2014 Conference on Decision and Control (CDC)
http://www.cds.caltech.edu/~murray/papers/yeu+14-cdc.html

Modeling the Effects of Compositional Context on Promoter Activity in an *E. coli* Extract based Transcription-Translation System

Enoch Yeung[1*], Andrew Ng[2*], Jongmin Kim[3*], Zachary Z. Sun[4], and Richard M. Murray[1,4]

***Abstract*— One of the fundamental challenges in implementing complex biocircuits is understanding how the spatial arrangement of biological parts impacts biocircuit behavior. We develop a set of synthetic biology parts for systematically probing the effects of spatial arrangement on levels of transcription. Our initial experimental assays prove that even the rearrangement of two biocircuit parts (comprised of a promoter, coding sequence, and terminator) into three spatially distinct orientations (convergent, divergent, and tandem orientation) can exhibit significantly different levels of transcription. These findings motivate the need for mathematical models to describe these spatial context effects. We pose a novel nonlinear mass-action kinetics based model that enables the integration of knowledge about spatial or compositional context and canonical descriptions of transcriptional dynamics. Our findings suggest that compositional context plays a role in biocircuit part performance and comprise an important piece of biocircuit interconnection theory.**

I. Introduction

Synthetic biology is an expanding discipline, with important applications in sustainable energy, human medicine, and food and chemical industries [1]. These applications often involve the implementation of synthetic networks of genes, also known as biocircuits. Recently, advancements in DNA synthesis techniques [2], [3] coupled with the simultaneous drop in sequencing and prototyping costs [4], [5] have enabled rapid development of new biocircuits. In principle, design iterations for a complex biocircuit can be performed in a matter of hours [4], leading to circuit synthesis in a matter of weeks.

In parallel to this research, multiple biological part repositories have been established, such as the Registry of Standard Biological Parts and AddGene. Such repositories provide a continually expanding library from which a combinatorial number of biocircuits can be constructed. However the state-of-the-art in biocircuit prototyping relies on brute-force exploration of the design-space, by fabricating every possible combination of biocircuit parts in search of a functional variant. Such an approach is tractable in genetic circuits with a few genes [6], but if the goal is to achieve complex biocircuits comprised of more parts, the combinatorial complexity of the design space becomes prohibitive.

An alternative strategy to a brute-force search of the design space is model-guided design, where existing models for biological parts in combination with results from biocircuit interconnection theory are used to construct predictive system models for each design prototype. However, modeling the interconnection of synthetic biological parts is not a simple task [7], since the inclusion of multiple biological parts can introduce various phenomena such as loading effects [8], [9], [10], resource competition [11], [12], and indirect crosstalk [13].

An additional factor in biocircuit interconnection theory has recently received attention: compositional context, the way in which biocircuit genes are spatially arranged in plasmid or genomic DNA. More generally, Cardinale and Arkin in [13] argue that a major source of variability and failure in biocircuit implementation is biocircuit context. They argue there are three types of biocircuit context that impact a circuit's dynamics: host context, environmental context, and compositional context. All three of these types of context are pervasive in *in vivo* systems — in this paper we study the effects of compositional context.

Existing work on compositional context has focused on naturally occurring *in vivo* systems, see [14] and [15] for two examples. Since our goal is to develop a modeling framework to describe interconnection in *synthetic* biological systems, we will use the tools of synthetic biology to experimentally address this question. Thus, we construct a simple series of biocircuits to drive our study of compositional context.

The rest of our paper is organized as follows: in Section II we discuss the details of our biocircuit design, synthesis, and experimental results to characterize compositional context effects. In Sections III.A, III.B, III.C we introduce models that incorporate supercoiling, R-Loop formation, and terminator leakage and show that these models are able to quantitatively recapitulate the trends in the experimental data.

II. A Simple Biocircuit to Study Compositional Context

The most basic circuit that enables study of compositional context has two genes. A minimum of two genes is required to explore the three possible spatial layouts, or orientations, of adjacent genes: convergent, divergent, and tandem. Convergent genes are transcribed towards each other, divergent genes are transcribed away from each other, and tandem genes are transcribed in the same direction. Thus, we will construct three versions of the same biocircuit to account for the three possible gene-pair orientations. Since gene (and promoter) orientation and spacing between transcriptional units are properties of the biocircuit DNA, we

*These authors contributed equally to this work. For correspondence, please contact eyeung@caltech.edu or andrew.ng@wustl.edu. 1-Control and Dynamical Systems, Caltech, 2-Department of Biomedical Engineering, Washington University in St. Louis, 3-Department of Systems Biology, Harvard University, 4-Division of Biology and Biological Engineering, Caltech.

use RNA based reporters to monitor transcriptional activity. The absence of ribosome binding sites is advantageous since it ensures minimal structural interference from ribosomes binding to nascent mRNA during the elongation phase of transcription. Thus, RNA based reporters eliminate any expression biases from translation machinery — they allow us to study the effects of compositional context on purely at the transcriptional level.

Since our goal is to discern variation in gene expression as a function of compositional context, we make a point to control for additional sources of gene expression variability. Specifically, we assembled all three versions of the biocircuit on the same plasmid backbone, to avoid confounding the effects of compositional context with plasmid copy number distribution. To control against variable intergenic spacing between the genes of the plasmid and our biocircuit, we inserted the genes encoding the biocircuit as a single fragment (using a parallel set of Gibson isothermal assembly reactions) at the same point in the backbone in each version. The terminators for each of the genes in our biocircuit were designed to be the same, since terminator efficiency has been known to affect downstream promoter activity in the tandem orientation and convergent orientation [16]. By using identical terminators, any expression interference through anti-termination, leaky termination, and steric occlusion of colliding elongation RNAP complexes (in the convergent orientation) would be equal across both genes. Finally, to avoid any crosstalk or competition effects for repressor or activator proteins, we chose distinct inducible promoters for the genes. This would ultimately facilitate our study of how compositional context affected gene activation and repression (data not shown in this paper).

Our biocircuit consisted of two genes: the malachite green (MG) [17] and mSpinach RNA aptamer [18]. The mSpinach aptamer fluoresces at a green wavelenth when bound to the dye 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), and the malachite green aptamer fluoresces at a red wavelength when bound to the malachite green oxalate dye. The separation between the excitation and emission spectra of these fluorescent RNA reporters allows us to study their gene expression simultaneously without spectral crosstalk. As mentioned above, we constructed three biocircuits for this study, one corresponding to each type of promoter-pair orientation (convergent, divergent, and tandem). In these circuits, the p_{Tet} promoter drives transcription of the MG RNA aptamer and the p_{Lac} promoter drives transcription of the mSpinach RNA aptamer. For the plasmid backbone, we used the pBEST plasmid backbone from [19] with a ColE1 replication origin and an AmpR antibiotic resistance marker.

For plasmid construction, we synthesized three gBlocks using Integrated DNA Technologies' DNA synthesis service, each implementing one of the three different orientations. We inserted each gBlock into the pBEST plasmid using Gibson isothermal assembly [3]. Sequences for the p_{Tet} and p_{Lac} promoters were taken from the Biobricks Parts Registry. Sequences for the MG and mSpinach RNA aptamer (including the tRNA scaffold) were taken from [17] and [18] respectively. Figure 1 shows a diagram of the three different versions of the biocircuit constructed for this study.

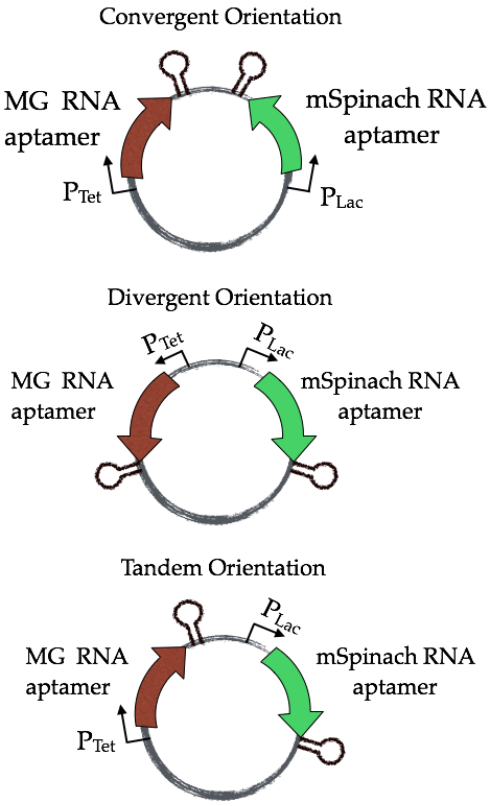

Fig. 1: A schematic illustrating the three types of compositional context that occur for our two-gene biocircuit. Two adjacent genes on a plasmid can be spatially arranged in convergent (top), divergent (middle), and tandem (bottom) orientation.

In choosing our assay to characterize expression we noted that *in vivo* assays are subject to large amounts of variability (due to growth conditions, growth history, temperature fluctuations, intracellular variability); these variability effects would make it more difficult to tease out the effects of compositional context from the effects of host or environmental context [13]. Furthermore, the MG aptamer relies on the MG-oxolate dye, which is known to be slightly cytotoxic to cells at standard working concentrations $(5-50\,\mu M)$. Thus, to isolate the effects of compositional context from these sources of variability, we chose an *in vitro* expression assay based on a BL21 Rosetta 2 *E. coli* S30 extract transcription-translation (TX-TL) system developed in [19]. The TX-TL system was also demonstrated as a biomolecular breadboard environment for rapid high-throughput prototyping of novel biocircuits [4]. Hence, any additional insight we gain about compositional context also informs prototyping efforts in the TX-TL system.

We performed TX-TL experiments in a 384-well plate format, using 10 μL reactions and equimolar concentrations of plasmid DNA miniprepped and PCR purified from overnight LB cultures. We monitored expression of the MG and mSpinach RNA aptamer using a Biotek Synergy HM1

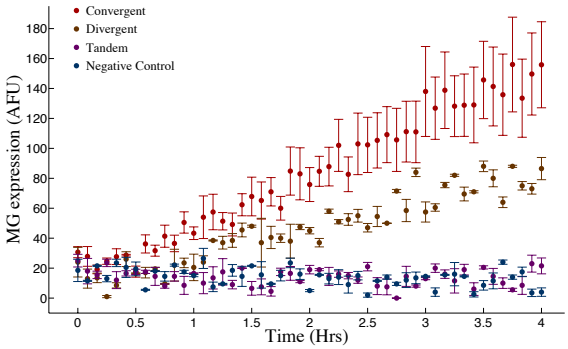

Fig. 2: Comparison of fluorescence intensity of malachite green aptamer between each biocircuit over a period of four hours.

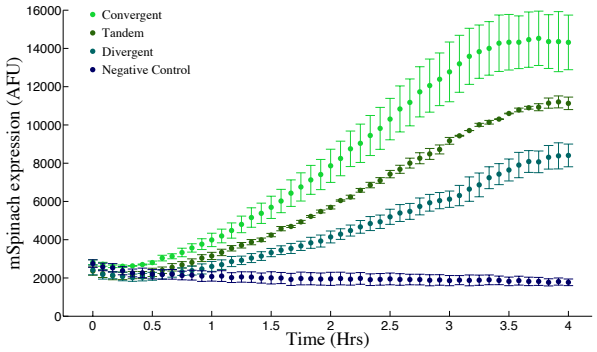

Fig. 3: Comparison of fluorescence intensity of mSpinach aptamer between each biocircuit over a period of four hours.

plate reader. Specifically, we collected time-series data for four hours at 29 °C in separate fluorescent channels, with excitation and emission wavelengths respectively of 1): 610 nm and 650 nm for MG aptamer detection, and 2) 470 nm and 510 nm for mSpinach aptamer detection. All experiments were performed in triplicate — the data from this assay is shown in Figures 2 and 3.

From the data, we see that the orientation of the gene affects both MG RNA aptamer and mSpinach expression in the TX-TL system. In the MG RNA aptamer fluorescent channel, convergent orientation expresses nearly two-fold more than divergent orientation, while expression in the tandem construct is negligible. In the mSpinach RNA aptamer fluorescent channel, convergent achieves two-fold more expression than divergent and nearly 1.5 fold more than tandem orientation.

III. A Transcription-Induced Supercoiling Mass Action Kinetics Model

While experiments to reverse-engineer a mechanistic explanation for these expression differences would be beyond the scope of this study (and is a field of research of its own), it is important to at least be aware of the biases introduced by compositional context in gene expression. Additionally, it is important to develop models (gray-box, black-box, or physical models) that capture these effects at a level of abstraction required for biocircuit interconnection theory. This is a challenge since existing work in this area focuses heavily on detailed physical simulations of plasmid supercoiling density as opposed to the phenomenological biases introduced in biocircuit performance. Thus, our goal in this section is to develop a modeling approach that can be generalized in future work to biocircuits of arbitrary complexity. Therefore, we seek simple representations that recapitulate the expression biases seen in experimental results, but that are grounded in the physical phenomena (relating to compositional context) driving these expression biases.

We will discuss three potential mechanistic phenomena from the literature that relate to compositional context: terminator leakage, transcription-induced supercoiling, and R-loop formation of the RNAP-DNA elongation complex. We argue that each of these phenomena when considered separately are insufficient to recapitulate at least one aspect in the data. We then pose a model where all three phenomena are considered simultaneously and show it is possible to predict the expression trends seen in the data.

The first and most commonly considered phenomena in synthetic biology models is terminator leakage. Terminator leakage occurs when the RNAP-DNA elongation complex is able to escape past the terminator region of a coding sequence and continue its transcript elongation using DNA downstream of the terminator. Notice that in such a model, we would expect minimal terminator leakage in the convergent orientation, since the RNAP-DNA elongation complex would have to leak through two T500 terminator hair-pins (which each have an experimentally measured efficiency of 98% [20]). In the divergent orientation, terminator leakage would only result in elongation through the plasmid backbone. Most likely, given the additional genes in the plasmid backbone and the length of the plasmid, transcription would terminate well before it reached the other gene in the biocircuit. Thus, from a transcriptional model based *purely* on terminator leakage we would expect no expression differences between the convergent and divergent versions of our biocircuit, which is inconsistent with the experimental data.

Alternatively, a model of tandem orientation incorporating supercoiling effects [21] would predict that MG expression would positively correlate with mSpinach expression and that mSpinach expression would have little effect on MG expression, since it transcribes in the opposite direction and propagates 1) negative supercoils back into the intergenic spacing region between the two genes and 2) positive supercoiling downstream with the expression of each mSpinach transcript [21]. Since MG is upstream of mSpinach, positive supercoils propagating downstream mSpinach would not have a significant effect on MG. On the other hand, negative supercoiling is shown in [15] and [22] to be beneficial for gene expression and thus we would expect MG expression to be significant in the tandem orientation.

Thus, a transcriptional model incorporating supercoiling and coexpression effects would be unable to explain the reduced MG expression (Figure 2) in the tandem orientation .

In the tandem orientation, negative supercoiling back-propagates from the p_{Lac} promoter into the $3'$ end of the coding sequence of MG aptamer. It is necessary, therefore, to consider the effect of negative supercoiling on transcription elongation. In particular, the authors in [23] review a series of experimental papers that show transcription-induced negative supercoiling from downstream genes can result in the formation of a R-loop structural complex between downstream negatively supercoiled DNA, the RNAP-DNA open complex and the nascent mRNA chain. This complex stalls the elongation process indefinitely and impedes subsequent transcription events. In this way, R loop complexes act to repress genes. It is only when we consider all three of these phenomena (leakage, supercoiling, and R-loop mediated stalling) that we are then able to recapitulate the expression biases seen in the experimental data.

Before we proceed, it will be useful to introduce several concepts from the supercoiling literature [14], [15], [21], [23], [22].

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.

For the purposes of our model, three regions of DNA 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. We will assume that the terminator regions do not collect supercoiling as readily as the surrounding DNA regions, since the T500 terminator forms a structurally stable hairpin. Further, we will not explicitly model the supercoiling density of the intergenic spacing region, however, our models will implicitly assume that the spacing region is able to maintain supercoils propagated from upstream or downstream transcription events.

For notation, 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 noncoding DNA between genes.

A. Convergent Orientation Model

In the convergent orientation, promoters face each other and as both genes express, positive supercoiling propagates from the transcription bubble into the intergenic spacing region and coding regions of the opposing gene. The chemical reaction network for this orientation is given as:

$$R + p_{\text{Lac}} \underset{k_r}{\overset{k_f(\sigma_{p,S})}{\rightleftharpoons}} EC_S \xrightarrow{k_{cat}(\sigma_{t,S})} mS + R + p_{\text{Lac}}$$

$$R + p_{\text{Tet}} \underset{k_r}{\overset{k_f(\sigma_{p,G})}{\rightleftharpoons}} EC_G \xrightarrow{k_{cat}(\sigma_{t,G})} MG + R + p_{\text{Tet}}$$

$$mS \xrightarrow{\delta_m} \emptyset$$

$$MG \xrightarrow{\delta_m} \emptyset$$

where R, EC_S denote

In this orientation, we suppose that terminator read-through is negligible, since it necessarily requires the open complex to pass through two T500 terminators and a non-coding intergenic spacing region. We now derive an expression for $\sigma_{p,S}(t)$ by first considering the effects of transcription on the supercoiling density of the transcript. The supercoiling density of the transcript region of mSpinach after the production of x transcripts of mSpinach (we assume for simplicity that there are no abortive transcription events) produced in the time interval $[t, t+\epsilon]$ can be expressed as:

$$\sigma_{t,S}(t+\epsilon) = \sigma_{t,S}(t) + x\frac{\Delta_{LN} h_o}{TL_S},$$

$$\sigma_{t,S}(t+\epsilon) = \sigma_{t,S}(t) + (mS^c(t+\epsilon) - mS^c(t))\frac{\Delta_{LN} h_o}{TL_S},$$

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

$$\frac{d(\sigma_{t,S})}{dt} = \left(\frac{d\,(mS)}{dt} + \delta_m mS\right)\frac{\Delta_{LN} h_o}{TL_S}\Omega.$$

Notice that the quantity $\dot{mS} + \delta_m mS$ 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, we know from [21] that gyrase relieves positive supercoiling of the transcript region at roughly $\gamma = 0.5$ turns per second, while topoisomerase relieves negative supercoiling of the transcript region at roughly $\tau = 0.25$ turns per second. Both enzymes act to maintain the natural physiological (negative) supercoiling density of $\sigma_0 = -0.65$. We incorporate these maintenance dynamics as follows:

$$\frac{d\,(\sigma_{t,S})}{dt} = \frac{\Delta_{LN} h_o}{TL_S}\Omega\left(\frac{d\,(mS)}{dt} + \delta_m mS\right) + \frac{\left(\tau \mathbf{1}_{\sigma_{t,S}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{t,S}>\sigma_0}\right) h_o}{TL_S},$$

where $\mathbf{1}_x$ is an indicator function for the Boolean condition x. 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 DNA template, it unwinds and displaces the supercoiling of a 17 base pair region, 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. The expression for $\sigma_{t,S}(t)$ then simplifies to

$$\begin{aligned}\dot{\sigma}_{t,S} = &-\frac{\Omega}{2}\left(k_{cat}(\sigma_{t,S})EC_S\right)\\ &+\frac{h_o}{TL_S}\left(\tau\mathbf{1}_{\sigma_{t,S}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,S}>\sigma_0}\right).\end{aligned}$$

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

$$\begin{aligned}\dot{\sigma}_{p,S} = &-\frac{\Omega}{2}(k_f(\sigma_{p,S})\, p_{\text{Lac}}\, R)\\ &+\frac{h_o}{PL_S}\left(\tau\mathbf{1}_{\sigma_{p,S}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{p,S}>\sigma_0}\right).\end{aligned}$$

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

$$\begin{aligned}\dot{\sigma}_{t,G} = &-\frac{\Omega}{2}\left(k_{cat}(\sigma_{t,G})\, EC_G\right)\\ &+\frac{h_o}{TL_G}\left(\tau\mathbf{1}_{\sigma_{t,G}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,G}>\sigma_0}\right),\\ \dot{\sigma}_{p,G} = &-\frac{\Omega}{2}(k_f(\sigma_{p,G})p_{\text{Tet}}\, R)\\ &+\frac{h_o}{PL_G}\left(\tau\mathbf{1}_{\sigma_{p,G}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{p,G}>\sigma_0}\right).\end{aligned}$$

Thus far, our derivation of the supercoiling state dynamics for the promoter and transcript regions are strictly the consequences of considering transcription induced-supercoiling for a single gene. However, in the convergent orientation, an additional perturbation of positive supercoiling impacts each gene transcript when the adjacent gene undergoes transcription. We model these perturbations as $\Delta^C_{G,S}$, $\Delta^C_{S,G}$ where $\Delta_{i,j}$ denotes the change in the supercoiling density in the transcript region of gene i from one transcription event of gene j. Using identical arguments as before we calculate the change in the linking number of transcript MG per transcription event of mSpinach to be $\frac{17}{2h_0}\frac{TL_S}{17}$. However, this perturbation is distributed across the transcript of MG, as well as the intergenic spacing region between the MG and mSpinach genes. Since the nature of the distribution for additional positive supercoiling turns is currently unknown and difficult to characterize experimentally, we model it using a distribution reflecting maximal uncertainty, i.e. a maximum entropy distribution. For this scenario, the maximum entropy distribution is the uniform distribution. Specifically, we assume the change in linking number is distributed uniformly across the region of interest.

The perturbation to the MG transcript region supercoiling density per transcription event of a molecule of mSpinach is thus given as

$$\Delta^C_{G,S} = \frac{TL_S}{2h_0}\frac{h_0}{TL_G + N_S} = \frac{TL_S}{2(TL_G + N_S)}$$

where N_S is the length of the intergenic spacing between the two constructs. Similarly, the perturbation term $\Delta^C_{S,G} = TL_G\backslash(2(TL_S + N_S))$. Thus, we can calculate the effect of mSpinach on MG RNA aptamer expression as $\Delta^C_{G,S}mS^c$ and vice versa, $\Delta^C_{S,G}MG^c$. Note that mS^c and MG^c are the discrete molecular counts of total mSpinach and MG RNA aptamer produced in the system respectively. Therefore, to include these terms in the dynamics of the supercoiling density states, we scale by volume to write them as state variables (defined as concentrations) and obtain the following expressions for the transcript supercoiling state dynamics of the mSpinach and MG genes:

$$\begin{aligned}\dot{\sigma}_{t,S} = &\frac{\Omega}{2}k_{cat}\left(\frac{TL_G}{TL_S + N_S}EC_G - EC_S\right)\\ &+\frac{h_o}{TL_S}\left(\tau\mathbf{1}_{\sigma_{t,S}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,S}>\sigma_0}\right),\\ \dot{\sigma}_{t,G} = &\frac{\Omega}{2}k_{cat}\left(\frac{TL_S}{TL_G + N_S}EC_S - EC_G\right)\\ &+\frac{h_o}{TL_G}\left(\tau\mathbf{1}_{\sigma_{t,G}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,G}>\sigma_0}\right).\end{aligned}$$

As mentioned above, transcription initiation k_f peaks when the supercoiling density is closest to $\sigma_0 = -0.65$. Indeed, the supercoiling density state across genes has been argued to be a form of global gene regulation [22], [23]. In [24], 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. Following this line of reasoning, we thus postulate that the transcription initiation rates can be approximated by a repression-based Hill function of the form:

$$k_{f,X}(t) = \zeta\frac{1}{|\sigma_{p,X}(t) - \sigma_0| + 1}, \tag{1}$$

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 the elongation/catalytic rates are defined by the functions

$$k_{cat,X}(t) = \frac{\beta}{TL_X}\frac{1}{|\sigma_{t,X}(t) - \sigma_0| + 1}, \tag{2}$$

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 the DNA and RNAP are constant in our *in vitro system*

$$\begin{aligned} R^{tot} &= R + EC_S + EC_G, \\ p_{\text{Lac}}^{tot} &= p_{\text{Lac}} + EC_S, \\ p_{\text{Tet}}^{tot} &= p_{\text{Tet}} + EC_G. \end{aligned}$$

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

$$\begin{aligned} \dot{mS} &= k_{cat,S}(\sigma_{t,S})EC_S - \delta_m mS, \\ \dot{MG} &= k_{cat,G}(\sigma_{t,G})EC_G - \delta_m MG, \\ \dot{EC_S} &= k_f(\sigma_{p,S})(R^{tot} - EC_S - EC_G)(p_{\text{Lac}}^{tot} - EC_S) \\ &\quad - (k_r + k_{cat}(\sigma_{t,S}))EC_S, \\ \dot{EC_G} &= k_f(\sigma_{p,G})(R^{tot} - EC_S - EC_G)(p_{\text{Tet}}^{tot} - EC_G) \\ &\quad - (k_r + k_{cat}(\sigma_{t,G}))EC_G, \\ \dot{\sigma}_{t,S} &= \frac{\Omega}{2}\left(\frac{TL_G}{TL_S + N_S}k_{cat,G}EC_G - k_{cat,S}EC_S\right) \\ &\quad + \frac{h_o}{TL_S}\left(\tau \mathbf{1}_{\sigma_{t,S}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{t,S}>\sigma_0}\right), \\ \dot{\sigma}_{t,G} &= \frac{\Omega}{2}\left(\frac{TL_S}{TL_G + N_S}k_{cat,S}EC_S - k_{cat,G}EC_G\right) \\ &\quad + \frac{h_o}{TL_G}\left(\tau \mathbf{1}_{\sigma_{t,G}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{t,G}>\sigma_0}\right), \\ \dot{\sigma}_{p,S} &= -\frac{\Omega}{2}(k_f(\sigma_{p,S})\, p_{\text{Lac}}\, R) \\ &\quad + \frac{h_o}{PL_S}\left(\tau \mathbf{1}_{\sigma_{p,S}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{p,S}>\sigma_0}\right), \\ \dot{\sigma}_{p,G} &= -\frac{\Omega}{2}(k_f(\sigma_{p,G})p_{\text{Tet}}\, R), \\ &\quad + \frac{h_o}{PL_G}\left(\tau \mathbf{1}_{\sigma_{p,G}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{p,G}>\sigma_0}\right). \end{aligned} \tag{3}$$

B. Divergent Orientation Model

In the divergent case, the chemical reaction network is identical as in the convergent case. However, since the promoters point away from each other and backpropagate negative supercoils whenever transcription initiation occurs, the supercoiling states of the promoters experience an additive negative perturbation while the supercoiling states for the transcript regions are unperturbed (any negative supercoiling can diffuse freely into the plasmid backbone). Thus, the divergent orientation supercoiling states $\sigma_{t,S}, \sigma_{t,G}$ have the same unperturbed dynamics as in the convergent case. Additionally, the promoters collect negative supercoiling from any transcription elongation. Therefore, the dynamics of $\sigma_{p,S}, \sigma_{p,G}$ are perturbed.

We use the notation $\Delta_{G,S}^D$ and $\Delta_{S,G}^D$ to signify the directed perturbation terms for the divergently oriented promoters' supercoiling states. This time, we argue that each mSpinach transcription initiation event backpropagates $-\frac{PL_S}{2h_0}$ negative turns while a MG transcription initiation event backpropagates $-\frac{PL_G}{2h_0}$ negative turns. Similarly, each transcription elongation event backpropagates $-\frac{TL_S}{2h_0}$ turns for mSpinach and $-\frac{TL_G}{2h_0}$ for MG aptamer. Thus, scaling by volume again, the perturbation at time t to the dynamics of $\sigma_{p,G}$ is given as

$$\begin{aligned} -\Omega\Delta_{G,S}^D k_f(\sigma_{p,S})p_{\text{Lac}}R &= -\Omega\left(\frac{PL_S}{2(PL_G + n_S)}k_f p_{\text{Lac}}\, R\right. \\ &\quad \left. + \frac{k_{cat,S}TL_S EC_S}{2(PL_G + PL_S + NS + TL_S)}\right) \\ &= \frac{-\Omega}{2}\left(k_f p_{\text{Lac}}\, R\, \frac{PL_S}{(PL_G + n_S)}\right. \\ &\quad \left. + \frac{k_{cat,S}EC_S TL_S}{PL_G + PL_S + NS + TL_S}\right). \end{aligned}$$

and by algebraic symmetry, the perturbation at time t to the dynamics of $\sigma_{p,S}$ is given as

$$\begin{aligned} -\Omega\Delta_{S,G}^D k_f(\sigma_{p,G})p_{\text{Tet}}\, R &= -\Omega\frac{PL_G}{2h_0}\frac{h_0}{PL_S + n_S}k_f p_{\text{Lac}}\, R \\ &= \frac{-\Omega}{2}\left(k_f p_{\text{Tet}}\, R\frac{PL_G}{(PL_S + n_S)}\right. \\ &\quad \left. + \frac{k_{cat,G}EC_G TL_G}{TL_G + PL_S + PL_G + n_S}\right). \end{aligned}$$

and the perturbed dynamics of $\sigma_{p,S}$ and $\sigma_{p,G}$ are given as

$$\begin{aligned} \dot{\sigma}_{p,S} &= -\frac{\Omega}{2}\left(k_f(\sigma_{p,S})p_{\text{Lac}} + \frac{k_f(\sigma_{p,G})p_{\text{Tet}}PL_G}{(PL_S + n_S)}\right. \\ &\quad \left. + \frac{k_{cat,S}EC_S TL_S}{R(TL_S + PL_G + PL_S + NS)}\right)R \\ &\quad + \frac{h_o}{PL_S}\left(\tau \mathbf{1}_{\sigma_{p,S}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{p,S}>\sigma_0}\right), \\ \dot{\sigma}_{p,G} &= -\frac{\Omega}{2}\left(k_f(\sigma_{p,G})p_{\text{Tet}} + \frac{k_f(\sigma_{p,S})p_{\text{Lac}}PL_S}{(PL_G + n_G)}\right. \\ &\quad \left. + \frac{k_{cat,G}EC_G TL_G}{R(TL_S + PL_S + PL_G + n_S)}\right)R \\ &\quad + \frac{h_o}{PL_G}\left(\tau \mathbf{1}_{\sigma_{p,G}<\sigma_0} - \gamma \mathbf{1}_{\sigma_{p,G}>\sigma_0}\right). \end{aligned}$$

while the other four state equations are as in the convergent model.

C. Tandem Orientation Model

In the tandem orientation, the two genes are adjacent in a way so that positive supercoiling propagates downstream first into the mSpinach promoter region and then the actual coding sequence of mSpinach. At the same time, negative supercoiling propagates upstream into the coding sequence for MG RNA aptamer and then into the promoter region for MG aptamer. The backpropagation forms an R-loop which stalls the elongation process and thus represses MG expression. Additionally, in the tandem orientation, infrequent terminator read-through events may occur where the open complex continues elongation, leaking past the terminator and into adjacent coding sequences. This typically results in correlation of downstream expression with upstream genes.

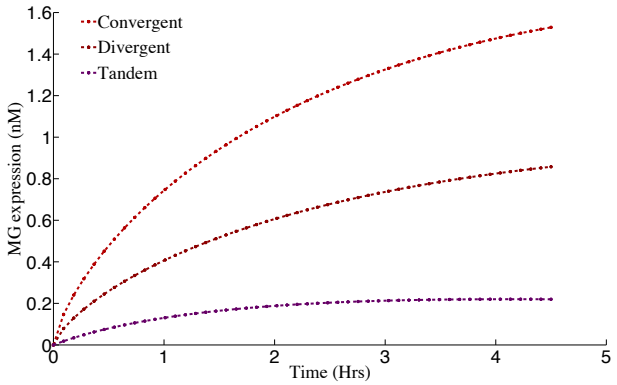

Fig. 4: A simulation of MG RNA aptamer expression in the convergent, divergent and tandem orientation model systems. Parameters for the simulation were specified as follows: $\delta_m = .05\,s^{-1}$ mRNA degradation is slightly slower than [25] because of aptamer structure and tRNA scaffold in mSpinach $R^{tot} = 1\mu\,M, p_{\text{Tet}}^{tot} = p_{\text{Lac}}^{tot} = 11\,nM$, $k_r = .01\,s^{-1}$, $\Omega = \frac{V_R}{V_C}$ where V_R is the reaction volume and V_C is the volume of a single cell (since parameters are quantified in an *in vivo* context with reaction volume V_C). $TL_S = 141$ bp [18], $TL_G = 38$ bp [17], $NS = 50$ bp, $PL_S = 40$ bp, $PL_G = 44$ bp as per DNA synthesis, $\zeta = 1\times10^4\,nM^{-1}s^{-1}$, $\beta = 5.4\times10^5 s^{-1}$, $k_l = .02\,s^{-1}$ [20], $k_{seq,max} = 1\,s^{-1}$, and $k_w = 1\,s^{-1}$.

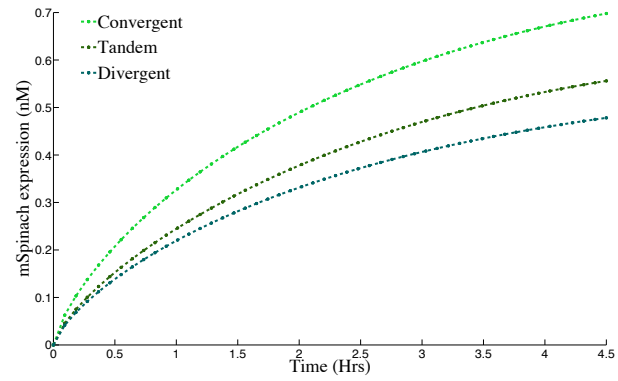

Fig. 5: A simulation of mSpinach RNA aptamer expression in the convergent, divergent, and tandem orientation. Parameters for the simulation are provided in Figure 4.

We model both of these phenomena with the following chemical reaction system:

$$R + p_{\text{Lac}} \underset{k_r}{\overset{k_f(\sigma_{p,S})}{\rightleftharpoons}} EC_S \xrightarrow{k_{cat}(\sigma_{t,S})} mS + R + p_{\text{Lac}}$$
$$R + p_{\text{Tet}} \underset{k_r}{\overset{k_f(\sigma_{p,G})}{\rightleftharpoons}} EC_G \xrightarrow{k_{cat}(\sigma_{t,G})} MG + R + p_{\text{Tet}}$$
$$mS \xrightarrow{\delta_m} \emptyset$$
$$MG \xrightarrow{\delta_m} \emptyset$$
$$EC_G \xrightarrow{k_{seq}(\sigma_{t,G})} RL : EC_G$$
$$EC_G \xrightarrow{k_l} EC_G : EC_S$$
$$EC_G : EC_S \xrightarrow{k_w} m_S + MG + R + p_{\text{Lac}} + p_{\text{Tet}}$$
$$k_{seq}(\sigma_{t,G}) \equiv k_{seq,max}\mathbf{1}_{\sigma_{t,G}<\sigma_0}\frac{|\sigma_{t,G}-\sigma_0}{1+|\sigma_{t,G}-\sigma_0|}$$

where $EC_G : EC_S$ is the read-through complex formed from terminator read-through during MG transcription, $RL : EC$ denotes the R-Loop complex formed by hyper-negatively supercoiled DNA in the transcript region and nascent mRNA strand. Moreover, the supercoiling state of the mSpinach promoter and MG transcript region depend on each others' dynamics, while the supercoiling states of MG promoter and mSpinach transcript are independent. Thus, the dynamics of $\sigma_{p,G}$ and $\sigma_{t,S}$ are given as:

$$\begin{aligned}\dot{\sigma}_{t,S} &= -\frac{\Omega}{2}\left(k_{cat}(\sigma_{t,S})EC_S\right) \\ &\quad + \frac{h_o}{TL_S}\left(\tau\mathbf{1}_{\sigma_{t,S}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,S}>\sigma_0}\right), \\ \dot{\sigma}_{p,G} &= -\frac{\Omega}{2}(k_f(\sigma_{p,G})p_{\text{Tet}}\,R) \\ &\quad + \frac{h_o}{PL_G}\left(\tau\mathbf{1}_{\sigma_{p,G}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{p,G}>\sigma_0}\right).\end{aligned}$$

Using similar arguments as in the convergent and divergent case, we can write the perturbed dynamics of $\sigma_{p,S}$ and $\sigma_{t,G}$ as

$$\begin{aligned}\dot{\sigma}_{p,S} &= \frac{\Omega}{2}\left(k_{cat}(\sigma_{p,G})EC_G\frac{TL_G}{2(PL_S+n_S)}\right. \\ &\quad \left. -k_f(\sigma_{p,S})p_{\text{Lac}}\,R - k_{cat}(\sigma_{T,S})EC_S\frac{TL_S}{PL_S+n_S}\right) \\ &\quad + \frac{h_o}{PL_S}\left(\tau\mathbf{1}_{\sigma_{p,S}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{p,S}>\sigma_0}\right), \\ \dot{\sigma}_{t,G} &= -\frac{\Omega}{2}\left(k_{cat}(\sigma_{t,S})EC_S\frac{TL_S}{PL_S+n_S+TL_G+TL_S}\right. \\ &\quad \left. +k_{cat}(\sigma_{t,G})EC_G + k_f(\sigma_{p,S})p_{\text{Lac}}\,R\frac{PL_S}{2(TL_G+nS)}\right) \\ &\quad + \frac{h_o}{TL_G}\left(\tau\mathbf{1}_{\sigma_{t,G}<\sigma_0} - \gamma\mathbf{1}_{\sigma_{t,G}>\sigma_0}\right).\end{aligned}$$

Due to the presence of additional complexes, the modified conservation laws are given as:

$$\begin{aligned}R^{tot} &= R + EC_S + EC_G + EC_G : EC_S, \\ p_{\text{Lac}}^{tot} &= p_{Lac} + EC_S + EC_G : EC_S, \\ p_{\text{Tet}}^{tot} &= p_{Tet} + EC_G + EC_G : EC_S.\end{aligned}$$

The remaining state dynamics are given as:

$$\begin{aligned}\dot{mS} &= k_{cat}(\sigma_{t,S})EC_S + k_w EC_G : EC_S - \delta_m mS, \\ \dot{MG} &= k_{cat}(\sigma_{t,G})EC_G + k_w EC_G : EC_S - \delta_m MG, \\ \dot{EC_S} &= k_f(\sigma_{p,S})(R^{tot} - X)(p_{\text{Lac}}^{tot} - EC_S - EC_G : EC_S), \\ &\quad - (k_r + k_{cat}(\sigma_{t,S}))EC_S \\ \dot{EC_G} &= k_f(\sigma_{p,G})(R^{tot} - X)(p_{\text{Tet}}^{tot} - EC_G - EC_G : EC_S), \\ &\quad - (k_r + k_{cat}(\sigma_{t,G}) + k_{seq}(\sigma_{t,G}) + k_l)EC_G, \\ \dot{EC_G : EC_S} &= k_l EC_G - k_w EC_G : EC_S,\end{aligned}$$

where $X = EC_S + EC_G + EC_G : EC_S$.

The results of a simulation for all three model systems are shown in Figures 4 and 5. We see that the model is able to recapitulate the trends in the experimental data, in particular, it shows that the convergentl construct obtained nearly two-fold the expression of the divergent construct, in both the mSpinach and MG fluorescence channel. Furthermore, the simulation shows that the tandem orientation achieves a level of mSpinach expression intermediate to the levels achieved by the divergently oriented system. In the MG fluorescence channel, the simulation faithfully shows that MG RNA aptamer expression in the tandem construct is practically abolished, consistent with experimental results.

IV. Conclusion and Future Work

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, compositional context should be a chief consideration during the design and prototyping process.

Our future research will involve experiments to further validate our modeling framework, investigate the role of intergenic spacing length on gene expression in regards to compositional context and how transcriptional regulator-regulatee relationships between adjacent genes vary depending on orientation. Our hope is that this research represents yet another step in the direction of a standardized set of engineering protocols for building complex biocircuits.

V. Acknowledgments

This material is based upon work supported in part by a National Science Foundation Graduate Fellowship, a National Defense Science and Engineering Fellowship and the Defense Advanced Research Projects Agency (DARPA/MTO) Living Foundries program, contract number HR0011-12-C-0065 (DARPA/CMO). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

References

[1] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline", *Molecular systems biology*, vol. 2, no. 1, 2006.

[2] C. Engler, R. Kandzia, and S. Marillonnet, "A one pot, one step, precision cloning method with high throughput capability", *PloS one*, vol. 3, no. 11, pp. e3647, 2008.

[3] D. G. Gibson, L. Young, R-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of dna molecules up to several hundred kilobases", *Nature methods*, vol. 6, no. 5, pp. 343–345, 2009.

[4] Z. Z. Sun, E. Yeung, C. A Hayes, V. Noireaux, and R. M. Murray, "Linear dna for rapid prototyping of synthetic biological circuits in an escherichia coli based tx-tl cell-free system", *ACS synthetic biology*, vol. 3, no. 6, pp. 387–397, 2014.

[5] J. Chappell, K. Jensen, and P. S. Freemont, "Validation of an entirely in vitro approach for rapid prototyping of dna regulatory elements for synthetic biology", *Nucleic acids research*, vol. 41, no. 5, pp. 3471–3481, 2013.

[6] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, "Genetic programs constructed from layered logic gates in single cells", *Nature*, vol. 491, no. 7423, pp. 249–253, 2012.

[7] J. AN Brophy and C. A. Voigt, "Principles of genetic circuit design", *Nature Methods*, vol. 11, no. 5, pp. 508–520, 2014.

[8] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag, "Modular cell biology : retroactivity and insulation", *Mol. Syst. Biol*, vol. 4, pp. 161, 2008.

[9] A. Gyorgy and D. Del Vecchio, "Retroactivity to the input in complex gene transcription networks", in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, Dec., pp. 3595–3601.

[10] A. Gyorgy and D. Del Vecchio, "How slaves affect a master module in gene transcription networks", in *Proc. of IEEE Conference on Decision and Control*, 2013.

[11] T. Ellis, X. Wang, and J. J. Collins, "Diversity-based, model-guided construction of synthetic gene networks with predicted functions", *Nature biotechnology*, vol. 27, no. 5, pp. 465–471, 2009.

[12] N. Cookson et al, "Queueing up for enzymatic processing: correlated signaling through coupled degradation", *Molecular Systems Biology*, vol. 7, no. 561, 2011.

[13] S. Cardinale and A. Arkin, "Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems", *Biotechnology Journal*, vol. 7, no. 7, pp. 856–866, 2012.

[14] J. O. Korbel, L. J. Jensen, C. Von Mering, and P. Bork, "Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs", *Nature biotechnology*, vol. 22, no. 7, pp. 911–917, 2004.

[15] M. L. Opel and G. Hatfield, "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", *Molecular microbiology*, vol. 39, no. 1, pp. 191–198, 2001.

[16] Y-J. Chen, P. Liu, A. AK Nielsen, J. AN Brophy, K. Clancy, T. Peterson, and C. A. Voigt, "Characterization of 582 natural and synthetic terminators and quantification of their design constraints", *Nature methods*, vol. 10, no. 7, pp. 659–664, 2013.

[17] J. R. Babendure, S. R. Adams, and R. Y. Tsien, "Aptamers switch on fluorescence of triphenylmethane dyes", *Journal of the American Chemical Society*, vol. 125, no. 48, pp. 14716–14717, 2003.

[18] J. S. Paige, K. Y. Wu, and S. R. Jaffrey, "Rna mimics of green fluorescent protein", *Science*, vol. 333, no. 6042, pp. 642–646, 2011.

[19] J. Shin and V. Noireaux, "An e. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells", *ACS synthetic biology*, vol. 1, no. 1, pp. 29–41, 2012.

[20] M. H. Larson, W. J. Greenleaf, R. Landick, and S. M. Block, "Applied force reveals mechanistic and energetic details of transcription termination", *Cell*, vol. 132, no. 6, pp. 971–982, 2008.

[21] L. F. Liu and J. C. Wang, "Supercoiling of the dna template during transcription", *Proceedings of the National Academy of Sciences*, vol. 84, no. 20, pp. 7024–7027, 1987.

[22] A. R. Rahmouni and R. D. Wells, "Direct evidence for the effect of transcription on local dna supercoiling in-vivo", *Journal of molecular biology*, vol. 223, no. 1, pp. 131–144, 1992.

[23] M. Drolet, "Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, r-loop formation and dna topology", *Molecular microbiology*, vol. 59, no. 3, pp. 723–730, 2006.

[24] V. L. Balke and J. D. Gralla, "Changes in the linking number of supercoiled dna accompany growth transitions in escherichia coli.", *Journal of bacteriology*, vol. 169, no. 10, pp. 4499–4506, 1987.

[25] Y. Taniguchi, P. J. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells", *Science*, vol. 329, no. 5991, pp. 533–538, 2010.