Submitted, IEEE Biomedical Circuits and Systems (BioCAS) Conference
http://www.cds.caltech.edu/~murray/papers/km11-biocas.html

Analysis and design of a synthetic transcriptional network for exact adaptation

Jongmin Kim* and Richard M. Murray*†
Departments of * Bioengineering, † Control and Dynamical Systems, California Institute of Technology,
Pasadena, CA 91125, USA Email: murray@cds.caltech.edu

***Abstract*— This paper presents a mathematical model for a synthetic transcriptional regulatory network *in vitro*. This circuit design resembles one of the well-known network motifs, the incoherent feed-forward loop, in which an activator regulates both a gene and a repressor of the gene. Through mathematical analysis, we show how the circuit can be controlled to demonstrate exact adaptation to input signals.**

I. Introduction

Living organisms and cells use their sensory systems to create new knowledge of their environment. One of the common features found in many sensory systems is exact adaptation in which the output upon change of input to a new constant level gradually returns to a steady level independent of the input [1], [2]. A well-known example is bacterial chemotaxis in which bacteria in search of nutrient sources are sensitive to spatial gradient but insensitive to attractant source levels. Previous studies uncovered gene regulatory networks composed of recurring interaction patterns called network motifs that form the basis of dynamic regulatory behavior of cells [3]. Understanding such network motifs underlying the exquisite dynamics of sensory systems is an important task. One network motif, the incoherent feed-forward loop (IFFL), is of particular interest because it can generate diverse dynamic features such as a temporal pulse, a band-pass filter, a fold-change detector including exact adaptation behavior [4], [5], [6]. In this work, we focus on the characteristics of an IFFL as a motif for exact adaptation and how this motif can be implemented in a synthetic *in vitro* transcription system. This exercise will highlight the flexibility and modularity of *in vitro* transcription system and the ease with which we can program biochemical systems in cell-free environment. In the future, the IFFL motif can be embedded in a regulatory network controlling artificial cells (e.g. liposomes programmed to release drugs) for biomedical applications.

II. Incoherent Feed-forward Loop (IFFL)

A type-1 IFFL is a regulatory pattern in which an input u serving as an activator controls a target gene y and also activates a repressor of that target gene, x (Figure 1(a)) [3]. For IFFL found in biological networks, the repressor and the activator are often transcription factors. We note that the abstract diagram in Figure 1(a) dictates a topological regulatory relationship rather than exact mathematical terms. In fact, several sets of equations with diverse parameter choices can be summarized by the same abstract diagram above. For

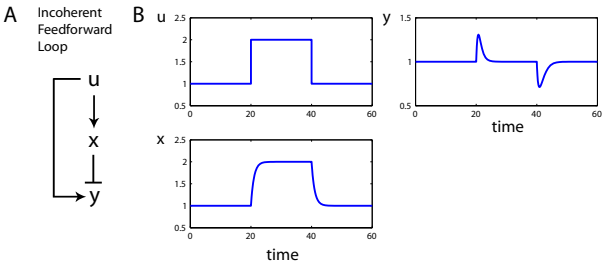

Fig. 1. An IFFL can provide exact adaptation. a) Schematic diagram of an IFFL. b) Simulation results of the dynamics of input u (activator for x and y), intermediate node x (inhibitor for y), and output y (eqs. (1)). Time is in arbitrary units.

instance, we analyze an IFFL termed a 'sniffer' [4], in which x enhances the degradation of y rather than repressing the production of y. The dynamics of a sniffer can be described by the following set of equations:

$$\dot{x} = u - x, \quad \dot{y} = u - x \cdot y. \tag{1}$$

Numerical simulation of this set of equations show an exact adaptation for y (Figure 1(b)). By setting $\dot{x} = 0$, the steady-state x value $x_s = u$ for all input $u > 0$. Letting $\dot{y} = 0$ with $x_s = u$, we have $y_s = 1$ for all input $u > 0$ as required for an exact adaptation.

As a generalized model of a 'sniffer', we consider the following set of equations:

$$\dot{x} = \alpha_1 \cdot u - \beta_1 \cdot x - k \cdot x \cdot y, \tag{2}$$

$$\dot{y} = \alpha_2 \cdot u - \beta_2 \cdot y - k \cdot x \cdot y. \tag{3}$$

Here, the α_1 and α_2 terms reflect generic asymmetry in the effectiveness of u as an activator for x and y. The β_1 and β_2 terms reflect generic degradation for x and y. One notable characteristic is the existence of the $k \cdot x \cdot y$ term for dynamics of both x and y: this type of accelerated degradation term can be chosen for the case in which x and y stoichiometrically react to annihilate each other. Assume that all constants are positive (α_1, α_2, β_1, β_2, $k > 0$). By setting $\dot{x} = 0$ and $\dot{y} = 0$, we have two relations for the steady-state values of x and y:

$$x_s = \frac{\alpha_1 \cdot u}{\beta_1 + k \cdot y_s}, \quad y_s = \frac{\alpha_2 \cdot u}{\beta_2 + k \cdot x_s}, \tag{4}$$

for all input $u > 0$. Let us investigate when the output y shows exact adaptation. If y_s is independent of the choice of u, we

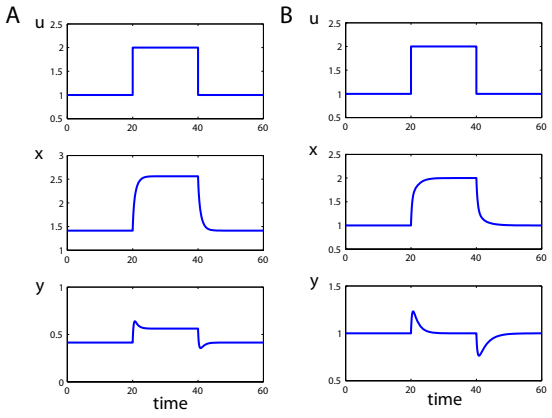

Fig. 2. Simulation results of the dynamics of eqs. (2) and (3). Whether this IFFL can provide exact adaptation depends on the choice of β_2. The parameter values are: $\alpha_1 = 2$, $\alpha_2 = 1$, $\beta_1 = 1$, $k = 1$, $\beta_2 = 1$ for a) and $\beta_2 = 0$ for b). Time is in arbitrary units.

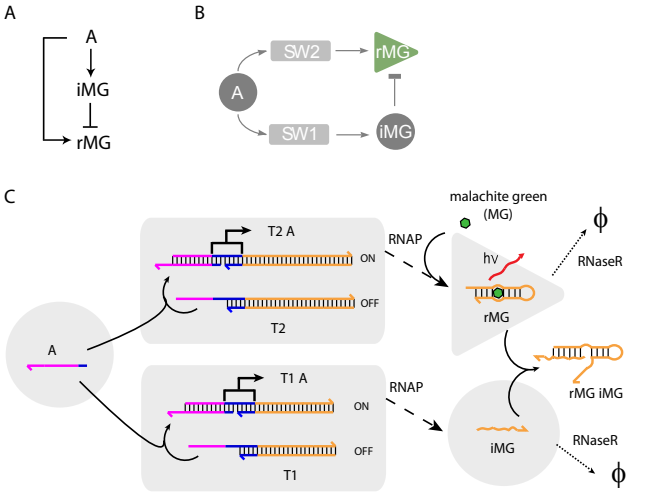

Fig. 3. a) An IFFL formed by transcriptional circuit components. b) Schematic block diagram of the transcriptional circuit components. c) Detailed molecular reactions in the transcriptional circuit as shown in b).

find that x_s is proportional to u. However, the numerator of y_s $(\alpha_2 \cdot u)$ changes linearly with u, while the denominator of y_s $(\beta_2 + k \cdot x_s)$ changes sublinearly with u. Thus, an exact adaptation is not possible when all constants are positive. On the other hand, if β_2 is zero with other constants being positive, the numerator and denominator for y_s are both proportional to u, satisfying the requirement of exact adaptation (if $\alpha_1 > \alpha_2$). The steady-state solutions are as follows:

$$x_s = \frac{(\alpha_1 - \alpha_2)u}{\beta_1}, \quad y_s = \frac{\alpha_2 \cdot \beta_1}{k(\alpha_1 - \alpha_2)}. \tag{5}$$

In Figure 2, numerical simulations of eqs. (2) and (3) are shown with different values of β_2. When β_2 is positive, the output y does not show exact adaptation (Figure 2(a)), while it shows exact adaptation with β_2 being zero (Figure 2(b)).

III. Design for a Transcriptional Circuit

A. Components of a transcriptional circuit

1) Transcriptional switches: We first introduce a synthetic *in vitro* switch design and lay groundwork for general circuit construction methods. The OFF state of the switch consists of a double-stranded (ds) DNA template ("T") with a partially single-stranded (ss) and thus incomplete T7 RNA polymerase (RNAP) promoter region. The switch is turned ON by the binding of a ssDNA activator ("A") that completes the RNAP promoter region (activation reaction). The resulting template ("T·A") has a nicked promoter but still transcribes well [7] (cf. Figure 3(c)). To provide a sharp threshold of activation, an inhibitor strand (ssDNA, "I") can bind to a complementary free-floating activator A, resulting in a functionally inert activator-inhibitor complex "A·I." (This reaction is not explicitly implemented for the circuit presented in this paper.) Using these design motifs for switches and signals, networks with arbitrary connectivity can be constructed modularly [7], [8], [9]. In principle, transcriptional circuits can be wired as continuous-time analog neural networks [8]. In a typical reaction network, RNA outputs will be produced by RNAP from upstream switches using NTP as fuel, and these outputs will serve as inputs for downstream switches. At the same time, the degradation of RNA signals by *Escherichia coli* ribonuclease H (RNase H) and/or ribonuclease R (RNase R) removes RNA signal, thereby undoing the regulatory effect of the RNA inputs (cf. Figure 3(c)).

2) Malachite Green aptamer: As an example of functional RNA that is amenable to real-time monitoring, we take an RNA aptamer for the chromophore malachite green (MG) as our output [10]. MG aptamer consists of a short RNA sequence whose central loop region serves as the binding pocket of MG. When MG is bound to the aptamer, it becomes highly fluorescent (cf. Figure 3(c)). However, any sequence change in the loop of the aptamer almost completely abolishes MG binding affinity. From our previous experiences, MG does not affect enzyme and nucleic acid hybridization reactions in *in vitro* circuits.

B. Transcriptional circuit design for IFFL

Using the synthetic transcriptional switch as the regulatory motif and MG aptamer as the output signal, we construct an IFFL. Because the central loop region of MG aptamer is the binding pocket for MG, disrupting the loop structure abolishes binding of MG to aptamer. Let's label the MG aptamer RNA transcript as "rMG" and the RNA transcript that complements part of rMG and opens loop structure as "iMG", short for inhibitor of rMG. If we want to construct an IFFL as shown in Figure 3(a), it is straightforward to design two transcriptional switches that share common input domains such that they are both activated by a single DNA activator "A" and transcribe different outputs, rMG and iMG (Figure 3(b)). (Detailed molecular reactions are shown in Figure 3(c).) Table 1 shows the list of hybridization and branch migration reactions and enzyme reactions. Here, we do not consider side-reactions or incomplete production and degradation products.

TABLE I

REACTION PATHWAYS FOR THE TRANSCRIPTIONAL CIRCUIT

Reaction type	Reaction		
Activation	T1 + A	$\xrightarrow{k_+}$	T1·A
Activation	T2 + A	$\xrightarrow{k_+}$	T2·A
Production	T1·A	$\xrightarrow{k_p}$	T1·A + iMG
Production	T2·A	$\xrightarrow{k_p}$	T2·A + rMG
Degradation	iMG	$\xrightarrow{\beta_1}$	ϕ
Degradation	rMG	$\xrightarrow{\beta_2}$	ϕ
Inhibition	rMG + iMG	$\xrightarrow{k}$	rMG·iMG

Further, enzyme reactions are treated as approximately first-order reactions. The dynamics of the *in vitro* circuit is described by the following four ordinary differential equations:

$$\dot{[\mathrm{T1A}]} = k_+[\mathrm{T1}][\mathrm{A}], \quad (6)$$

$$\dot{[\mathrm{T2A}]} = k_+[\mathrm{T2}][\mathrm{A}], \quad (7)$$

$$\dot{[\mathrm{iMG}]} = k_p[\mathrm{T1A}] - \beta_1[\mathrm{iMG}] - k[\mathrm{iMG}][\mathrm{rMG}], \quad (8)$$

$$\dot{[\mathrm{rMG}]} = k_p[\mathrm{T2A}] - \beta_2[\mathrm{rMG}] - k[\mathrm{iMG}][\mathrm{rMG}]. \quad (9)$$

The system preserves the conservation relation, $[\mathrm{T}i^{\mathrm{tot}}]$ = $[\mathrm{T}i]$ + $[\mathrm{T}i\mathrm{A}]$, and similarly for $[\mathrm{A}^{\mathrm{tot}}]$, where the superscript tot indicates that all species involving the given strands are being counted. Using these conserved quantities, the remaining variables, [T1], [T2] and [A], are directly calculated from the concentrations of other species. Let $u = [\mathrm{A}^{\mathrm{tot}}]$, $x = [\mathrm{iMG}]$, and $y = [\mathrm{rMG}]$. Then, we obtain the same set of ODEs as in eqs. (2) and (3) for eqs. (8) and (9) as follows:

$$\dot{x} = \alpha_1 \cdot u - \beta_1 \cdot x - k \cdot x \cdot y, \quad (10)$$

$$\dot{y} = \alpha_2 \cdot u - \beta_2 \cdot y - k \cdot x \cdot y, \quad (11)$$

where α_1 and α_2 are functions of k_p, $[\mathrm{T1}^{\mathrm{tot}}]$, and $[\mathrm{T2}^{\mathrm{tot}}]$. In the next section, we discuss how to tune parameters to achieve desired dynamic behaviors.

C. Tuning parameter values

1) Production rates (α_1 and α_2): Since T1 and T2 have identical input domains, the activator A will bind to T1 and T2 with the same affinity. We also expect that the binding reaction will be fast and practically irreversible because of large gain in thermodynamic energy upon binding. If $[\mathrm{T1}^{\mathrm{tot}}] + [\mathrm{T2}^{\mathrm{tot}}] > [\mathrm{A}^{\mathrm{tot}}]$, at steady-state, all activator will be bound to either of the templates: $[\mathrm{A}] \simeq 0$ and $[\mathrm{T1A}] + [\mathrm{T2A}] \simeq [\mathrm{A}^{\mathrm{tot}}]$. Thus, $[\mathrm{T1A}] = [\mathrm{A}^{\mathrm{tot}}]\frac{[\mathrm{T1}^{\mathrm{tot}}]}{[\mathrm{T1}^{\mathrm{tot}}]+[\mathrm{T2}^{\mathrm{tot}}]}$ and $[\mathrm{T2A}] = [\mathrm{A}^{\mathrm{tot}}]\frac{[\mathrm{T2}^{\mathrm{tot}}]}{[\mathrm{T1}^{\mathrm{tot}}]+[\mathrm{T2}^{\mathrm{tot}}]}$ after transient change in the concentration of A. Therefore, we can calculate α_1 and α_2: $\alpha_1 = k_p\frac{[\mathrm{T1}^{\mathrm{tot}}]}{[\mathrm{T1}^{\mathrm{tot}}]+[\mathrm{T2}^{\mathrm{tot}}]}$ and $\alpha_2 = k_p\frac{[\mathrm{T2}^{\mathrm{tot}}]}{[\mathrm{T1}^{\mathrm{tot}}]+[\mathrm{T2}^{\mathrm{tot}}]}$. We can tune the parameter ratio α_1/α_2 by adjusting template concentrations and scale both parameters by adjusting RNAP concentrations that control k_p.

2) Degradation rates (β_1 and β_2): Here, we focus on the use of RNase R as the enzyme for degradation. RNase R is a processive, 3′ to 5′ hydrolytic exoribonuclease that plays an important role in the degradation of structured RNAs [11]. An interesting feature of RNase R distinct from other RNases is that it can by itself degrade RNAs with extensive secondary structure provided that a single-stranded 3′ overhang is present. Duplex RNAs with no overhang or with only a 4-nt 3′ overhang bind to RNase R with a dissociation constant K_d ($K_d = k_{OFF}/k_{ON}$) greater than 5 μM, in stark contrast to single-stranded RNAs binding to RNase R with a K_d of 2 nM [11]. Therefore, employing different secondary structures for RNA substrates at their 3′ ends offers an opportunity to tune degradation rates by RNase R by about three orders of magnitudes. Fortunately, in our current *in vitro* circuit design, the fluorescent output molecule rMG has significant secondary structure with little or no overhang at its 3′ end, while the inhibitor for output iMG has no significant secondary structure. Therefore, we expect that the degradation rate for iMG (β_1) is roughly three orders of magnitude larger than the degradation rate for rMG (β_2) without extra sequence design efforts. Note that, even if the native RNA sequences do not provide such distinctive structural differences as in this example, it is always possible to extend the 3′ ends of the target RNA species in the system to tune the degradation rates by RNase R.

3) Inhibition rate (k): The inhibition rate k can be tuned by adjusting the "toehold", a single-stranded overhang beyond the helical domain of rMG to be placed on its 5′ end (to minimize the 3′ overhang that can be recognized by RNase R). Strand displacement reactions for various toehold lengths have been characterized: the completion rate of branch migration depends exponentially on toehold length [12]. Therefore, by adjusting the toehold length for rMG where iMG can initiate a branch migration process to open the loop of rMG, the inhibition rate k can span roughly six orders of magnitude from $1/\mathrm{M/s}$ to $10^6/\mathrm{M/s}$.

4) Input (u): We can always increase input u by simply adding more activator A into the reaction mixture. On the other hand, specifically removing certain species from the reaction mixture is non-trivial. Therefore, we rely on an inhibition strategy that renders A inert similar to the case for rMG and iMG. Although not explicitly shown in Table 1, we can add an inhibitor I that is fully complementary to the activator A to the system if we want to decrease u, in which case u should be reinterpreted as $u = [\mathrm{A}^{\mathrm{tot}}] - [\mathrm{I}^{\mathrm{tot}}]$. The introduced inhibitor I will stoichiometrically reduce the amount of A bound to the templates through toehold-mediated branch migration reactions [12].

IV. MATHEMATICAL ANALYSIS

A. Simulation results

For a plausible choice of parameters, the simulation results indicate that a dynamic behavior close to exact adaptation can be achieved for the output [rMG] (Figure 4(a)). We characterize steady-state fold change of output y ([rMG]) upon step increase of u ($[\mathrm{A}^{\mathrm{tot}}]$) by 2-fold: $F = y_{s(final)}/y_{s(initial)}$.

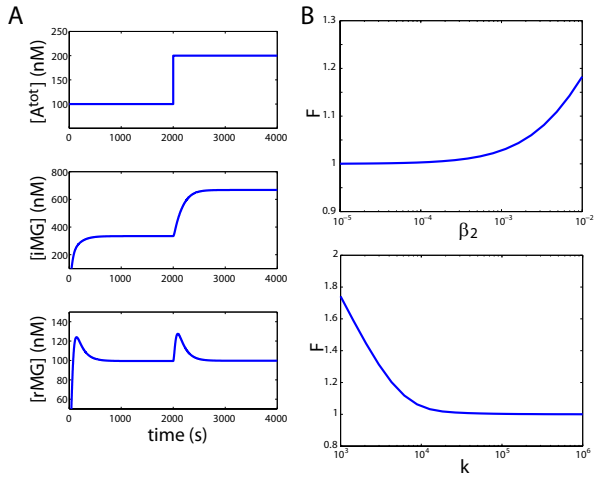

Fig. 4. a) Simulation results of the transcriptional circuit. The input $[\mathrm{A}^{\mathrm{tot}}]$ is increased from 100 nM to 200 nM at 2000s. The parameter values are: $[\mathrm{T1}^{\mathrm{tot}}] = 200$ nM, $[\mathrm{T2}^{\mathrm{tot}}] = 100$ nM, $k_+ = 10^5/\mathrm{M/s}$, $k_p = 0.1/\mathrm{s}$, $\beta_1 = 0.01/\mathrm{s}$, $\beta_2 = 0.0001/\mathrm{s}$, and $k = 10^5/\mathrm{M/s}$. b) The fold change response in [rMG] upon 2-fold change of $[\mathrm{A}^{\mathrm{tot}}]$ as functions of β_2 and k.

First, we analyze the effect of changing β_2. In practice, β_2 is not zero, but rather a few orders of magnitude slower than β_1. From the simulation results, we find that close to exact adaptation behavior ($F < 1.01$) can be achieved with $\beta_2 < 3.4{*}10^{-4}/\mathrm{s}$ (Figure 4(b)). Second, we analyze the effect of adjusting k. In contrast to degradation rates which can be tuned continuously by adjusting RNase R concentrations, the inhibition rate (k) can be tuned for several orders of magnitude but with difficulty in continuous fine adjustments [12]. We observe that close to exact adaptation behavior ($F < 1.01$) can be achieved with $k > 3{*}10^4/\mathrm{M/s}$ (Figure 4(c)). Together, the simulation results show that the exact adaptation behavior can be achieved for this *in vitro* circuit with reasonable parameters.

B. Limitations of the model

One of the limitations in the current dynamical model is the assumed first-order enzyme reactions. A more realistic model is Michaelis–Menten enzyme reactions, where the available enzyme concentrations can be calculated as follows:

$$[\mathrm{RNAP}] = \frac{[\mathrm{RNAP}^{\mathrm{tot}}]}{1 + \frac{[\mathrm{T1A}]}{K_M} + \frac{[\mathrm{T2A}]}{K_M}} \simeq \frac{[\mathrm{RNAP}^{\mathrm{tot}}]}{1 + \frac{[\mathrm{A}^{\mathrm{tot}}]}{K_M}}, \quad (12)$$

$$[\mathrm{RNase\ R}] = \frac{[\mathrm{RNase\ R}^{\mathrm{tot}}]}{1 + \frac{[\mathrm{iMG}]}{K_{M,R,1}} + \frac{[\mathrm{rMG}]}{K_{M,R,2}}} \simeq \frac{[\mathrm{RNase\ R}^{\mathrm{tot}}]}{1 + \frac{[\mathrm{iMG}]}{K_{M,R,1}}}, \quad (13)$$

where we assume high $K_{M,R,2}$ (e.g. $K_{M,R,2} > 5\ \mu\mathrm{M}$). The Michaelis constants, the affinity of substrates to the enzymes, are calculated as $K_M = \frac{k_{OFF}+k_{cat}}{k_{ON}}$. Note that the available enzyme concentrations are functions of their substrates; the available enzyme concentrations diminish as the concentrations of substrates increase. Then, the system dynamics can be rewritten as:

$$\dot{x} = \alpha_1 \cdot f(u) - \beta_1 \cdot g(x) - k \cdot x \cdot y, \quad (14)$$

$$\dot{y} = \alpha_2 \cdot f(u) - k \cdot x \cdot y, \quad (15)$$

where $f(u)$ and $g(x)$ are Michaelis–Menten type sub-linear functions of u and x as above. We note that, as long as $g(x)$ is roughly linear in x, the exact adaptation result still applies.

Another potentially important reaction not included in our current model is the degradation of rMG·iMG complex. The rMG·iMG complex have single-stranded 3′ overhangs for both rMG and iMG, making them good substrates for RNase R. Ideally, the two components of the complex (rMG·iMG) would degrade at similar speed, in which case the current dynamic equations that ignore the rMG·iMG complex after it is formed is valid. On the other hand, if one component, rMG or iMG, is degraded much faster than the other component of the complex, the released single-stranded RNA component provides an effective production term in the system dynamics. These limitations may require further refinement of models through careful experimental measurements.

V. Discussion

Through mathematical analysis, we were able to demonstrate that the *in vitro* transcriptional circuit designed to have an IFFL motif can provide exact adaptation behavior with reasonable parameter choices. We further outlined several strategies to tune parameters for experimental realization of exact adaptation. This new circuit is currently under experimental investigation. The upcoming experimental results will allow further validation of the model parameters and may reveal as yet hidden issues in this *in silico* prediction.

Acknowledgment

We thank Elisa Franco and Enoch Yeung for discussion.

References

[1] S. M. Smirnakis, M. J. Berry, D. K. Warland, W. Bialek, and M. Meister, "Adaptation of retinal processing to image contrast and spatial scale," *Nature*, vol. 386, pp. 69-73, 1997.

[2] Y. Tu, T. S. Shimizu, and H. C. Berg, "Modeling the chemotactic response of *Escherichia coli* to time-varying stimuli," *Proc Natl Acad Sci USA*, vol. 105, pp. 14855-14860, 2008.

[3] R. Milo, et al., "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, pp. 824-827, 2002.

[4] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell," *Curr Opin Cell Biol*, vol. 15, pp. 221-231, 2003.

[5] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, "Defining network topologies that can achieve biochemical adaptation," *Cell*, vol. 138, pp. 760-773, 2009.

[6] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon, "The incoherent feedforward loop can provide fold-change detection in gene regulation," *Mol Cell*, vol. 36, pp. 894-899, 2009.

[7] J. Kim, K. S. White, and E. Winfree, "Construction of an *in vitro* bistable circuit from synthetic transcriptional switches," *Mol Syst Biol*, vol. 2, pp. 68, 2006.

[8] J. Kim, J. J. Hopfield, and E. Winfree, "Neural network computation by *in vitro* transcriptional circuits," *Advances in Neural Information Processing Systems*, vol. 17, pp. 681-688, 2004.

[9] J. Kim and E. Winfree, "Synthetic *in vitro* transcriptional oscillators," *Mol Syst Biol*, vol. 7, pp. 465, 2011.

[10] D. Grate and C. Wilson, "Laser-mediated, site-specific inactivation of RNA transcripts," *Proc Natl Acad Sci USA*, vol. 96, pp. 6131-6136, 1999.

[11] H. A. Vincent and M. P. Deutscher, "Substrate recognition and catalysis by the exoribonuclease RNase R," *J Biol Chem*, vol. 281, pp. 20769-29775, 2006.

[12] B. Yurke and A. P. Mills Jr., "Using DNA to power nanostructures," *Genet Program Evol Mach*, vol. 4, pp. 111-122, 2003.