682218 (pEYD3A), GEO: KX682219 (pEYD4A), GEO: KX682220 (pEYD5A), GEO: KX682221 (pEYmRC3A), GEO: KX682222 (pEYmRD3A), GEO: KX682223 (pEYmRT3A), GEO: KX682224 (pEYRC.C1A), GEO: KX682225 (pEYRC.D1A), GEO: KX682226 (pEYRC.T1A), GEO: KX682227 (pEYRCC1C), GEO: KX682228 (pEYRCD1C), GEO: KX682229 (pEYRCT1C), GEO: KX682230 (pEYsCFPA), GEO: KX682231 (pEYsRFPA), GEO: KX682232 (pEYT1A), GEO: KX682233 (T1-Linear), GEO: KX682234 (pEYT2A), GEO: KX682235 (pEYT3A), GEO: KX682236 (pEYT4A), GEO: KX682237 (pEYT5A), GEO: KX682238 (pGC011), and GEO: KX682239 (pGC014). Experimental data files are also available at www.github.com/YeungRepo/context_models.