Top-level dynamics and the regulated gene response of feed-forward loop transcriptional motifs

Michael Mayo,[1,*] Ahmed Abdelzaher,[2] Edward J. Perkins,[1] and Preetam Ghosh[2]
[1]*Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, Mississippi 39180, USA*
[2]*Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia 23284, USA*
(Received 13 February 2014; published 10 September 2014)

Feed-forward loops are hierarchical three-node transcriptional subnetworks, wherein a top-level protein regulates the activity of a target gene via two paths: a direct-regulatory path, and an indirect route, whereby the top-level proteins act implicitly through an intermediate transcription factor. Using a transcriptional network of the model bacterium *Escherichia coli*, we confirmed that nearly all types of feed-forward loop were significantly overrepresented in the bacterial network. We then used mathematical modeling to study their dynamics by manipulating the rise times of the top-level protein concentration, termed the induction time, through alteration of the protein destruction rates. Rise times of the regulated proteins exhibited two qualitatively different regimes, depending on whether top-level inductions were "fast" or "slow." In the fast regime, rise times were nearly independent of rapid top-level inductions, indicative of biological robustness, and occurred when RNA production rate-limits the protein yield. Alternatively, the protein rise times were dependent upon slower top-level inductions, greater than approximately one bacterial cell cycle. An equation is given for this crossover, which depends upon three parameters of the direct-regulatory path: transcriptional cooperation at the DNA-binding site, a protein-DNA dissociation constant, and the relative magnitude of the top-level protien concentration.

DOI: 10.1103/PhysRevE.90.032706 PACS number(s): 87.16.A−, 82.40.Bj, 87.10.Rt

I. INTRODUCTION

Networks map how related nodes are connected together by links and are widely employed throughout the physical, life, and social sciences to visualize and interpret information about large, self-interacting systems. In just one example from genetics, a network's nodes can be identified with the genes of DNA, which are interconnected by links that denote causal correlations between the two expression levels. Such networks are termed gene regulatory networks [1]. They have been shown to harbor subnetworks, termed network motifs [2,3], which are found more abundantly than in randomized networks with the same degree sequence. This fact begs the question of whether or not any special functionality can be attributed to these peculiar patterns.

Addressing this question theoretically, Milo and collaborators hypothesized that network motifs function as "simple building blocks of complex networks" [2]. One such motif is the three-node hierarchical feed-forward loop, wherein a top-level gene affects the activity of a target through two regulatory paths: a direct path with a single link, and an indirect path composed of two sequential links [Fig. 1(a)]. These loops have been implicated in some specialized functions, such as fold-change detection [4–7], noise buffering [8], a nonmonotone input dependence [9], and pulse-like "biphasic" qualities [10,11].

Two types of feed-forward loop can be identified, depending on whether the overall regulatory effect of the direct path is the same (coherent) or opposite (incoherent) as the indirect paths' effect on the target gene. The coherent loops are thought to function as delay elements for expression of the regulated gene (node 3; Fig. 1), when compared against the case without the coregulator interaction, termed "simple regulation" in Ref. [10]. In contrast, the incoherent type feed-forward loops are thought to accelerate the expression of the target protein [10]. These conclusions were justified on the basis of computer experiments, wherein each feed-forward loop was perturbed with a binary, ON/OFF transcriptional regulator concentration [10]. However, some predictions of these models have since been confirmed experimentally [12–14].

Here we further investigated the dynamical consequences for protein levels of the regulated gene for each of the eight types of feed-forward loop motif. More specifically, we used mathematical modeling to manipulate rise times of the top-level regulator, here termed the induction time, while also holding its initial and final concentration levels fixed. Although the topology is fixed, the pattern of regulatory interactions varies among feed-forward loops. We show below how these differences result in qualitatively different behaviors depending on if the top-level induction time is either "fast" or "slow."

II. TRANSCRIPTIONAL NETWORKS

A. *Escherichia coli* model network

The *Escherichia coli* (*E. coli*) bacterium provides a prototypical transcriptional network as data sets are well documented [3,15]. We extracted an *E. coli* network from the software package GeneNetWeaver [16]. This network hosted $L = 3648$ links and $N = 1564$ nodes and was sparse: $L/N \approx 2.3$ links per node. GeneNetWeaver refers to a link's regulatory action as stimulatory ("+"), inhibitory ("–"), biregulatory ("+–"), or undocumented ("?"). This extracted network contained some disconnected components; we removed these by pruning, leaving only its largest connected component for further analysis.

B. Feed-forward loops

The feed-forward loop is a hierarchical three-node subnetwork composed of two gene coregulators, wherein one

*Michael.L.Mayo@usace.army.mil

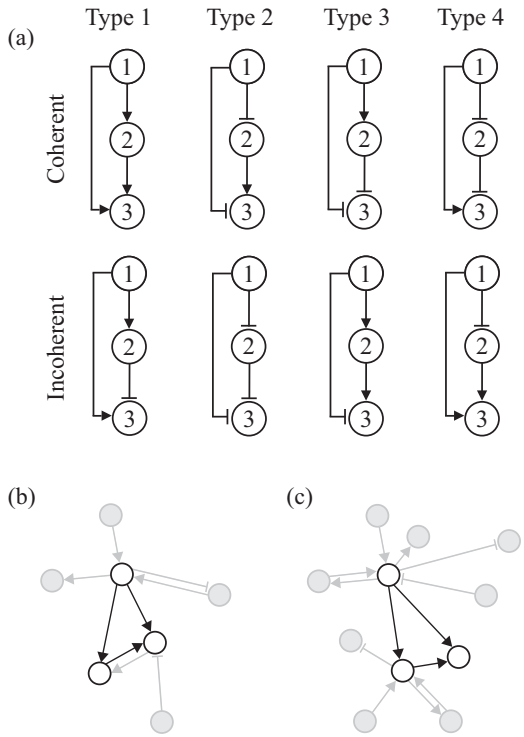

FIG. 1. (a) All eight feed-forward loop transcriptional motifs. The composite sign of the indirect path, (1,2) and then (2,3), is identical to the direct path (1,3) for the "coherent" loops, but is opposite for the "incoherent" loops [10]. The regulatory action of the single (2,3) link differentiates the coherent from the incoherent loops of each type. The search algorithm included counts of (b) embedded feed-forward loops or excluded them to count only the (c) canonical loops.

regulates the other. This structure was originally shown to be overrepresented in *E. coli* gene-regulatory networks inferred from the RegulonDB database [3].

There are eight types of feed-forward loop, each accounting for a possible combination of its three links, which may either up- or down-regulate target expression, as illustrated in Fig. 1(a). The coherent loops are those wherein the "direct path" from node 1 to 3 contributes to expression of the regulated gene (node 3) in the same manner as the whole of the "indirect path," which spans from nodes 1 to 3 but involves the node 2 waypoint. Thus, the incoherent loops are those wherein the direct and indirect paths regulate the gene oppositely.

Herein we differentiate between two classes of feed-forward loop present within a transcriptional network. The first class, termed the "embedded" feed-forward loop, consists of feed-forward loop subnetworks potentially exposed to intramotif interactions [Fig. 1(b)]. The second class, termed the "canonical" feed-forward loop, are those loops that lack any additional links between their nodes [Fig. 1(c)].

C. Randomized networks

We constructed randomized versions of the *E. coli* network, which preserved its degree sequence, to compare against feed-forward loop counts obtained from *E. coli* and used the well-known configuration model for this purpose (e.g., see Ref. [17] for a review). Briefly, each network link was "cut," to leave a number of directed "stubs" adjoined to each node. Links were reconnected at random with uniform probability from the number of remaining stubs. Although this process automatically preserves the degree of each node, it may otherwise admit multiple or self-loops that we are compelled to avoid because they are not present in the feed-forward loop structure. Although autoregulation among *E. coli* genes is common, we did not consider them here because the feed-forward loop is free of self-loops.

The configuration model, as implemented above, does not preserve the regulatory identity of each link. The *E. coli* network used here has 1966 up-regulating (+) links, 1466 down-regulating (–) links, 201 links of mixed regulation (+–), and 15 links of undetermined regulatory identity (?). We developed two strategies to assign these identities to links of each random network. In the first strategy, we kept these totals constant but distributed them uniformly throughout a network, with equal probability until their numbers were exhausted. We term this strategy *random regulation* (RR). In the second strategy, the regulatory identity of each link was simply preserved from that of the "source" node. This strategy was termed *source-preserved regulation* (SPR).

III. THE MATHEMATICAL MODEL

We focus our modeling efforts on the canonical feed-forward loops. At the transcriptional level, we assumed an infinite RNA polymerase pool, so that transcription is rate limited by the levels of DNA-binding proteins. The transition from an open to closed promoter occurs rapidly [18], so we model RNA production as a constant rate [19]. We further assumed an infinite pool of ribosomes, so that protein levels, $S_i(t)$, are limited purely by the transcriptional activity [20]. Protein concentrations can therefore be modeled using the equation

$$\frac{d}{dt}S_i(t) = f_i(\{S\};t) - k_i^{\deg}S_i(t). \tag{1}$$

Here $f_i(\{S\};t)$ expresses the number of proteins created per unit time and volume, which may depend on a number of regulator protein levels, denoted by the set $\{S\} = \{S_j : j = 1,\ldots\}$; $1/k_i^{\deg}$ denotes lifetime of the ith protein species.

A. Modeling protein production

If expression of the ith gene varies with the jth transcriptional activator protein concentration, denoted by $a_{j\to i}$, then f_i can be modeled using a sigmoid-type equation: $f_i(a_{j\to i}) = V_i(a_{j\to i}/K_{ji})^n/[1+(a_{j\to i}/K_{j\to i})^n]$ [10,21]. This is the well-known Hill equation from enzyme kinetics [22]. The parameter V_i sets the (constant) maximum protein production flux, n measures protein cooperation at the promoter, and $K_{j\to i}$ measures the dissociation constant (=1/affinity) of the jth regulator to the ith target promoter/operator [10,21,23]. Repressor protein concentrations, $r_{j\to i}$, inhibit the transcriptional activity, and this effect can be modeled with a similar equation: $f_i = V_i/[1+(r_{j\to i}/K_{j\to i})^n]$ [10,21].

While these equations account for the action of single activator or repressor species, they can be altered to include contributions from multiple activator or repressor species. For

example, the equation for modeling contributions of many activators is given by

$$f_i(\{a_{j\to i}\}_j;t) = V_i \frac{\sum_j (a_{j\to i}(t)/K_{j\to i})^n}{1+\sum_j (a_{j\to i}(t)/K_{j\to i})^n} \quad (2)$$

and employs OR-gate logic [10,24,25]. A gene regulated only by multiple repressors may be modeled using the equation

$$f_i(\{r_{j\to i}\}_j;t) = \frac{V_i}{1+\sum_j (r_{j\to i}(t)/K_{j\to i})^n}. \quad (3)$$

Finally, both types of regulators can act concurrently to modify the expression of a gene. In this case, we employ the AND-gate logic to account for these dual actions:

$$f_i(\{a_{j\to i},r_{j\to i}\}_j;t) = \frac{V_i}{1+\sum_j (r_{j\to i}(t)/K_{j\to i})^n} \times \frac{\sum_j [a_{j\to i}(t)/K_{j\to i}]^n}{1+\sum_j [a_{j\to i}(t)/K_{j\to i}]^n}. \quad (4)$$

The OR and AND logic gates have been used extensively before in similar modeling situations [10,24,25].

B. Modeling assumptions and approximations

Dynamics of each feed-forward loop have been assumed to operate near stable steady-state fixed points, in the long-time limit, because transcriptional networks are more likely to include dynamically stable motifs rather than oscillatory, or source-like, fixed points [26]. The degradation rate, $k_i^{\rm deg}$, or the protein turnover time, $1/k_i^{\rm deg}$, accounts for dilution effects from cell division from a bacterial population, primarily because bacterial proteins are long-lived [27]. Additionally, experimental measurements of transcript half-lives tend on the order of hours (e.g., see Ref. [28]), which roughly coincide with the *E. coli* doubling time, approximately 18–180 min [29,30]. Thus, we take all $1/k_i^{\rm deg} = 1/k^{\rm deg}$ being equal, and this value as the *E. coli* cell-cycle time.

We point out that Eqs. (2)–(4) were not obtained from any reaction mechanism, although similar equations, such as the Hill and Michaelis-Menten kinetic equations, can be derived from a single-site occupancy model [22]. Our primary modeling assumptions are that gene transcription limits the protein yield, and that the protein flux can be parameterized using sigmoid-type equations, which is in agreement with experimental observations (see Ref. [31] and references therein). Although the QSSA reduction may not always agree well with the full model near the steady state [32,33], our phenomenological models are sufficiently simple wherein such variance may not be large. Such models have been applied successfully to dynamically model the results from several experimental systems [12,23,27].

Finally, we note that protein rise times obtained from different coregulator interactions may be directly compared. For example, Mangan and Alon [10] employed similar models of the feed-forward loop to compare its dynamics to the case of "simple regulation," which they take as the feed-forward loop absent the coregulator interaction. However, their models begin from an initial concentration of $\tilde{S}_i(0) = 0$, which may be directly validated with experiments using florescent reporter proteins. Alternatively we analyze feed-forward loop dynamics from the angle of perturbation from the nominal biological steady state, $\tilde{S}_i(0) = 1$, which is representative of *in vivo* conditions. Mangan and Alon [10] normalized the final protein levels to infer dynamic consequences from changing topology. Our results can be similarly normalized after first shifting the maximum protein levels regulated by the loop: $\max\{\tilde{S}_i(\tilde{t})\} - 1$, or by equating steady-state levels between topologically different motifs. This does not change the value of the rise time, but rather dilates or contracts the fold-change levels. This shifting may, however, lead to fold-change levels < 0 when the net result of the loop is repressive of the protein concentration; this effect is not captured by the earlier modeling studies that initialize the protein concentration to zero, such as in Ref. [10].

C. Scaling the model equations

We have scaled each protein concentration, $\tilde{S}_i = S_i/S_i^{\rm bio}$ relative to the value of an initial biological state, $S_i^{\rm bio}$. Conceptually, this concentration is the mean value measured over a living bacterial population operating in the nominal (unperturbed) state. Dissociation constants were similarly scaled: $\tilde{K}_{j\to i} = K_{j\to i}/S_j^{\rm bio}$. Concentrations evolve in elapsed time, $\tilde{t} = tk^{\rm deg}$, where, as outlined above, the quantity $1/k^{\rm deg}$ is understood as a typical *E. coli* cell cycle time (Sec. III B). Finally, the maximum transcriptional activity for the ith species or node was scaled relative to these parameters: $\tilde{V}_i = V_i/k^{\rm deg}S_i^{\rm bio}$.

We note that $\tilde{V}_i$ is not, in principle, independent of the other parameters, because of the steady-state condition: $\tilde{f}_i(\{\tilde{S}(\infty)\};\infty) = 1$. As an example, we applied this restriction to Eq. (4). Solving for $\tilde{V}_i$, gives the equation

$$\tilde{V}_i = \underbrace{\left[1+\sum_l (1/\tilde{K}_{l\to i})^n\right]}_{\text{repressors}} \underbrace{\frac{1+\sum_j (1/\tilde{K}_{j\to i})^n}{\sum_j (1/\tilde{K}_{j\to i})^n}}_{\text{activators}}. \quad (5)$$

A similar result may be obtained for the cases of Eqs. (2) and (3).

These scaling relationships express the protein concentrations in terms of a relative measure: the ratio of the absolute concentration level, $S_i(t)$, to the initial level, $S_i^{\rm bio}$, termed the fold change. There are some advantages to expressing concentrations in these relative units. For example, incoherent feed-forward loops are thought to be superior detectors of the fold-change response over changes in the absolute levels [4]. This ability may contribute to improved reliability in the gene response, given the inherently noisy cellular environment [4,34].

D. Parameter values

Biological fluxes are ubiquitously sigmoid in their dependence on protein concentrations [22], which holds for transcriptional activity [21]. Given this shape, we might expect "normal" values of the scaled dissociation constants to reside near the inflection point, $\tilde{K}_{j\to i} = (n+1/n-1)^{1/n}$, because it offers a staging ground for rapid response to smaller changes in the regulator concentration. For a typical value

of $n = 2$, this equation works out to be $\tilde{K}_{j\to i} \approx 1.73$. So it appears reasonable to experimental measurements to reside near $\tilde{K}_{j\to i} \sim 1$.

This estimate can be refined by appealing to eukaryotic data, as they are readily available. The number of transcription factors within a eukaryotic cell is approximately 7.22×10^4 (95% CI is $[4.2 \times 10^4, 1.24 \times 10^5]$, assuming the data are log-normally distributed) [37]. The volume of a typical nucleus is approximately 337×10^{-18} m^3 (95% CI is $[178 \times 10^{-18}$ m^3, 638×10^{-18} m$^3]$, again assuming a logarithmic distribution for the data) [35,36]. Considering only the means, the concentration of a typical transcription factor protein is approximately 356 nM. Given that a species-wide nonspecific dissociation constant is $\sim$1000 nM [37], we may estimate $\tilde{K}_{j\to i} \sim 1000\ \text{nM}/356\ \text{nM} = 2.81$. Note that these estimates have assumed the differences between eukaryotic and prokaryotic transcription (e.g., chromatin remodeling) does not significantly affect the protein-DNA binding kinetics.

As with these transcription factor data, many biological data are log-normally distributed [38]; so the dissociation constants may be similarly distributed. The above arguments suggest a mean value of approximately $\ln 2.81 = 1.03$. We may use this value to construct a probability density function, $P(\tilde{K}_{i\to j})$, from which to sample parameter values. Data were unavailable to support an estimation for the variance; so, we instead assumed a logarithmic standard deviation of one base. This choice is consistent with parameter values employed in other studies [10].

IV. COMPUTER SIMULATIONS

A. Manipulating top-level regulator rise times

Each feed-forward loop was perturbed by manipulating the time, $\tilde{\tau}_1$, needed for node 1 to rise to half its maximum level, which we refer to here as the induction time. Protein concentrations for node 1 evolve in response to an instantaneous, steplike concentration increase of its activator. The induction time can be manipulated by setting a new value for the degradation constant: $\ln 2/\tilde{\tau}_1$. Thus, $\tilde{S}_1(\tilde{t})$ evolves according to the equation (see Appendix A)

$$\frac{d}{d\tilde{t}}\tilde{S}_1(\tilde{t}) = \frac{\ln 2}{\tilde{\tau}_1}[\tilde{S}_{\max} - \tilde{S}_1(\tilde{t})], \tag{6}$$

wherein $\tilde{S}_{\max}$ is the steady-state value for node 1: $\tilde{S}_1(\infty) = \tilde{S}_{\max}$. Equation (6) can be solved exactly, to give

$$\tilde{S}_1(\tilde{t}) = \tilde{S}_{\max} - (\tilde{S}_{\max} - 1)2^{-\tilde{t}/\tilde{\tau}_1}. \tag{7}$$

Our primary dynamical metric for measuring the expression of the regulated gene (node 3), is the rise time of its concentration profile, $\tilde{\tau}_3$; i.e., the time needed for the protein concentration to reach half of its maximum level.

Figure 2 depicts how the top-level regulator concentration varies dynamically with the induction times, achieved by manipulating the degradation or dilution rate of the node 1 proteins (see Appendix A). Without manipulation, induction time of node 1 would be fixed to $\tilde{\tau}_1 = \ln 2 \approx 0.693$ (thick solid line). All rise times were measured from the point of the instantaneous, steplike activation of the node 1 regulator (dotted line).

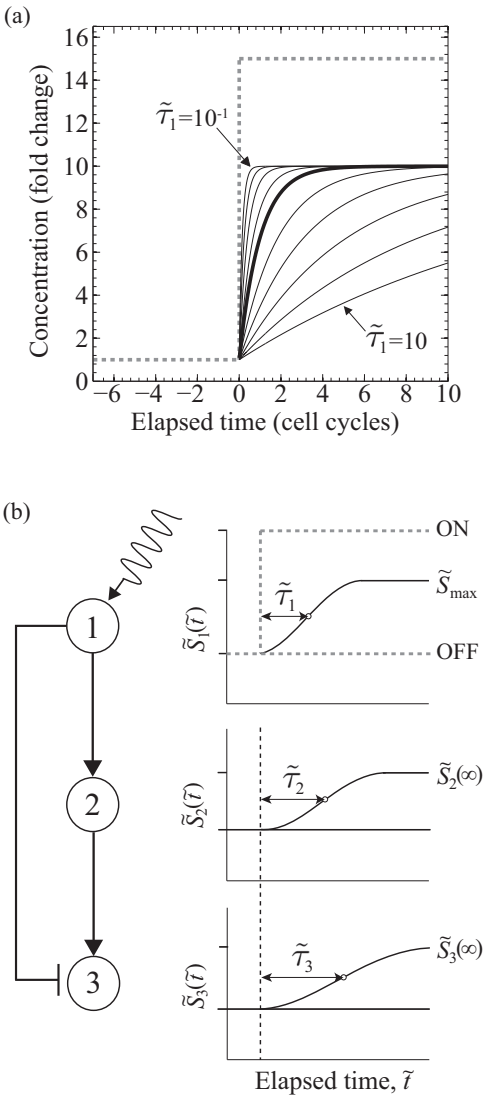

FIG. 2. (a) Concentrations for the top-level regulator, $\tilde{S}_1(\tilde{t})$, of each feed-forward loop, under conditions of varying induction time values, $\tilde{\tau}_1$. Steplike activation of the node 1 regular (dotted line), the point at which all other expression responses are measured, confers a graded response in the node 1 protein production (solid lines). Without manipulation, induction time is fixed, $\tilde{\tau}_1 = \ln 2$ (thick solid line). Here, $n = 2$ and $\tilde{S}_{\max} = 10$. (b) Measurements of rise times for protein concentrations in the feed-forward loop from a steplike activator of node 1 (wavy line).

B. Singular regulation of protein levels

We have also compared dynamics of each feed-forward loop (see Fig. 1) to that of the regulated gene of each loop (node 3), which we consider to be controlled by a single transcription factor. This single-regulated protein level constitutes the direct path of each feed-forward loop (from node 1 to 3). This theoretical setup eliminates contributions to gene 3 expression from the indirect path (node 1 to 2 to 3). To emphasize this comparison, we kept node labels for the single-regulated gene from each of the eight feed-forward loops; i.e., node 3 was controlled only by the concentration of node 1.

C. Expression thresholds and detection limits

Intrinsic and extrinsic noise are two components contributing to variation of the gene expression profile in living bacteria. Extrinsic noise refers to sources of variation such as in the location and concentration of biomolecules or the states of cells, while intrinsic noise refers to variation contributed from microscopic events governing reaction rates of the gene expression machinery [39]. Here we refer to the total noise level, η, which incorporates both noise types, as the ratio of the standard deviation of the signal subject to intrinsic and extrinsic noise to the mean value of the signal [39].

This total noise level affects protein concentrations, and determines to which stimuli the cells respond. Thus, the fold-change level of a transcriptional activator (inhibitor) should rise above (below) a threshold value, at the steady state, for the noise inherent in the transcriptional and translational processes. The overall noise level in *E. coli* cultures has been measured for mutant and wild type genes [39]. We calculated the mean for these reported values, 0.23, by estimating a probability distribution function from the data.

This value can be used to estimate a threshold for the steady-state concentration value of a "detectable" signal. The fold change in a steady-state concentration should be greater than a standard deviation from the initial state, $1+\eta$; however, the lower envelope of the perturbed state is $\tilde{S}_i(\infty)(1-\eta)$. Thus, for a concentration to rise above the ambient cellular noise level, the steady-state concentration for the perturbed state should be large enough such that the difference between them is greater than zero: $\tilde{S}_i(\infty)(1-\eta)-1-\eta>0$. This condition gives a lower bound on the fold-change level of an overexpressed gene:

$$\tilde{S}_i(\infty) > \frac{1+\eta}{1-\eta}. \tag{8}$$

A similar bound can be identified for underexpressed genes, wherein the fold-change level should be reduced in an amount that falls below a value:

$$\tilde{S}_i(\infty) < \frac{1-\eta}{1+\eta}. \tag{9}$$

Putting the averaged value of $\eta = 0.23$ into Eq. (8) gives $\tilde{S}_i(\infty) > 1.6$ for activated genes, or, using Eq. (9), $\tilde{S}_i(\infty) < 0.63$ for the inhibited ones. This value is consistent with bioinformatic estimates that put a threshold on differential expression of approximately 1.5–2-fold [40].

D. Stochastic simulations

For each feed-forward loop, we carried out a total of 10^3 simulations for each value of the induction time, chosen from the interval $[10^{-3},10^3]$. Values for the dissociation constants, $\tilde{K}_{i\to j}$, were sampled randomly, per simulation, from the log-normal distribution of Sec. III D. Rise times of the regulated gene (node 3) were then calculated, and a normal kernel smoothing method [41] was used to estimate a probability density function, $P(\tilde{\tau}_3;\tilde{\tau}_1)$. Finally, a mean rise time, $\langle\tilde{\tau}_3\rangle$, was estimated for each value of the induction time using this density function:

$$\langle\tilde{\tau}_3\rangle = \int_{-\infty}^{\infty} \tilde{\tau}_3 P\left(\tilde{\tau}_3;\tilde{\tau}_1\right) d\tilde{\tau}_3. \tag{10}$$

Brackets, $\langle\cdot\rangle$, thus denote an average over all values for the dissociation constants.

V. RESULTS

A. Feed-forward loop overrepresention in *E. coli*

Table I shows that many feed-forward loops are overrepresented in the model *E. coli* transcriptional network. Statistical significance was determined by calculating a z score for its abundance, α. This was carried out using counts obtained from 100 random networks of each type (RR or SPR). We expressed the z score as the difference between the *E. coli* abundance of an FFL and its mean as counted in the random networks, in units of standard deviations: $(\ln\alpha - \langle\alpha\rangle)/\sqrt{\sigma}$. The quantities $\langle\alpha\rangle$ and σ are, respectively, the mean and variance of the log-normally distributed counts from each type of randomized network.

A majority of the canonical feed-forward loops show counts similar to the embedded loops, indicating that many feed-forward loops found using our algorithm were free of additional internal interactions. However, there were two instances wherein the counts greatly differed. Specifically, we counted 204 embedded incoherent type 2 loops with only 34 canonical ones, but also counted 336 embedded incoherent type 4 loops compared to only 18 canonical ones. Even more surprising was that the embedded incoherent type 4 loops were significantly overrepresented when compared with the both types of random networks, while the canonical incoherent type 4 loop was significantly underrepresented. These numbers indicate that extraneous interactions dominate the counts of embedded I-2 and I-4 loops.

Counts from Table I can be directly compared to those reported from previous studies. For example, Mangan and Alon curated gene-regulatory networks from the literature, for both the *E. coli* bacterium and the model yeast *Saccharomyces cerevisiae* [10]. Their counts substantially differ from those we report in Table I, which can probably be explained by the incompleteness of their networks. However, the relative ordering of their coherent and incoherent feed-forward loops by abundance are substantially different from that of Table I; thus, not all of the count variation can be explained purely by our use of a proportionally larger model network.

Finally, we note that reports of feed-forward loops and their abundances counted from transcriptional networks are somewhat inconsistent. For example, several studies collated all feed-forward loop types and reported the aggregate counts [2,26,42]. In many of these cases the authors use only smaller subsets of the larger available transcriptional networks. There are also differences in the underlying biological definitions of a feed-forward loop. For example, Ref. [43] refers to a feed-forward loop as a four-node structure, with nodes defined as either the DNA promoters or protein regulators. However, we and others (e.g., Refs. [3,10]) refer to feed-forward loops as interactive patterns between the regulated products (i.e., RNA and proteins).

B. Linearity of the rise time: Two dynamic response regimes

Figure 3 illustrates the relationship between the mean rise time, $\langle\tilde{\tau}_3\rangle$, and the induction time, $\tilde{\tau}_1$, for each of the eight

TABLE I. Coherent (C) and incoherent (I) type feed-forward loops (FFLs) for the *E. coli* transcriptional network, counted for both embedded and canonical FFLs. Regulatory interactions were either stimulating, +, or inhibitory, –. For example, C-1 labels the coherent type 1 feed-forward loop. z scores reported for FFL distributions obtained from 100 random networks built using either random-regulation (RR) or source-preserved regulation (SPR) methods. $*$ denotes statistical significance, wherein all z scores > 1.96 ($p < 0.05$).

	Links			Embedded motifs			Canonical motifs		
Type	(1,2)	(2,3)	(1,3)	Abundance, α	RR	SPR	Abundance, α	RR	SPR
C-1	+	+	+	305*	5.22	4.30	264*	5.03	3.96
C-2	–	+	–	186*	4.91	4.84	126*	3.81	3.42
C-3	+	–	–	38	0.5	1.55	36	−0.47	1.56
C-4	–	–	+	91*	2.46	3.06	72	1.90	2.62
I-1	+	–	+	191*	4.78	3.91	166*	4.35	3.66
I-2	–	–	–	201*	5.82	5.95	34	0.38	1.26
I-3	+	+	–	56	0.22	0.78	51	−0.31	0.72
I-4	–	+	+	336*	6.34	5.52	18*	−3.85	−2.49

feed-forward loop types shown in Fig. 1. These curves were found numerically by first applying the cutoff for differential expression [Eqs. (8)–(9), with $\eta = 0.23$), and then calculating an average value for each time point (see Sec. IV D).

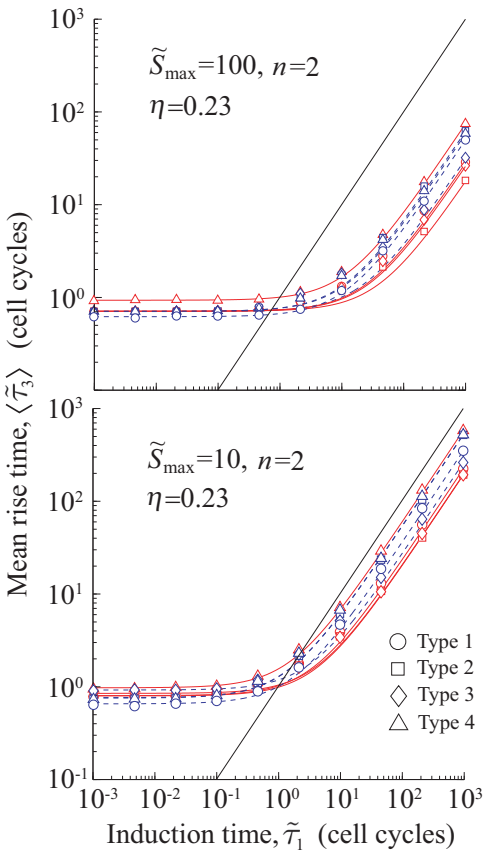

FIG. 3. (Color online) Rise times for the four coherent (red) and incoherent (blue) loops are plotted against induction times for the top-level regulator node 1 (see Fig. 1), and illustrated on a logarithmic scale. Dotted (incoherent FFLs) or solid (coherent FFLs) lines show best-fit solutions to a linear model. The solid black line denotes $\tilde{\tau}_3 = \tilde{\tau}_1$.

The relationship between the induction and rise times were, on average, empirically linear for all eight feed-forward loops (solid and dotted lines), which is consistent with type of response from the case of a single-regulated gene [Eq. (12)]. This result holds despite a larger spread in the top-level concentrations. Values for the fitted parameters have been collected in Table II. Similar relationships have been reported using a detailed mathematical model of gene transcription and protein translation [44].

C. Role of the direct and indirect regulation paths in feed-forward loop dynamics

The particular value of the induction time has direct consequences on the qualitative properties of the feed-forward loop. For example, Fig. 4(a) shows that a large pulse develops dependent on the timing of the inductions; faster inductions show no pulse, while the protein concentration exhibits a large pulse for intermediate induction times. Development of such a pulse can be understood in terms of independent contributions

TABLE II. Parameters for the empirical curves of Fig. 3, which are best fitted to the equation: $\langle\tilde{\tau}_3\rangle = m\tilde{\tau}_1 + B$.

	$\tilde{\tau}^R = B/m$		Slope, m		y intercept, B	
	$\tilde{S}_{max}$		$\tilde{S}_{max}$		$\tilde{S}_{max}$	
Type	10	100	10	100	10	100
C-1	3.59	26.0	0.221	0.0273	0.794	0.711
C-2	4.36	41.7	0.193	0.0171	0.839	0.712
C-3	4.13	28.3	0.191	0.0251	0.787	0.712
C-4	1.67	12.7	0.578	0.0734	0.964	0.935
I-1	1.86	12.6	0.348	0.0492	0.647	0.621
I-2	1.45	11.6	0.521	0.0610	0.754	0.710
I-3	3.53	23.3	0.258	0.0303	0.911	0.706
I-4	1.43	12.4	0.521	0.0572	0.746	0.710

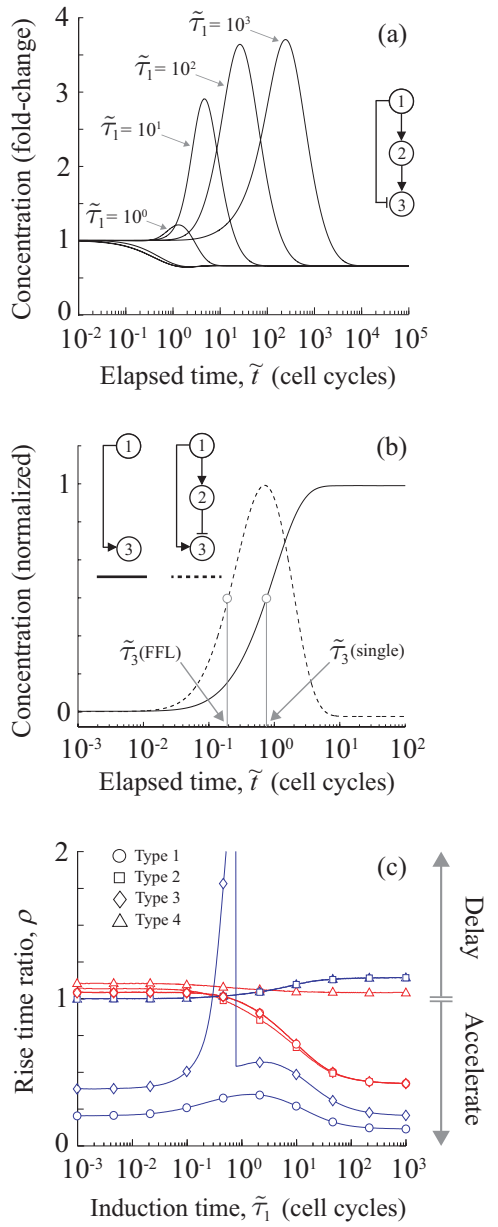

FIG. 4. (Color online) (a) Pulsing dynamics of the I-3 feed-forward loop. (b) Comparison in normalized response between the full I-1 loop and the direct linkage of the I-1 loop. (c) Ratio of rise time in a feed-forward loop, $\tilde{\tau}_3$(FFL), to that of the single, direct linkage, $\tilde{\tau}_3$(single), for coherent (red) and incoherent (blue) feed-forward loops. Values calculated using $n = 2$, $\tilde{S}_{\max} = 10$, and all $\tilde{K} = 2.81$.

from the individual paths of the feed-forward loop topology, and their interaction at the regulation target, gene 3.

Consider the incoherent type 3 feed-forward loop of Fig. 4(a) as an example. Here we may hypothesize that the direct-regulatory path, (1,3), contributes primarily to the down-regulation observed in the expression of gene 3 for rapid inductions, Fig. 4(a), because this link contributes the only down-regulating action. However, as inductions become slower, the early response of the gene 3 profile is to rise, possible only from contributions stemming from the indirect path. Ultimately these rising concentrations are suppressed by the down-regulating action of the direct link.

To test this hypothesis we introduce a new metric, the rise time ratio, ρ, which we define here as the ratio of the regulated genes' rise time from the full feed-forward loop, $\tilde{\tau}_3$(FFL), to the rise time obtained from just the direct path contribution, $\tilde{\tau}_3$(single):

$$\rho = \frac{\tilde{\tau}_3(\text{FFL})}{\tilde{\tau}_3(\text{single})} \tag{11}$$

and calculated for all eight feed-forward loops using the parameter values $\tilde{K}_{j\to i} = 2.81$, $\tilde{S}_{\max} = 10$, and $n = 2$. The direct-path contribution was calculated using the same type of feed-forward loop, after elimination of the indirect-path topology. An example of how each system responds is given by Fig. 4(b).

As shown in Fig. 4(c), this ratio approaches unity in nearly all loops exposed to faster induction times. Thus, most feed-forward loops appear to behave as if the single direct path dominates rise times of the regulated gene under faster inductions. For slower inductions, this ratio either increases or decreases, which we attribute, respectively, to a delay or acceleration of the regulated gene response in the full feed-forward loop relative to that for the single, direct linkage.

For the strong-pulsing incoherent type 3 loop [Fig. 4(a)], the gene-response metric switches from the rise time (time to half-maximum level) to the response time (time to half steady-state level) with increasing induction time. This occurs because the apex of the emerging pulse rises above the steady-state level. When faced with two metrics of gene response, we take the smallest (fastest) one: $\tilde{\tau}_3(\text{FFL}) = \min\{\text{rise time, response time}\}$, which explains the abrupt switching from delay to acceleration in ρ for the type 3 incoherent loop [Fig. 4(c)]. It is not clear whether or not this has any consequence for biology, although it is interesting to note that the incoherent type 3 loop was the only incoherent loop not significantly overexpressed in our *E. coli* network (Table I).

D. Robustness of rise times to rapid inductions of less than one cell cycle

An equation can be found for the approximate rise time, $\tilde{\tau}_j$, of a gene j, up-regulated by a single transcription factor i. As shown in the appendices [Appendix B, Eq. (B21)], this equation is linear:

$$\tilde{\tau}_j = \mu\tilde{\tau}_i + \beta. \tag{12}$$

The slope, μ, and y intercept, β, can be expressed in terms of the parameters $\tilde{S}_{\max}$, $\tilde{K}$, and n. The linearity of Eq. (12) is consistent with results from other studies, such as rise times reported for a single negative autoregulated gene [23].

If we write the right-hand side (RHS) of Eq. (12) as $\beta[(\mu/\beta)\tilde{\tau}_i + 1]$, then the rise time is nearly constant when the induction time falls approximately below a threshold: $\tilde{\tau}_i < \tilde{\tau}^R = \beta/\mu$. This calculation can be carried out using equations of the appendix to give

$$\tilde{\tau}^R = n\ln\sqrt{2}\frac{\tilde{K}^n + 2}{\tilde{K}^n + 1}\left[\frac{\tilde{S}_{\max}}{(\tilde{K}^n + 2)^{1/n}} - 1\right], \tag{13}$$

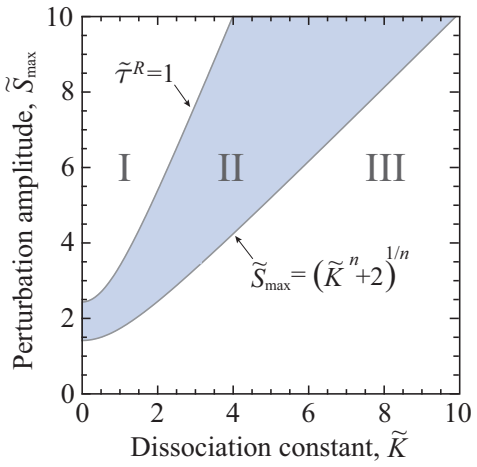

FIG. 5. (Color online) Diagram for regimes of the robustness threshold, $\tilde{\tau}^R$ [Eq. (13)]. Regions I and II illustrate parameters associated with a responsiveness to perturbation, and those of region III lead to approximately unresponsive dynamics. The shaded area, region II, shows parameter values such that $\tilde{\tau}^R$ resides within one cell cycle (fixed $n = 2$).

under conditions of $\tilde{S}_{max} \gg 2^{1/n}$. One equation in the appendices, Eq. (B5),

$$\tilde{S}_{max} > (\tilde{K}^n + 2)^{1/n}, \tag{14}$$

places an approximate bound on the perturbation value that elicits a gene response. This equation quantifies an intuition that transcriptional activators with small DNA-binding affinity require larger fold-change concentration changes to elicit differential expression of the target gene.

An average robustness threshold, $\langle\tilde{\tau}^R\rangle$, can be found with Eq. (13), by calculating over all biological values of $\tilde{K}$, and using the lognormal probability density function from Sec. III D, $\Pi(\tilde{K})$:

$$\langle\tilde{\tau}^R\rangle = \int_0^\infty \tilde{\tau}^R(\tilde{K})\Pi(\tilde{K})d\tilde{K} \approx 2.34. \tag{15}$$

Here we have used values $n = 2$ and $\tilde{S}_{max} = 10$. This result is consistent with the numerical results of Fig. 3 (bottom panel). Note that also $\langle\tilde{\tau}^R\rangle = \langle\beta\rangle/\langle\mu\rangle = \langle\beta/\mu\rangle$.

Figure 5 plots Eq. (13) across the parameter space, assuming dimerized transcriptional regulators ($n = 2$). The dissociation constant has a value fixed by biology, and from the heuristic arguments of Sec. III D, may be near to $\tilde{K} = 2.81$ for most DNA-binding proteins. Thus, the anecdotal value of $\tilde{S}_{max} = 10$ that we have used throughout this work appears poised near, but above, a robustness transition of one cell cycle. Indeed, solving the equation $\tilde{\tau}^R(\tilde{S}_{max}, \tilde{K} = 2.81, n = 2) = 1$ gives $\tilde{S}_{max} \approx 7.23$.

VI. CONCLUSIONS

We used a simplified model of the protein production rate limited by gene expression to manipulate the dynamics of the top-level regulator (node 1) of all eight types of feed-forward loop transcriptional motif. By controlling how soon the concentration of the top-level regulator rises to its maximum value, we observed two qualitatively different regimes in protein level rise times. In the first, rise times remained constant despite rapid top-level inductions, and in the other, rise times varied proportionally with slower induction times. Bacteria may leverage the remarkable stability of rise times attributed to dynamics taking place within a cell-cycle time, which our models suggest provides fault tolerance and reliability of the gene response in the presence of transcriptional noise, a feature noted before in the incoherent loops [8].

By modeling a single activator-gene system representative of the direct link of the feed-forward loop motif, we found the crossover into robustness depends on only three parameter values: transcription factor cooperation at the promoter, n; a dissociation constant, $\tilde{K}$ (=1/DNA-binding affinity); and the fold-change level of the top-level perturbation, $\tilde{S}_{max}$. Given biological parameter values, this threshold works out to approximately one cell-cycle time and is consistent with experimental results [23,27]. Additionally, mathematical models incorporating both transcription and translation exhibit similar trends [44].

Analysis of these models show that the direct-regulation path of nearly all loops, the (1,3) link, determines rise times for faster top-level inductions, while the indirect path contributes with slower inductions. The incoherent type 1 and 3 loops, however, appear to be exceptions to this rule. That dynamics of a single regulatory link can determine the response a downstream gene has implications beyond functionality of single motifs. For example, it may speak to the modularity hypothesis of feed-forward loops: If faster signals are limited to nearest neighbors, then small-node subnetworks may remain dynamically localized, preserving their function despite the presence of long-range interactions. This hypothesis is reminiscent of packet flooding in decentralized wireless networks, wherein information is often transmitted indiscriminately to neighboring nodes, which nearly guarantees reception at the target [45].

Two factors contribute to the timing of the regulated protein yield: the length of time for the transcription factor (top-level) proteins to accumulate beyond a "promoter threshold" of gene activation [27] and the time needed for cellular machinery to first produce the RNA molecules. The latter is a function of the parameter values and is thus fixed. The former depends on the protein lifetime, which we explored here as a variable. For example, if the top-level protein concentration rapidly rises, then the protein production rate is limited by the (fixed) time needed to produce protein by the transcription and translation processes. Thus, the rise time bears almost no dependence on the induction time. If the accumulation of top-level proteins is instead "slow," regulated protein production will be limited by the shorter lifetime of the top-level proteins. The rise time regime is therefore a function of which contribution emerges as the rate-limiting factor the regulated protein yield.

These results are not strictly limited to the *E. coli* bacterium. Indeed, the field of synthetic biology is devoted to engineering modular cell-free systems from constituent biological parts, such as DNA, RNA, and others [46]. Results from these efforts promise the ability to leverage biological "operating principles" to solve novel nonbiological problems. Because our results suggest that dynamical robustness arises automatically from a single gene-transcription factor interaction, without any feedback loop, there is great potential to use current methods of synthetic biology to put these predictions directly to the test.

ACKNOWLEDGMENTS

We thank Ananthram Swami and Christopher Warner for insightful discussions throughout the course of this research. Funding was provided by the US Army's Environmental Quality and Installations 6.1 Basic Research program. Opinions, interpretations, conclusions, and recommendations are those of the author(s) and are not necessarily endorsed by the U.S. Army.

APPENDIX A: DYNAMICAL EQUATION FOR MANIPULATING THE RISE TIMES OF $\tilde{S}_1(\tilde{t})$

In this paper we manipulated rise times of node 1 for each of the feed-forward loops shown in Fig. 1 and observed concentrations of the regulated gene, node 3. Concentrations of node 1, $S_1(t)$, responded to a steplike activator concentration, instantaneously rising from the nominal level to a maximum, as shown in Fig. 2. Node 1 thus responded by moving from the nominal level to its own maximum, $S_{\max}$. Thus, we varied the amplitude of node 1, rather than its steplike activator concentration, because they are both directly related through an equation of type 2. Finally, if node 1 degrades at a rate parameterized by $k'^{\deg}$, then together with Eq. (1), concentrations of node 1 evolved according to

$$\frac{d}{dt}S_1(t) = V_1 - k'^{\deg}S_1(t), \tag{A1}$$

wherein V_1 is constant but depends on the activator concentration. The degradation constant can be expressed in the cell-cycle lifetime $1/k^{\deg}$ (Sec. III B):

$$\frac{d}{dt}S_1(t) = V_1 - \frac{k'^{\deg}}{k^{\deg}}k^{\deg}S_1(t). \tag{A2}$$

Thus, the quantity $k'^{\deg}/k^{\deg}$ refers to a multiple of the cell-cycle time.

The scaling arguments of Sec. III C can be applied to Eq. (A2), to give the equation

$$\frac{d}{d\tilde{t}}\tilde{S}_1(\tilde{t}) = \frac{k'^{\deg}}{k^{\deg}}[\tilde{S}_{\max} - \tilde{S}_1(\tilde{t})]. \tag{A3}$$

Inspection between the exact solution of this equation and $\tilde{S}_1(\tilde{t}) = \tilde{S}_{\max} - (\tilde{S}_{\max} - 1)2^{-\tilde{t}/\tilde{\tau}_1}$, shows that $k'^{\deg}/k^{\deg} = \ln 2/\tilde{\tau}_1$.

APPENDIX B: EQUATION FOR THE RISE TIME OF A SINGLE GENE-TRANSCRIPTION FACTOR INTERACTION

Here we derive an equation relating the response time, $\tilde{\tau}_j$, of an up-regulated gene j to the induction time, $\tilde{\tau}_i$, of its single regulator i. The concentration, $\tilde{S}_i(\tilde{t})$, of the top-level regulator followed the sigmoid of Eq. (7):

$$\tilde{S}_i(\tilde{t}) = \tilde{S}_{\max} - (\tilde{S}_{\max} - 1)2^{-\tilde{t}/\tilde{\tau}_i}, \tag{B1}$$

and the dynamical equation for concentration of the regulated gene, given by Eqs. (1)–(2):

$$\frac{d}{d\tilde{t}}\tilde{S}_j(\tilde{t}) = \frac{1+\tilde{K}^n}{1+[\tilde{K}/\tilde{S}_i(\tilde{t})]^n} - \tilde{S}_j(\tilde{t}). \tag{B2}$$

We dropped labels on dissociation constants because we understand that gene j is regulated by only one activator i.

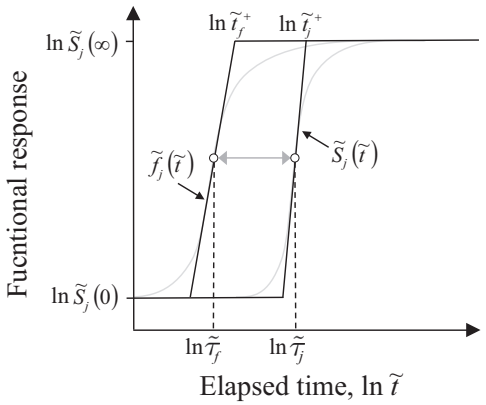

FIG. 6. Illustration of the three-regime piecewise approximations for $\tilde{f}_j(\tilde{t})$ and $\tilde{S}_i(\tilde{t})$. Both axes are logarithmically scaled, wherein straight lines correspond to power laws.

Given Eq. (B2), we first explain some simplifying approximations imposed on Eq. (B2) (Sec. B 1). We next establish a relationship linking $\tilde{\tau}_j$ to $\tilde{\tau}_i$ (Sec. B 2). We further explain how the resulting equation can be solved to get an analytical solution for $\tilde{\tau}_j(\tilde{\tau}_i)$ (Sec. B 3). Finally, we validate the analytic equation by comparing it against the numerical solution (Sec. B 4).

1. Simplifying approximations

The equation

$$\tilde{f}_j(\tilde{t}) = \frac{1+\tilde{K}^n}{1+[\tilde{K}/\tilde{S}_i(\tilde{t})]^n}$$

is sigmoid and so has three distinct power-law regimes (Fig. 6): (1) it is constant for smaller times, $\tilde{t} \leqslant \tilde{t}_f^-$; (2) it is constant for larger times, $\tilde{t} \geqslant \tilde{t}_f^+$, due to promoter saturation; and (3) it approximately obeys a power law in the intermediate region, $\tilde{t}_f^- < \tilde{t} < \tilde{t}_f^+$. We can therefore expand $\tilde{f}_j(\tilde{t})$ for the intermediate regime in a Taylor series about $\tilde{t} = \tilde{\tau}_f$ to first order in the log scale, which gives

$$\tilde{f}_j(\tilde{t}) = \begin{cases} 1 & \tilde{t} \leqslant \tilde{t}_f^- \\ \frac{1}{2}(\tilde{K}^n+2)\left[\frac{\tilde{t}}{\tilde{\tau}_f}\right]^\gamma & \tilde{t}_f^- < \tilde{t} < \tilde{t}_f^+, \\ 1+\tilde{K}^n & \tilde{t} \geqslant \tilde{t}_f^+ \end{cases} \tag{B3}$$

wherein the exponent, γ, is given by

$$\gamma = \frac{n}{2}\left(\frac{\tilde{K}^n}{\tilde{K}^n+1}\right)\left[\frac{\tilde{S}_{\max}}{(\tilde{K}^n+2)^{1/n}} - 1\right] \times \ln\left[\frac{\tilde{S}_{\max}-1}{\tilde{S}_{\max}-(\tilde{K}^n+2)^{1/n}}\right]. \tag{B4}$$

Note that Eqs. (B3) and (B4) place an lower bound on the experimental "control" parameter $\tilde{S}_{\max}$:

$$\tilde{S}_{\max} > (\tilde{K}^n+2)^{1/n}. \tag{B5}$$

This inequality sets the approximate size of the stimulus in the responsive gene-expression regime. For example, given

a high-affinity transcription factor ($\tilde{K} = 0$) that dimerizes ($n = 2$), the smallest amplitude is $\tilde{S}_{\max} = \sqrt{2} \approx 1.41$.

An approximate solution to Eq. (B2) can be found by noting $\tilde{S}_j(\tilde{t})$ is empirically sigmoid in kind with $\tilde{f}_j$. As with $\tilde{f}_j$, we expand $\tilde{S}_j$ in a first-order Taylor series to give

$$\tilde{S}_j(\tilde{t}) = \begin{cases} 1 & \tilde{t} \geqslant \tilde{t}_j^+ \\ \frac{1}{2}(\tilde{S}_j(\infty)+1)\left[\frac{\tilde{t}}{\tilde{\tau}_j}\right]^\xi & \tilde{t}_j^- < \tilde{t} < \tilde{t}_j^+ \\ \tilde{S}_j(\infty) & \tilde{t} \leqslant \tilde{t}_j^- \end{cases}. \tag{B6}$$

Using Eq. (B2), the steady-state amplitude, $\tilde{S}_j(\infty)$, can be written in terms of $\tilde{S}_{\max}$, the promoter dissociation constant $\tilde{K}$ (=1/affinity), and the transcription factor cooperativity, n:

$$\tilde{S}_j(\infty) = \tilde{S}_{\max}^n \frac{\tilde{K}^n + 1}{\tilde{K}^n + \tilde{S}_{\max}^n}. \tag{B7}$$

The exponent ξ is given by

$$\xi = \tilde{\tau}_j \left[\frac{2\tilde{f}_j(\tilde{\tau}_j)}{\tilde{S}_j(\infty)+1} - 1 \right]. \tag{B8}$$

Equation (B8) relies on $\tilde{f}_j$; we avoid further approximation by using the full form of $\tilde{f}_j$ rather than its approximation, Eq. (B4). Thus, an expression for $\tilde{f}_j(\tilde{t} = \tilde{\tau}_j, \tilde{\tau}_i)$ is given by

$$\tilde{f}_j(\tilde{\tau}_j, \tilde{\tau}_i) = \frac{\tilde{K}^n + 1}{1 + (\tilde{K}/\tilde{S}_{\max})^n / G_n(\tilde{\tau}_j, \tilde{\tau}_i)}, \tag{B9}$$

which can be expressed in terms of attenuating function, G_n, that is dependent on the induction and rise times:

$$G_n(\tilde{\tau}_j, \tilde{\tau}_1) = \left[1 + \left(\frac{1 - \tilde{S}_{\max}}{\tilde{S}_{\max}} \right) 2^{-\tilde{\tau}_j/\tilde{\tau}_i} \right]^n. \tag{B10}$$

In summary, we have assumed the transcriptional flux, $f_j(\tilde{t})$ [Eq. (B3)], and the concentration $\tilde{S}_j(\tilde{t})$ [Eq. (B6)] are both sigmoid curves, and exploited this reasoning to approximate them using piecewise equations. In the following section we leverage these approximations to establish a direct relationship between the induction and rise times.

2. Linking the induction and rise times

As can be seen from Fig. 6, the difference between the upper cutoffs is approximately equal to the timing delay between the respective curves for $\tilde{f}_j$ and $\tilde{S}_j$: $\ln \tilde{t}_3^+ - \ln \tilde{t}_f^+ = \ln \tilde{\tau}_3 - \ln \tilde{\tau}_f$. We thus obtain

$$\tilde{t}_3^+ = \frac{\tilde{t}_f^+ \tilde{\tau}_3}{\tilde{\tau}_f}. \tag{B11}$$

It remains to find expressions for $\tilde{t}_f^+$ and $\tilde{t}_3^+$. The former satisfies $\tilde{f}(\tilde{t}_f^+) = \tilde{K}^n + 1$, which can be inverted to give

$$\tilde{t}_f^+ = \tilde{\tau}_f \left(2\frac{\tilde{K}^n + 1}{\tilde{K}^n + 2} \right)^{1/\gamma}, \tag{B12}$$

while the latter can be found by inverting the equation $\tilde{S}_j(\tilde{t}_j^+) = \tilde{S}_j(\infty)$:

$$\tilde{t}_j^+ = \tilde{\tau}_j \left[2\frac{\tilde{K}^n + 1}{\tilde{K}^n(\tilde{S}_{\max}^n + 1)/\tilde{S}_{\max}^n + 2} \right]^{1/\xi(\tilde{\tau}_j)}. \tag{B13}$$

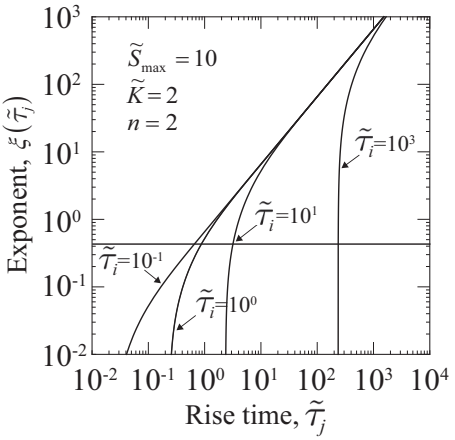

FIG. 7. Plot of the exponent ξ against $\tilde{\tau}_j$ for various values for the induction times, $\tilde{\tau}_i$. Horizontal line indicates the position of the right-hand side of Eq. (B14).

Finally, putting Eqs. (B12) and (B13) into Eq. (B11) gives

$$2\tilde{\tau}_j \left\{ \left(1 - \frac{\tilde{K}^n}{\tilde{K}^n + \tilde{S}_{\max}^n} + \frac{1}{\tilde{K}^n + 1} \right) \left[1 + \frac{(\tilde{K}/\tilde{S}_{\max})^n}{G_n(\tilde{\tau}_j, \tilde{\tau}_i)} \right] \right\}^{-1} - \tilde{\tau}_j$$
$$= \gamma \frac{\ln\left[2(\tilde{K}^n + 1)/(\tilde{K}^n(\tilde{S}_{\max}^n + 1)/\tilde{S}_{\max}^n + 2)\right]}{\ln\left[2(\tilde{K}^n + 1)/(\tilde{K}^n + 2)\right]}. \tag{B14}$$

We note that the right-hand side of this equation is independent of both $\tilde{\tau}_j$ and $\tilde{\tau}_i$; the relationship between them is given entirely by its left-hand side, which is the exponent $\xi(\tilde{\tau}_j)$. The remainder of this appendix is devoted to solving Eq. (B14).

3. Finding a solution for $\tilde{\tau}_j(\tilde{\tau}_i)$

Figure 7 plots the left-hand side of Eq. (B14) across several decades of $\tilde{\tau}_i$. We can see that for larger $\tilde{\tau}_i$, a discontinuity in $d\xi(\tilde{\tau}_j)/d\tilde{\tau}_i$ is approximately coincident with the intersection of $\xi(\tilde{\tau}_j)$ and the right-hand side of Eq. (B14). So, finding the discontinuities provides a good approximation of $\tilde{\tau}_j$ for "larger" $\tilde{\tau}_i$. Carrying this out leads to

$$\tilde{\tau}_j \sim \mu \tilde{\tau}_i, \tag{B15}$$

wherein μ is given by the equation

$$\mu = \frac{1}{\ln 2} \ln \left[\frac{\tilde{S}_{\max} - 1}{\tilde{S}_{\max} - \{\cdot\}^{1/n}} \right]. \tag{B16}$$

Here we have abbreviated

$$\{\cdot\} = \frac{(\tilde{K}^n + 1)\frac{\tilde{S}_{\max}^n + 1}{\tilde{S}_{\max}^n - 1} + 1}{(\tilde{K}^n + 1)\frac{\tilde{S}_{\max}^n + 1}{\tilde{S}_{\max}^n - 1} - \tilde{K}^n}. \tag{B17}$$

Note that $\{\cdot\} \approx \tilde{K}^n + 2$ when $\tilde{S}_{\max}$ is large in the sense of Eq. (B5).

The other asymptote of Eq. (B14), valid near $\tilde{\tau}_i = 0$, can be obtained by putting $G_n(\tilde{\tau}_j, \tilde{\tau}_i) = 1$ into Eq. (B14), and then solving directly for $\tilde{\tau}_j$:

$$\tilde{\tau}_j \sim \beta, \tag{B18}$$

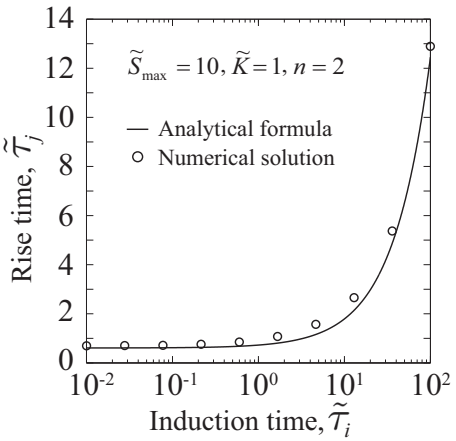

FIG. 8. Plot of rise time, $\tilde{\tau}_j$, against the induction time, $\tilde{\tau}_i$, obtained from numerical integration of Eq. (B2) (circles) or from the analytic formula of Eq. (B21) (line).

wherein β is given by the equation

$$\beta = \frac{n}{2}\left[\frac{\tilde{S}_{\max}}{(\tilde{K}^n+2)^{1/n}} - 1\right]\frac{(\tilde{K}^n+1)^{\frac{\tilde{S}^n_{\max}+1}{\tilde{S}^n_{\max}-1}}+1}{\tilde{K}^n+1} \times \frac{\ln(\cdot)_1 \ln(\cdot)_2}{\ln[2(\tilde{K}^n+1)/\tilde{K}^n+2]}. \quad \text{(B19)}$$

Here we have abbreviated

$$(\cdot)_1 = \frac{\tilde{S}_{\max}-1}{\tilde{S}_{\max}-(\tilde{K}^n+2)^{1/n}} \text{ and}$$

$$(\cdot)_2 = 2 \times \frac{\tilde{S}^n_{\max}-1}{\tilde{S}^n_{\max}} \times \frac{\tilde{K}^n+1}{(\tilde{K}^n+1)^{\frac{\tilde{S}^n_{\max}+1}{\tilde{S}^n_{\max}-1}}+1}.$$

Equation (B19) reduces to

$$\beta = \frac{n}{2}\left[\frac{\tilde{S}_{\max}}{(\tilde{K}^n+2)^{1/n}} - 1\right]\left(\frac{\tilde{K}^n+2}{\tilde{K}^n+1}\right)\ln(\cdot)_1, \quad \text{(B20)}$$

when $\tilde{S}_{\max} \gg 2^{1/n}$. As it should be, β is independent of $\tilde{\tau}_i$. Finally, a solution to Eq. (B14) can be given by the the sum of equations (B15) and (B18):

$$\tilde{\tau}_j = \mu\tilde{\tau}_i + \beta. \quad \text{(B21)}$$

4. Validity of the solution

Equation (B21) is validated numerically by Fig. 8. We integrated Eq. (B5) to estimate the rise time for a number of different induction times (circles). The analytic formula (line), Eq. (B21), overlays the numerical results in very good agreement.

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[3] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[4] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon, Mol. Cell **36**, 894 (2009).
[5] O. Shoval, L. Goentoro, Y. Hart, A. Mayo, E. Sontag, and U. Alon, Proc. Natl. Acad. Sci. USA **107**, 15995 (2010).
[6] Y. Hart, Y. E. Antebi, A. E. Mayo, N. Friedman, and U. Alon, Proc. Natl. Acad. Sci. USA **109**, 8346 (2012).
[7] Y. Hart, A. E. Mayo, O. Shoval, and U. Alon, PLoS ONE **8**, e57455 (2013).
[8] M. Osella, C. Bosia, D. Cora, and M. Caselle, PLoS Comput. Biol. **7**, e1001101 (2011).
[9] S. Kaplan, A. Bren, E. Dekel, and U. Alon, Mol. Syst. Biol. **4**, 203 (2008).
[10] S. Mangan and U. Alon, Proc. Natl. Acad. Sci. USA **100**, 11980 (2003).
[11] D. Kim, Y.-K. Kwon, and K.-H. Cho, Bioessays **30**, 1204 (2008).
[12] S. Mangan, A. Zaslaver, and U. Alon, J. Mol. Biol. **334**, 197 (2003).
[13] S. Kalir, S. Mangan, and U. Alon, Mol. Syst. Biol. **1**, 2005.0006 (2005).
[14] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon, J. Mol. Biol. **356**, 1073 (2006).
[15] H.-W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer, and A.-P. Zeng, Nucleic Acids Res. **32**, 6643 (2004).
[16] T. Schaffter, D. Marbach, and D. Floreano, Bioinformatics **27**, 2263 (2011).
[17] M. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH Verlag, New York, 2005), pp. 35–68.
[18] M. Shea and G. Ackers, J. Mol. Biol. **181**, 211 (1985).
[19] H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. USA **94**, 814 (1997).
[20] D. Nierlich, Science **184**, 1043 (1974).
[21] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC Press, Boca Raton, 2006).
[22] I. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems* (John Wiley & Sons, New York, 1993).
[23] N. Rosenfeld, M. Elowitz, and U. Alon, J. Mol. Biol. **323**, 785 (2002).
[24] J. Hasty, D. McMillen, and J. Collins, Nature (London) **420**, 224 (2002).
[25] N. E. Buchler, U. Gerland, and T. Hwa, Proc. Natl. Acad. Sci. USA **100**, 5136 (2003).
[26] R. Prill, A. Iglesias, and A. Levchenko, PLoS Biol. **3**, e343 (2005).
[27] U. Alon, Nature (London) **8**, 450 (2007).
[28] E. Yang, E. van Nimwegena, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J. Darnell, Genome Res. **13**, 1863 (2003).
[29] H. Bremer, J. Gen. Microbiol. **128**, 2865 (1982).
[30] J. Krebs, E. Goldstein, and S. Kilpatrick (eds.), *Lewin's GENES XI* (Jones and Bartlett Learning, Burlington, 2014).

[31] H. D. Jong, J. Comput. Biol. **9**, 67 (2002).
[32] M. Bennett, D. Volfson, L. Tsimring, and J. Hasty, Biophys. J. **92**, 3501 (2007).
[33] A. Ciliberto, F. Capuani, and J. Tyson, PLoS Comput. Biol. **3**, e45 (2007).
[34] J. E. Ferrell, Mol. Cell **36**, 724 (2009).
[35] G. Maul and L. Deaven, J. Cell Biol. **73**, 748 (1977).
[36] K. Monier, J. Armas, S. Etteldorf, P. Ghazal, and K. Sullivan, Nat. Cell. Biol. **2**, 661 (2000).
[37] M. Biggin, Dev. Cell **21**, 611 (2011).
[38] E. Limpert, W. Stahel, and M. Abbt, BioScience **51**, 341 (2001).
[39] M. Elowitz, A. Levine, E. Siggia, and P. Swain, Science **297**, 1183 (2002).
[40] I. Yang, E. Chen, J. Hasseman, W. Liang, B. Frank, S. Wang, V. Sharov, A. Saeed, J. White, J. Li, N. Lee, T. Yeatman, and J. Quackenbush, Genome Biol. **3**, research0062 (2002).
[41] A. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis* (Oxford University Press, Oxford, 1997).
[42] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Pinter, U. Alon, and H. Margalit, Proc. Natl. Acad. Sci. USA **101**, 5934 (2004).
[43] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, Science **298**, 799 (2002).
[44] R. Murugan, PLoS ONE **7**, e41027 (2012).
[45] K. Viswanath and K. Obraczka, Comput. Commun. **29**, 949 (2006).
[46] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, Mol. Syst. Biol. **2**, 2006.0028 (2006).